

**University of Alberta**

**On Integrating Element Based Heat Capacity Correlations into Process Simulation  
Software - Tackling Pseudo Component Definition**

by

**Sepideh Rajaeirad**

A thesis submitted to the Faculty of Graduate Studies and Research

in partial fulfillment of the requirements for the degree of

**Master of Science**

in

**Chemical Engineering**

**Chemical and Materials Engineering**

©Sepideh Rajaeirad

Fall 2013

Edmonton, Alberta

Permission is hereby granted to the University of Alberta Libraries to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only. Where the thesis is converted to, or otherwise made available in digital form, the University of Alberta will advise potential users of the thesis of these terms.

The author reserves all other publication and other rights in association with the copyright in the thesis and, except as herein before provided, neither the thesis nor any substantial portion thereof may be printed or otherwise reproduced in any material form whatsoever without the author's prior written permission.

## **Abstract**

In the present work, available direct and indirect methods for isobaric liquid heat capacity estimation are evaluated relative to experimental data for diverse hydrocarbon mixtures for which elemental analysis, refinery and molecule based characterization are experimentally accessible. The study focuses on identification of systemic and random deviations between specific calculation approaches and experimental data, based on characterization method so that their preferred range of use in process simulators can be identified. The roles for new elemental composition based predictive heat capacity correlations in particular are explored. The Dastgostar-Shaw correlation for liquid heat capacities (direct) and the Lastovka-Shaw correlation for ideal gases + a departure function (indirect), both based on elemental composition, provide accurate heat capacity values for a broad range of fluids at saturation and are, for example, preferred over the widely used Lee-Kesler correlation for liquids based both on accuracy and range of application. In order to implement these element-based heat capacity correlations into commercial chemical engineering process simulator software, the number of atoms per unit mass, must either be available from experimental measurement, or be estimated from available property data with little deviation because both heat capacity correlations are sensitive to this value.

An API method, available in the literature, systematically underestimates carbon and heteroatom content and leads to biased heat capacity estimates with significant error for the element composition based correlations. Two approaches for estimating the number of atoms per unit mass are reported: one based on regressing liquid density values at 25 °C, molar mass and boiling point, and the other approach is based on applying a feed forward neural network with one hidden layer to the same property data. A data set obtained from the NIST TRC chemistry

webbook comprising 154 organic liquid compounds comprising n-alkanes, n-alkenes, n-alkynes, naphthenics, aromatics, and heteroatom containing compounds are used to train the correlations and the neural networks. 97 compounds were used to regress correlation coefficients. 70, 15, and 15 percent of the data were used to train, test, and validate the neural network, respectively. Both approaches correct the bias and reduce the deviation of the heat capacity estimates and niches for the application of direct and indirect element-based heat capacity correlations for liquids are identified.

## **Acknowledgement**

I express the deepest appreciation to my supervisor, Professor John M. Shaw, who gave me the opportunity to work in his research group and benefit from his guidance, experience and knowledge. His cheery character made these past two years of my life full of joy and good memories. The completion of this thesis would not have been possible without his help.

I thank Dr. Marco Satyro, (Virtual Materials Group) for his advice and suggestions and Dr. Herbert Loria (Virtual Material Group) for his unrelenting support and help. Moreover I would like to acknowledge Dr. Nafiseh Dadgostar, which research was a key element for my thesis, for providing me resources and helping answer all my questions. Also Ms. Mildred Becerra for her help and guidance in the lab.

I am grateful to my friend, Nima Sammaknejad, PhD candidate in Process Control Engineering, for tips and suggestions regarding correlative model development.

I also would like to thank the support of administrative staff and my colleagues of the Petroleum Thermodynamic Group and the Chemical & Materials Engineering Department at the University of Alberta, especially, Ms. Linda Kaert for her help and kindness.

My deepest gratitude goes to my sisters, Dr. Raheleh Rajaeirad and Dr. Fatemeh Rajaeirad, whose encouragement, and support was a constant source of motivation and inspiration. Also, my sweet niece and nephew, Ava and Arvin Alinejad, who brought light and joy to my life.

Last and but not least, I would like to thank my guardian angels, my parents Monir Tavakolibazaz and Mohammadtaghi Rajaeirad, who devoted their love and youth to support and encourage me through all of life's Challenges. I love you, and owe you everything.

# Table of Contents

List of Figures

List of Tables

Nomenclature

1	Introduction	1
1.1	Unconventional Heavy Oil Reserves Overview	1
1.2	Characterization of Oil Sand Ore and Heavy Oil	2
1.3	Heat Capacity Predictive Correlations	5
1.4	Process Simulator Software	5
2	Literature Review	6
2.1	Constant Pressure Liquid Heat Capacity	6
2.2	Techniques for Liquid Heat Capacity Estimation	8
2.2.1	Group Contribution Methods	9
2.2.2	Corresponding State Methods	14
2.2.3	Indirect Method	
2.2.4	Estimation of Isobaric Liquid Heat Capacity in Simulation Software	15
2.2.5	Heat Capacity Prediction Based on Elemental Composition	15
2.3	Elemental Analysis Based on API Approach	18
2.4	Artificial Neural Networks	19
2.5	Data Regression for Coefficients Calculations	22
2.6	Objectives	23
3	Experiments	25
3.1	Methodology	25
3.1.1	Technical Review	25
3.1.2	Principles	26
3.1.3	Applications	30
3.2	Setaram TG-DSC 111 Description	31
3.3	Calibration	34
3.4	Heat Capacity Calculation	35

4	On Transferring New Constant Pressure Heat Capacity Computation Methods to Engineering Practice	37
4.1	Introduction	37
4.2	Experimental Procedure and Set up Conditions	37
4.3	Sample Preparation	40
4.4	Available Method to Predict Similarity Variable	42
4.5	Experimental Heat Capacity for the Model Mixtures	44
4.6	Comparison Between Available Methods for Predicting Isobaric Liquid Heat Capacity	50
4.7	Conclusions and Recommendations	59
5	Predictive Correlation for, $\alpha$ , a Composition Similarity Variable for Organic Compounds	61
5.1	Introduction	61
5.2	Correlation Development to Predict Similarity Variable	63
5.2.1	Dependence of Similarity Variable on Physical Properties	63
5.2.2	Similarity Variable Prediction by Correlation	70
5.2.3	Similarity Variable Database	72
5.2.4	Results	72
5.3	Artificial Neural Network Model	77
5.3.1	Network Design	77
5.4	Comparison	85
5.5	Conclusions	88
6	Overall Conclusions and Future Work	90
6.1	Conclusions and Recommendations	90
6.2	Future Work	91
	References	92
	Appendices	99

## List of Figures

Figure 2-1: Single input neuron(node) schematic structure	20
Figure 2-2: Multiple input neuron schematic structure	20
Figure 2-3: A neural network schematic structure	21
Figure 3-1: TG-DSC 11 apparatus	32
Figure 3-2: Schematic view of the experimental set-up: TG-DSC 111 Setaram	33
Figure 3-3: Experimental step example	38
Figure 4-1: Water isobaric liquid heat capacity: ▲, Data from literature [46]; Δ, Experimental data for the sample mass of 89.30 mg; □, Experimental data for the sample mass of 54 mg	39
Figure 4-2: Water isobaric liquid heat capacity: ■, data from literature [46]; ●, experimental data for the sample mass 89.30 mg with inert gas; Δ, experimental data for the sample mass 89.30 mg without inert gas.	40
Figure 4-4: The deviation of the API procedure in estimating $\alpha$ for different families: ■, n-alkyne; ▲, $C_{10}H_{20}$ isomers; Δ, Aromatic; □, n-alkene; ○, Naphthenic	44
Figure 4-5: Sapphire heat capacity obtained from: —, literature data using the Archer equation; - - -, Experimental data without offset; . . ., Optimized data using offset for a) Mixture 1, b) Mixture 2, c) Mixture 3, and d) Mixture 4.	46
Figure 4-6: Liquid heat capacity for Mixture 1: —, ..., Literature data calculated for ideal mixture (component liquid $C_p$ obtained from NIST [82])	47
Figure 4-7: Liquid heat capacity for Mixture 2: —, ..., Literature data calculated for ideal mixture (component liquid $C_p$ obtained from NIST [82])	48
Figure 4-8: Liquid heat capacity for Mixture 3: —, Optimized experimental data..., Literature data calculated for ideal mixture (component liquid $C_p$ obtained from NIST [82])	49
Figure 4-9: Liquid heat capacity for Mixture 4: —, ..., Literature data calculated for ideal mixture (component liquid $C_p$ obtained from NIST [82])	50
Figure 4-10: Computational matrix for the possible direct and indirect approaches for calculating isobaric liquid heat capacity of liquids	52
Figure 4-11: Isobaric liquid heat capacity of Mixture 1 calculated by various methods: —, Experimental data with the error of $0.02 \text{ J.K}^{-1}.\text{g}^{-1}$ shown as a shaded area; - - -, IGC <sub>p</sub> +APR base departure function (known compound); ▲, IGC <sub>p</sub> +APR base departure function (petroleum cut); □, Lastovka-Shaw IGC <sub>p</sub> +APR base departure function (known compound); —, Lastovka-Shaw IGC <sub>p</sub> +APR base departure function (petroleum cut); Δ, Dadgostar-Shaw $C_p$ (Actual $\alpha$ ); . . ., Dadgostar_Shaw $C_p$ (API base $\alpha$ ); ○, Lastovka-Shaw IGC <sub>p</sub> (API base	55

alpha)+APR base departure function;  $\_ \cdot \_$  , Lee-Kesler  $C_p$

Figure 4-12: Isobaric liquid heat capacity of Mixture 2 calculated by various methods:  $\text{---}$ , 56  
Experimental data with the error of  $0.02 \text{ J.K}^{-1} \cdot \text{g}^{-1}$  shown as a shaded area;  $\text{-- --}$ , IGC<sub>p</sub>+APR  
base departure function (known compound);  $\blacktriangle$ , IGC<sub>p</sub>+APR base departure function  
(petroleum cut);  $\square$ , Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (known compound);  
 $\text{---}$ , Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (petroleum cut);  $\Delta$ , Dadgostar-Shaw  
 $C_p$  (Actual  $\alpha$ );  $\dots$ , Dadgostar\_Shaw  $C_p$  (API base alpha);  $\circ$ , Lastovka-Shaw IGC<sub>p</sub> (API  
base  $\alpha$ ) +APR base departure function;  $\_ \cdot \_$  , Lee-Kesler  $C_p$

Figure 4-13: Isobaric liquid heat capacity of Mixture 3 calculated by various methods:  $\text{---}$ , 57  
Experimental data with the error of  $0.02 \text{ J.K}^{-1} \cdot \text{g}^{-1}$  shown as a shaded area;  $\text{-- --}$ , IGC<sub>p</sub>+APR  
base departure function (known compound);  $\blacktriangle$ , IGC<sub>p</sub>+APR base departure function  
(petroleum cut);  $\square$ , Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (known compound);  
 $\text{---}$ , Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (petroleum cut);  $\Delta$ , Dadgostar-Shaw  
 $C_p$  (Actual  $\alpha$ );  $\dots$ , Dadgostar\_Shaw  $C_p$  (API base alpha);  $\circ$ , Lastovka-Shaw IGC<sub>p</sub> (API  
base  $\alpha$ ) +APR base departure function;  $\_ \cdot \_$  , Lee-Kesler  $C_p$

Figure 4-14: Isobaric liquid heat capacity of Mixture 4 calculated by various methods:  $\text{---}$ , 58  
Experimental data with the error of  $0.02 \text{ J.K}^{-1} \cdot \text{g}^{-1}$  shown as a shaded area;  $\text{-- --}$ , IGC<sub>p</sub>+APR  
base departure function (known compound);  $\blacktriangle$ , IGC<sub>p</sub>+APR base departure function  
(petroleum cut);  $\square$ , Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (known compound);  
 $\text{---}$ , Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (petroleum cut);  $\Delta$ , Dadgostar-Shaw  
 $C_p$  (Actual  $\alpha$ );  $\dots$ , Dadgostar\_Shaw  $C_p$  (API base alpha);  $\circ$ , Lastovka-Shaw IGC<sub>p</sub> (API  
base  $\alpha$ ) +APR base departure function;  $\_ \cdot \_$  , Lee-Kesler  $C_p$

Figure 5-1: a)  $C_p$  predicted by the D-S correlation for:  $\text{---}$ ,  $\alpha=0.15 \text{ mol} \cdot \text{g}^{-1}$ ;  $\text{-- --}$ , 62  
 $\alpha=0.085 \text{ mol} \cdot \text{g}^{-1}$ ;  $\text{---}$ ,  $\alpha=0.215 \text{ mol} \cdot \text{g}^{-1}$ . Absolute (b) and relative (c) deviations from  
 $C_p$  values predicted using  $\alpha = 0.15 \text{ mol} \cdot \text{g}^{-1}$ :  $\text{-- --}$ , for  $\alpha=0.085 \text{ mol} \cdot \text{g}^{-1}$  and  $\text{---}$   
 $\alpha=0.215 \text{ mol} \cdot \text{g}^{-1}$ .

Figure 5-2: Dependence of  $\alpha$  on: a)  $T_b$ , b) MW, c) density . 69

Figure 5-3: Changes of MW,  $T_b$ , and density versus sample number 71

Figure 5-4: The quality of the model (Equation 5-2) over the training data set 74

Figure 5-5: The quality of the model (Equation 5-2) over the test data set 75

Figure 5-6: The absolute residual of the model (Equation 5-2) over the training data set 75

Figure 5-7: The absolute residual of the model (Equation 5-2) over the test data set 76

Figure 5-8: The quality of the model (Equation 5-3) over the test data set 76

Figure 5-9: The absolute residual of the model (Equation 5-3) over the test data set 77

Figure 5-10: ANN regression graph with hidden layer size of 5 81

Figure 5-11: ANN regression graph with a hidden layer size of 6 81



Figure 5-12: The evaluation of the ANN developed to predict  $\alpha$  84

Figure 5-13: Experimental  $C_p$ ,  $\alpha$ , and  $C_p$  predicted by the Dadgostar-Shaw correlation with 88  
similarity variable estimated by; - -, API procedure; . . ., ANN; —, PLS (Equation 5-2) and  
—, actual  $\alpha$  for mixtures a) 1, b) 2, c) 3, d) 4 (see Table 4-1 for compositions).

## List of Tables

Table 2-1: Universal coefficients for Dadgostar-Shaw correlation	16
Table 2-2: Elemental percentages of liquid hydrocarbons based on degree of API	19
Table 3-1: Coefficient for the Archer equation at temperatures higher than 20 °C	36
Table 4-1: The composition of model hydrocarbon mixtures	41
Table 4-2: Some properties of the model mixtures	41
Table 4-3: The temperature range and the mass of sample for each mixture used in DSC experiments	41
Table 4-4 : The compounds used for calculating the API method deviation in $\alpha$ estimation	42
Table 4-5: Deviation of liquid phase constant pressure heat capacity computational approaches from experimental data for mixtures 1-4	59
Table 5-1: Property database used for similarity variable correlation <sup>1</sup>	64
Table 5-2: The RMSE for different combinations of inputs	73
Table 5-3: Universal coefficient for the Equation 5-2 and 5-3	74
Table 5-4: Error evaluation for different input configurations	79
Table 5-5: Error evaluation for different hidden layer sizes	80
Table 5-6: Error evaluation for trained ANN with different training algorithms	82
Table 5-7: Error evaluation for trained ANN with different transfer functions	83
Table 5-8: RMSD and Bias value for Equation 5-2, 5-3 and ANN models over the training data set	85
Table 5-9: RMSD and Bias value for PLS, ANN, and API models over the test data set	86

## Nomenclature

$C_p$	Constant-pressure liquid heat capacity
$H$	Enthalpy
$T$	Temperature
$X_{wi}$	mole or mass fraction
$C_{pl}$	Enthalpy changes with regard to temperature at constant pressure
$C_{\delta l}$	Enthalpy variation of a saturated liquid with temperature
$C_{sat,l}$	Required energy for changing the temperature while keeping the liquid in a saturated state
$V_{\delta l}$	Volume of saturation liquid
$P$	Pressure
$n_i$	Number of groups of type $i$
$w$	Acentric factor
$C_p^\circ$	Ideal gas heat capacity
$T_r$	Reduced temperature
$P_r$	Reduced pressure
$R$	Universal gas constant
$T_c$	Critical temperature
$\alpha$	Similarity variable
$W$	Weight connecting the input to the neuron
$\theta$	Transfer function
$O$	Output
$\emptyset$	Heat flow rate
$A$	Cross section area
$\lambda$	Thermal conductivity
$\Delta l$	Distance between the measured temperature point and the furnace
$S$	Sample
$R$	Reference
$F$	Furnance
$\beta$	Average heating rate
$\emptyset_{FS}$	Heat flow rate from the furnace to the sample
$\emptyset_r(t)$	Time dependent heat flow rate

# 1 Introduction

## 1.1 Unconventional Heavy Oil Reserves Overview

Conventional and unconventional oil are among the main world energy sources. The techniques used for producing or extracting unconventional oil are more complex and energy intensive than those needed for conventional oil [1]. New techniques such as steam injection for the oil sand industries have been developed to increase the efficiency of unconventional oil production [2]. According to OPEC, the largest proven oil reserves including non-conventional oil deposits are in Venezuela (20% of global reserves), Saudi Arabia (18%), Canada (13%) and Iran (9%) [3]. The scale of the unconventional resources dwarfs conventional and more readily produced resources. For example, Canada's oil sands deposits contain estimated 1.7 to 2.5 trillion barrels of heavy oil in place [4]. However, using present technologies and estimated economy conditions, only about 10% (173 billion bbl) can be recovered [4]. What makes Canada's non-conventional oil resource use grow quickly compared to other oil fields in the world is the political stability and the size of the Canadian resource. The unconventional resources are also diverse. Over 95% of Canadian oil reserves comprise 15 separate oil sand deposits located in the province of Alberta (OSAs). The three main deposits are the Athabasca Wabiskaw-McMurray (commonly referred to as the Athabasca Oil Sands), the Cold Lake Clearwater, and the Peace River Bluesky-Gething deposit which occupy an area of 142,000 km<sup>2</sup>. Production methods also vary by location. For the Athabasca oil sands deposit, 80% is expected to be produced by in situ methods and the other 20 % by open-pit mining [5]. In situ bitumen production is currently performed using Steam-Assisted Gravity Drainage (SAGD) a process developed by Roger Butler in 1970 [5]. Other production methods are at various stages of

development and implementation. Thus while these resources are lumped together as unconventional, they are diverse in nature, present in diverse geological environments, and subject to diverse production methods. These resources also pose numerous challenges with respect to technology development and the environment more broadly.

## **1.2 Characterization of Bitumen and Heavy Oil**

Thermophysical property knowledge of heavy oil, bitumen and their fractions not only helps to improve existing technologies. It also helps identify innovative methods for extraction or that recover more oil from reserves, and in particular those approaches that are economically viable because property knowledge is used to populate process and property models. Three types of property knowledge and models underly process development:

1. Transport properties and models (e.g.: viscosity, mass and thermal diffusivity)
2. Equilibrium phase behavior properties and models (PVT phase diagrams, phase density, phase compositions, solubilities, ...)
3. Energy measurements and models ( heat capacity, enthalpies of mixing, vaporization, ...)

Since characteristics and properties of heavy oil are dependent on the level of solvent or water washing arising in situ or ex situ, bacterial degradation of the resource, thermophysical properties are expected to differ with the locality, and the depth of reservoirs. Thermophysical properties of produced fluids are also expected to diverge based on production method and surface facilities employed for separation, transport and refining. Since collecting data for all possible heavy oil, their fractions and various mixtures is impossible, predictive models that permit interpolation and more importantly extrapolation are essential. Typically, these models

comprise correlations that are built upon oil characterization. Measured composition or property data comprise input parameters for these models.

Conventional hydrocarbon mixtures and their fractions can be categorized as well-defined and ill-defined mixtures. In each case, there is a distinct approach for their characterization. A well-defined mixture is a mixture with a known set of molecules present or a mixture defined as a petroleum cut with a narrow spread of possible molecular structures. Physical properties of such mixtures can be obtained from properties of the model components/homologous groups (light n-paraffins, iso-paraffins, olefins, naphthenics, monoaromatics and polyaromatics) by their specific simple mixing rules. Moreover, for well-defined mixtures, specialized structure-property correlations for pure compounds can be used [5]. The compositions of ill-defined petroleum fluids or fractions is not known. Boiling point ( $T_b$ ) based fractionation and bulk property measurements incorporating specific gravity (SG), viscosity, refractive index, carbon-to-hydrogen weight (CH) ratio, or average molar mass of the whole fluid and its fractions are used to characterize them. Fluid properties are then identified in terms of the number of pseudo-components (e.g.: boiling ranges) and generalized empirical correlations in terms of  $T_b$ . Specific gravity is then used for estimation of various properties required for thermodynamic calculations, such as molar mass, critical constants, acentric factor, etc [5].

For fractions with unknown distillation data ( $M > 300 \text{ g}\cdot\text{mol}^{-1}$ ), bulk measurements such as molar mass or viscosity may be used together with specific gravity to estimate basic parameters and physical properties. If specific gravity is not available, refractive index or CH ratio may be employed as an alternate characterization parameter. For fluids possessing a large mass fraction of non-distillable material, this approach becomes less appropriate and modeling of such fluids becomes more empirical, often requiring introduction of fluid specific or reservoir related tuning

parameters. Use of these fluid specific models beyond their narrow scope of development (in another reservoir or at out of scope conditions within the same reservoir) is not advised.

Conventional characterization methods and estimation techniques developed for light oils and well-defined oils do not work for heavy oils and bitumen. Heavy oils contain complex mixtures of hydrocarbon compounds with a broad distribution of molecular sizes and unknown molecular structures. Up to 50% of the heavy oil fluid can be non-distillable. These non-distillable fractions comprise large aromatic, O, N, S and heavy metals containing and frequently multifunctional compounds, and are typically polar. Compositions, molar mass ranges (from hundreds to thousands of  $\text{g}\cdot\text{mol}^{-1}$ ) and molecular structures of the constituents are uncertain and expected to be very complex. Some sub-fractions originated from solubility classes (SARA) or from chromatography are also complex. Measured bulk properties may exhibit hysteresis (depending on the nature of the property). Additionally, since heavy oil fractions are thermally unstable at temperatures higher than 250 – 300 °C, property data are not measurable in a precise manner. Thus, heavy oil, bitumen and their fractions are classified as ill-defined materials and use of correlations for constant-pressure liquid heat capacity such as the Lee-Kesler correlation [6] which rely on  $S_G$ , measured under conditions where bitumen and heavy oil are semi solid, and  $T_b$ , recorded under conditions where bitumen and heavy oil are chemically reactive, or extrapolated from some other property are likely to be unreliable or subject to significant systematic error. Extending correlation and measurement techniques to include accurate representation of ill-defined hydrocarbon fluids is a significant undertaking. The key is to find robust and easily measurable properties that are available for both well-defined and ill-defined fluids alike and that correlate with properties of interest. In this work, the focus is on constant

pressure heat capacity. Not the development of a correlation but on the implementation issues arising in process simulation, as this introduced some surprising challenges and pitfalls.

### **1.3 Heat Capacity Predictive Correlations**

A series of predictive correlations for constant pressure heat capacity of crystalline organic solids, liquids and ideal gases were recently reported [7-10]. Their primary application is to predict the thermal behavior of ill-defined hydrocarbons, where elemental analysis is one of a few certain composition characteristics available that bridge the gap between well-defined and ill-defined hydrocarbon mixtures. Examples include, boiling cuts or solubility classes such as asphaltenes or maltenes, where no models or only primitive and imprecise models are available for estimating heat capacity. However, the precision and accuracy of these correlations has warranted further evaluation, including applications arising in light or well-defined hydrocarbon mixtures, where for example indirect calculation approaches for the heat capacity of liquids (ideal gas heat capacity + an equation of state based departure functions [11]) or direct correlations (such as the Lee-Kesler correlation [6]) are currently implemented in process simulators. Each of these approaches for predicting liquid heat capacity have advantages and disadvantages, and possess different input data requirements. Identification of hierarchies and the best niches for diverse combinations of methods is challenging, as is conveying this complexity to users.

### **1.4 Process Simulation Software**

Commercial chemical engineering software is used to design and optimize chemical processing systems in industries as diverse as oil and gas processing, petrochemicals, bitumen processing, refining, etc. The software is used to model mass and energy flows within equipment



and processes, and to size equipment. Thermodynamic and transport properties are computed iteratively and have a significant impact on outcomes. Accurate correlations with reliable inputs are needed to obtain process or equipment designs that work in the field. These may be based on standard correlations, look up tables or artificial neural networks.

## 2 Literature Review

### 2.1 Constant Pressure Liquid Heat Capacity

Constant-pressure liquid heat capacity, defined as Equation 2-1. [12]:

$$C_p = \left(\frac{\partial H}{\partial T}\right)_p \quad 2-1$$

is one of the characteristic thermophysical properties of fluids. In general, heat capacity is directly related to temperature derivatives of basic thermodynamic functions. Heat capacity values and trends with temperature are widely used in engineering for writing energy balances, in thermodynamics for calculating entropy and enthalpy values, and in thermochemistry for obtaining reaction enthalpies, which are also functions of temperature. Heat capacity knowledge is also needed for evaluating temperature effects for phase and reaction equilibria [13]. Unexpected variation in apparent heat capacity helps detect phase transitions and changes in the structure of solutions. Heat capacity may also be applied for calculating the temperature derivative of vapor pressure and can be used to extrapolate of vapor pressure [14].

For well-defined hydrocarbon mixtures, the heat capacity of the mixture, remote from the critical point for all components,  $Tr < 0.95$ , approaches a simple summation for neighboring members of a homologous series [15]:

$$C = \sum_{i=1}^n X_{w_i} C_i \quad 2-2$$

where  $X_{w_i}$  is the mole or mass fraction and  $C_i$  is the heat capacity of component  $i$  per mole or mass of component, respectively in a mixture with  $n$  components.

Three different terms for liquid heat capacity are normally used:

- $C_{pL}$ : Enthalpy changes with regard to temperature at constant pressure
- $C_{\delta L}$ : Enthalpy variation of a saturated liquid with temperature
- $C_{sat,L}$ : Required energy for changing the temperature while keeping the liquid in a saturated state

The relationship among these three heat capacity definitions is:

$$C_{\delta L} = \frac{dH_{\delta L}}{dT} = C_{pL} + [V_{\delta L} - T \left( \frac{\partial V}{\partial T} \right)_p] \left( \frac{\partial P}{\partial T} \right)_{\delta L} = C_{sat,L} + V_{\delta L} \left( \frac{\partial P}{\partial T} \right)_{\delta L} \quad 2-3$$

All three definitions of liquid heat capacity have similar values except at high reduced temperature where the values diverge. Normally either  $C_{\delta L}$  or  $C_{pL}$  is estimated while  $C_{sat,L}$  is the property measured experimentally [16].

## 2.2 Liquid Heat Capacity Estimation

Methods and techniques for measuring or estimating liquid heat capacity depend on the type of liquid, whether it is pure or a mixture, defined or ill-defined, below or above boiling point, etc. There are published experimental liquid heat capacity data for many pure hydrocarbons and homologous series [13]. There are several techniques for estimating liquid heat capacity directly. They were categorized into four general groups by Reid et al. [16]: theoretical, group contribution, corresponding states, and Watson thermodynamic cycle [17]. Some of these techniques are reviewed here. Liquid heat capacity can also be calculated indirectly using the ideal gas heat capacity and a departure function [28]. This latter approach is the default one, for example, in chemical process simulators.

### 2.2.1 Group Contribution Methods

These methods assume that molecules are composed of interchangeable fragments or groups. Each group contributes to the total molar heat capacity, irrespective of the molecule in which it is found. Interaction among groups are ignored. Johnson and Huang [18] published a liquid phase heat capacity prediction model based on atomic contributions. Chueh and Swanson [19] published an equation for predicting  $C_{pL}$  values based on more complex constituent groups. The error for this method is 2 to 3 %, within a limited temperature range ( $0.7 < T_r < 0.95$ ). Missenard proposed a method based on other structural groups with temperature sensitive values with a limited range of application compositionally (i.e.: it cannot be applied for components with double bonds), and thermally (the temperature range is limited to the range between the freezing point  $\sim 0.4 T_r$  and  $T_r < 0.75$ ) and the error is  $\pm 5$  % [16]. Newer methods, sometimes called second-order additive schemes [20, 21], which account for dissimilar contributions, have been developed. In these methods what an atom is bonded to is considered. An example for these methods is that of Ruzicka and Domalski [22-24]. This type of method allows equation development for estimating liquid heat capacity in the range from the melting point to the boiling point. The general form of this method is:

$$C_{pL} = R \left[ A + B \frac{T}{100} + D \left( \frac{T}{100} \right)^2 \right] \quad 2-4$$

Where R is gas constant and T is the temperature in K. The following are equations for A, B, and D parameters:

$$A = \sum_{i=1}^k n_i a_i \quad , \quad B = \sum_{i=1}^k n_i b_i \quad , \quad D = \sum_{i=1}^k n_i d_i \quad 2-5$$

Where  $n_i$  is the number of groups of type  $i$ ,  $k$  is the total number of different kinds of molecular groups.  $a_i$ ,  $b_i$ , and  $d_i$  are available for 114 various groups listed in [16]. Liquid heat capacity at higher temperatures is not covered by this method.

Good knowledge of the molecular structure of components is necessary for applying these techniques for estimating liquid heat capacity. Therefore, group contribution methods are not applicable for ill-defined hydrocarbon, such as bitumen and heavy oil.

### 2.2.2 Corresponding State Methods

According to van der Waals, the theorem of Corresponding State Principle (CSP) mentions that all fluids, when compared at the same reduced temperature and reduced pressure have almost the same compressibility factor and the deviation from ideal gas behavior are almost the same [25]. Accordingly, methods for predicting liquid heat capacity have been developed:

$$C_p - C_p^\circ = (\Delta C_p)^{(0)} + \omega(\Delta C_p)^{(1)} \quad 2-6$$

where  $(\Delta C_p)^{(0)}$  is the simple fluid contribution,  $(\Delta C_p)^{(1)}$  is the deviation function,  $C_p^\circ$  is the ideal gas heat capacity,  $C_p$  is the liquid heat capacity, and  $\omega$  is the acentric factor. Utilizing these values, the heat capacity departure function  $(C_p - C_p^\circ)$  can be estimated.

Several equations were developed by Lee and Kesler [6, 28] based on Corresponding State Methods. In this method the accuracy of Johnson-Grayson enthalpy correlation had been investigated particularly for high pressure condition and those near critical region. The same relationship was used for the heat capacity estimation of the liquids. Promising results were published for the analytical form of the Lee-Kesler heat capacity departure function for liquid hydrocarbon heat capacity calculations [28]. The improved correlation for heat capacity of liquid

for petroleum fraction is widely used for hand calculations [6]. Normal boiling point and specific gravity, which are generally more available properties, are used in this correlation to characterize petroleum fraction for  $T_r < 0.85$ :

$$C_p = A_1 + A_2T + A_3T^2 \quad 2-7$$

$$A_1 = -1.17126 + (0.023722 + 0.024907 \text{ spgr})K + \frac{(1.14982 - 0.046535K)}{\text{spgr}} \quad 2-8$$

$$A_2 = (-10^{-4})(1.0 + 0.82463K)(1.12172 - \frac{0.27634}{\text{spgr}}) \quad 2-9$$

$$A_3 = (-10)^{-8}(1.0 + 0.82463K)(2.9027 - \frac{0.70958}{\text{spgr}}) \quad 2-10$$

Where  $T_r$  is reduced temperature ( $T/T_{pc}$ );  $T_{pc}$  is pseudo-critical temperature in degree Rankine;  $K$  is Watson characterization factor and  $\text{spgr}$  is specific gravity 60 F/60 F. The errors for this correlation are discussed in Chapter 4.

Based on the theorem of CSP, Watson [17] developed a two parameter correlation for thermodynamic properties of liquids. Later, Reid and Sobel [29] expanded the Watson equation for heat capacity to a three-parameter correlation with the critical compressibility as the third correlating parameter to calculate heat capacity around a critical region. Moreover, a modified method of Watson [17] was developed by Chueh and Swanson [19]

Tyagi [25] observed that expressing the term  $(H^\circ - H_{sv})$  in the Reid and Sobel equations and the Chueh and Swanson equations as a function of reduced temperature and reduced pressure would yield more accurate value for saturated liquid heat capacity. He then proposed an analytical procedure utilizing the enthalpy departure function from the ideal state suggested by Lee and Edmister [30] and Stevens and Thodos [31] to predict the values of  $dH_{\delta L}/dT$  and

$(\frac{dQ}{dT})_{satL}$ . He introduced three methods among which Method 1 seemed to be more promising and described below:

$$C_{\delta L} = \frac{dH_{\delta L}}{dT} = \frac{d}{dT}(H_{satL} - H^{\circ}) + C_p^{\circ} \quad 2-11$$

where  $H_{\delta L}$  is the enthalpy of saturated liquid. The following generalized equation is for isothermal enthalpy difference for pure hydrocarbon described by Lee and Edmister [30]:

$$\begin{aligned} \frac{(H_{satL} - H^{\circ})}{RT_c} = & A_2 - A_3T_r - 2A_4T_r^3 - 6A_5T_r^7 + (A_6 - A_7T_r - 2A_8T_r^3)P_r - 3A_9T_r^4 \\ & + \omega(A_{10}T_r^2 + A_{11} + A_{12}P_r - 3A_{13}T_r^4P_r^2) \end{aligned} \quad 2-12$$

where  $A_i$ s are the generalized constants:  $A_1=6.32873$ ;  $A_2=-8.45167$ ;  $A_3=-6.90287$ ;  $A_4=1.87895$ ;  $A_5=-0.33448$ ;  $A_6=-0.018706$ ;  $A_7=-0.2286517$ ;  $A_8=0.18940$ ;  $A_9=-0.002584$ ;  $A_{10}=8.7015$ ;  $A_{11}=-11.201$ ;  $A_{12}=-0.05044$ ;  $A_{13}=0.002255$

Differentiating Equation 2-12 with respect to temperature gives

$$\begin{aligned} \frac{d(H_{satL} - H^{\circ})}{dT} = & R(-A_3 - 6A_4T_r^2 - 42A_5T_r^6 \\ & + P_r(-A_7 - 6A_8T_r^2) - (12A_9T_r^3) + \omega(2A_{10}T_r - 12A_{13}T_r^3P_r^2)) \end{aligned} \quad 2-13$$

A group contribution method suggested by Rihani and Doraiswamy [32] is used for calculating ideal gas heat capacity,  $C_p^{\circ}$ , in the above method:

$$C_p^{\circ}(group) = a + bT + cT^2 + dT^3 \quad 2-14$$

$$C_p^\circ(\text{compd.}) = \sum a + \sum bT + \sum cT^2 + \sum dT^3 \quad 2-15$$

where  $C_p^\circ(\text{compd.})$  is the ideal gas heat capacity of a compound,  $C_p^\circ(\text{group})$  is the ideal gas heat capacity of a group, and a, b, c, and d are the characteristics of that particular group and the values for different groups are tabulated by Rihani and Doraiswamy [32]. The summation in Equation 2-32 is over all groups in the compound. One can calculate saturated liquid heat capacity by substituting Equation 2-15 and 2-13 in Equation 2-11. The limitation for this approach is, it is good for  $T_r$  between 0.4 and 1 and  $P_r$  value between saturation pressure and 10.

Tyagi [25] also suggested two other methods (details on them can be found in [25]); however, method 1 is slightly superior among others since it proved to be more accurate around critical point. Tyagi's method can be used in a reduced temperature range of 0.4 to 1 for various organic compounds. There are no graphical computations as in the previous Reid-Sobel method, so no additional error is introduced. The relative deviation for this method from the experimental data is reported to be less than 3%, while for the Reid-Sorel method, the average error is more than 10% for the compounds with  $Z_c$  less than 0.23. Furthermore, the error of 8.7% was observed using the Reid-Sobel method for all the compounds the Tyagi method was tested for at reduced temperature range of 0.7-0.95. It should be emphasized that the enthalpy departure function used in this method was developed for pure components. As a result, this approach may not be suitable for ill-defined hydrocarbons and mixtures.

Application of thermodynamic models based on the CSP to predict liquid heat capacity of ill-defined hydrocarbons is also challenging, since the critical properties are not available, the mean molar mass is not known, and the heat capacity of ill-defined hydrocarbon in the ideal gas state, the reference state for such calculations, is unknown. Typically, the molecular structure is



required for application of the methods for estimating these properties, and also in addition, the range of estimated values diverges as molecular mass increases, even in the case of pure n-alkanes [45].

### 2.2.3 Indirect Methods

As it is mentioned, in the indirect method the departure function can be calculated based on an equation of state and the ideal gas heat capacity of the mixture can be estimated from the existing correlations; i.e Equation 2-31.

Bessieres et al [14] conducted a comparison in which, a departure function, estimating the heat capacity of real gases [16], was applied to predict heat capacity ( $C_p$ ) of heavy distillation cuts. Equations of state are needed to get derivatives in the following equation:

$$C_p(P, T) = C_p^\circ(T) + \int_0^V T \left( \frac{\partial^2 P}{\partial T^2} \right)_V dV - T \frac{\left( \frac{\partial P}{\partial T} \right)_V^2}{\left( \frac{\partial P}{\partial V} \right)_T} - R \quad 2-16$$

where R is the gas constant and  $C_p^\circ$  is the ideal gas heat capacity.

Almost all models applied in that study showed that the prediction accuracy decreased as the distillation cut's mean molecular mass (boiling point) increased. Additionally, it was observed that those equations of state with parameter fitted on vapor pressure of high molecular mass of compounds were more reliable in estimating heat capacity of heavy distillation cuts. Consequently, Bessieres et al. [14] proposed that an adjustment of equations of state parameters to heavy compounds properties enhances the accuracy of heat capacity estimates.

## 2.2.4 Estimation of Isobaric Liquid Heat Capacity in Simulation Software

Prediction of thermodynamic properties can be performed by different process simulators; amongst these thermodynamic properties is the isobaric liquid heat capacity. One of those simulators is VMGSim developed by Virtual Materials Group Inc [46]. It has different approaches for calculation of the isobaric liquid heat capacity in terms of defined or ill-defined materials. For defined materials, the isobaric liquid heat capacity is calculated from the departure function method based on a selected property package. In the case of ill-defined materials, defined as pseudo-components in VMGSim, the ideal gas heat capacity is first calculated with the use of pseudo-component user-defined specific gravity and average boiling point by an ideal gas heat capacity estimation method, like the Lee-Kesler correlation described in section 2-2-2; then, the isobaric liquid heat capacity is calculated by the same method used for well-defined materials; i.e. EOS based departure function plus ideal gas heat capacity (Equation 2-33) described in section 2.2.3.

## 2.2.5 Heat Capacity Prediction Based on Elemental Composition

Laštovka and Shaw [10] developed a correlation for ideal gas heat capacity, which is based on the assumption that the ideal gas heat capacity of large molecules is primarily a function of the number of vibrations per mass of the molecule. With a direct relationship between the numbers of vibration modes per mass of a molecule and the number of atoms per mass of the molecule for large molecules, a similarity variable can be defined as:

$$\alpha = \frac{N}{M} = \frac{\sum_{i=1}^n \nu_i}{\sum_{i=1}^n \nu_i M_i} = \frac{\sum_{i=1}^n x_i}{\sum_{i=1}^n x_i M_i} = \frac{\sum_{i=1}^n \frac{w_i}{M_i}}{\sum_{i=1}^n w_i} \quad 2-17$$

Where  $v_i$  is the stoichiometric coefficient for element  $i$  in a compound consisting of  $N$  atoms,  $n$  is the number of elements in the compound,  $M_i$  is the molar mass of chemical element  $i$  (in  $\text{mol.g}^{-1}$ ),  $x_i$  is the mole fraction of element  $i$  in the compound, and  $w_i$  is the mass fraction of element  $i$ .

Based on this similarity variable concept, Dadgostar and Shaw [7] developed a predictive correlation for the isobaric specific heat capacity of liquids applicable for pure organic compounds and ill-defined mixture such as heavy oil, bitumen, and boiling cuts. This correlation employs temperature and similarity variable as inputs, along with six universal coefficients:

$$C_{pL} = a_1 + (a_{21}\alpha + a_{22}\alpha^2)T + (a_{31}\alpha + a_{32}\alpha^2)T^2 \quad 2-18$$

For  $T > 200$  K:

$$a_1 = (a_{11}\alpha + a_{12}\alpha^2) * 24.5 \quad 2-19$$

Six universal coefficients appearing in Equation 2-18 and 2-19 are reported in Table 2-1:

Table 2-1: Universal coefficients for Dadgostar-Shaw correlation

Coefficient	Value
$a_{11}$	-0.3416
$a_{12}$	2.2671
$a_{21}$	0.1064
$a_{22}$	-0.3874
$a_{31}$	-9.8231E-05
$a_{32}$	4.182E-04

The advantage of this correlation over others is the fact that only elemental composition of a liquid is required and the structural information is shown to be of the secondary importance. The correlation was examined using a test data set including liquid organic compounds and the average absolute deviation was calculated to be 0.067 J/g/K. The correlation was also used for ill-defined hydrocarbon liquids, and their heat capacity was estimated to within 6 and 2.8% in the temperature range of 325 to more than 500 K.

In order to compare these new correlations with other widely used methods for calculating liquid phase heat capacity, Virtual Material Group Inc. published a report [47] comparing five different ideal gas heat capacity estimation methods either direct or indirect. For the direct methods the Lee-Kesler correlation for ideal gases [28], Lastovka-Shaw, Twu-Black [95], and API [96] were used and for the indirect method the Dadgostar-Shaw correlation for liquids was used to estimate liquid heat capacity and the heat capacity residual value was estimated by different property packages. Five different assays in VMGSim were characterized in order to estimate their ideal gas heat capacity. The  $\alpha$  which is a required input for the both Lastovka - Shaw and Dadgostar-Shaw correlation was estimated based on API method described in the following section. Accordingly, the average relative deviation of the Dadgostar-Shaw correlation was between 9 to 10 % for the low boiling point pseudo-components to almost 16% for high boiling point compounds with the Advanced Peng-Robinson property package.

### 2.3 Elemental Analysis Based on API Approach

For implementing the Dadgostar-Shaw correlation in simulator software, Virtual Material Group suggested a way for estimating chemical formulae which can be further used for calculating the similarity variable. In this approach based on MW and API gravity of a pseudo-component, if the component has specific gravity lower than 0.85 (35°API) at 15 °C, it is considered as a paraffin-type component and a typical alkane formula is used. If the specific gravity is greater than 0.85 (35°API), the formula is calculated from the percentage of S, N and C/H ratio which are functions of the API gravity of the pseudo-component [15] as shown in Table 2-2. The other approach implemented in the VMGsim software is so-called “content curve”, where the formulas are calculated based on the elemental curves added in the Content Curves tab, and the  $\alpha$  can be calculated accordingly.

Table 2-2: Elemental percentages of liquid hydrocarbons based on degree of API [15]

Gravity (Degree of API)	Sulfur (Percentages by weight)	Inerts (Percentages by weight)	Carbon-to-hydrogen weight ratio
0	2.95	1.15	8.80
5	2.35	1.00	8.55
10	1.80	0.95	8.06
15	1.35	0.85	7.69
20	1.00	0.75	7.65
25	0.70	0.70	7.17
30	0.40	0.63	6.79
35	0.30	0.60	6.50

## 2.4 Artificial Neural Network

Artificial Neural Networks (ANN) as presented by McCulloch and Pitts [48] are created from an approach for developing intelligent systems by modeling the biological structure and functions of human brain which has neurons and axons [49,50]. ANNs are collections of small individual processing units named neurons (nodes) and the information is passed through the neurons by interconnections (axons) [50]. The network is typically consists of three layers: input layer, some hidden layers, and output layer [51].

A single input neuron, shown in Figure 2-1, consists of 5 terms which are all connected to each other by Equation 2-20.

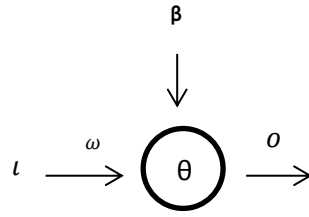


Figure 2-1: Single input neuron(node) schematic structure

$$O = \theta(iw + \beta) \tag{2-20}$$

Where  $i$  is the input,  $w$  is the weight connecting the input to the neuron,  $\theta$  is the transfer function,  $\beta$  is the biased term and  $O$  is the output. Figure 2-2 illustrates a neuron with multiple inputs represented by Equation 2-21.

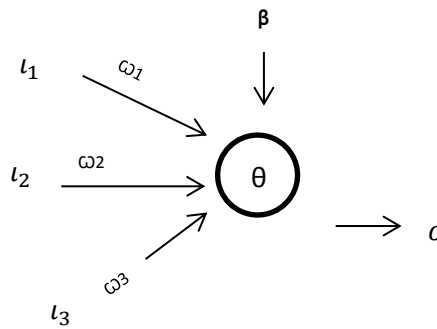


Figure 2-2: Multiple input neuron schematic structure

$$O = \theta(i_1w_1 + i_2w_2 + i_3w_3 + \beta) \tag{2-21}$$

A neural network with one hidden layer is shown in Figure 2-3.

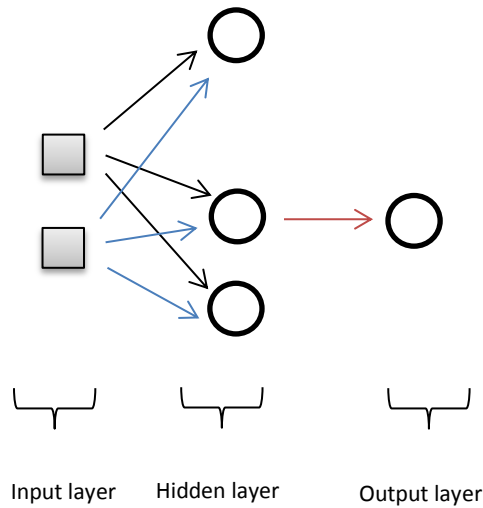


Figure 2-3: A neural network schematic structure

The relationship of inputs and outputs can be represented by a network with biases more easily than a network without biases. There are various transfer functions which can be either linear or non-linear; however, storing non-linear relationships between the input and output can be performed by non-linear transfer functions [50].

Developing a neural network consists of a main stage named training. In the training step the inputs are introduced to the network together with the desired outputs. At this stage, weight terms are adjusted in a way that the desired output can be resulted. The training stage stops when the satisfactory values for the weights are found and the network uses these weights to make decision, to recognize pattern, or to define associations in the test data set [49]. There are several learning algorithms (training function) which can be used to train a network, such as the ones suggested by Haykin [52] or Neocleous [53]; however, the most widely used is the back propagations (BP) and its variants [54, 55]. The training of all patterns of a training data set is



called an epoch. The training set has to cover all the collections of input–output examples. BP training is a gradient descent algorithm. It attempts to improve the performance of the neural network by reducing the total error by varying the weights along their gradients.

Artificial intelligent systems are considered as a technology which can be applied to solve complex and non-linear problems [49]. Nowadays, ANN are used to estimate thermodynamic properties of compounds where it is complex and difficult to predict the properties using analytical equations [56-59]. The advantages of ANN compared to conventional methods are simplicity, speed, and ability to learn from examples.

In this work, ANN will be applied to forecast elemental composition which is a required input for the recent elemental composition based correlations. With the ANN model, simulator software would be able to implement element base correlations for estimating liquid heat capacity with high accuracy; that is why, application of an artificial forecasting system which is able to estimate the property with acceptable error is examined.

## **2.5 Data Regression for Coefficients Calculations**

The most frequently used method in data fitting is the least squares method. In the Least Squares approach, the sum of the squares of the residuals from the equation application is minimized. The residuals are the difference between the observed value and the value predicted by a fitting model [60]. Independent variables in this method can be either single or multiple, and the output of the least squares regression is an equation which is a function of the independent variables and some universal coefficients [61]. A simple regression and least squares method cannot be utilized when there are substantial uncertainties in the independent variables.

The least squares methods are divided into two categories in terms of linearity; linear and non-linear least squares. The regression is called linear when the model includes linear combination of the coefficients; on the other hand, i.e. it is called non-linear when the derivative of model with respect to each coefficient is neither constant nor dependent only on the value of the independent variables [60,61]. Other categories for the least squares method include multiple least square and partial least square. When the independent variables are few in the number, are not collinear, and have a well-understood relationship to the dependent variables, multiple linear regression (MLR) can be a good way to develop a model. However, if any of these three conditions is not met, Partial Least Square (PLS) is used to develop a predictive model. It is important to say that understanding the underlying relationship between independent and dependent variables is not a goal of PLS, i.e. the factors which have negligible effect on the response are not considered by applying the PLS method [62].

## **2.6 Objectives**

Both group contribution based models and corresponding state based models are not applicable to estimate liquid heat capacity for ill-defined hydrocarbons. An element based correlation (Dadgostar-Shaw) which is a function of  $\alpha$ , based on elemental analysis, and temperature was developed to estimate heat capacity of ill-defined liquids directly. Liquid phase heat capacity can also be estimated indirectly from an element based ideal gas correlation [10] + a departure function (indirectly). The dissonance between the values obtained by indirect and direct calculation of liquid phase heat capacity, observed during the implementation of the elemental analysis based heat capacity correlations in VMGSIM, was surprising and led to the inception of this project. The principal objectives of this project are to determine the source of the dissonance and to resolve it. A case study approach is adopted where the test fluids may be

described on a molecular basis, an element basis, or a refinery (boiling range) basis, and where the impacts of elemental composition estimation methods, such as the API method, and departure functions may be evaluated.

## **3 Experiments**

### **3.1 Methodology**

Differential Scanning Calorimetric (DSC) was used in this study to measure the heat capacity of compounds experimentally. It should be emphasized that this method requires careful calibration before each experiment to yield accurate and reliable data.

#### **3.1.1 Technical Review**

Differential Scanning Calorimetric is a well-known thermal analysis technique, which is applied in a wide temperature range in various areas of research, quality inspection and development. The DSC device works based on measuring the difference in the heat flow rate to the sample cell and reference cell while they are subjected to a controlled temperature program. DSC can easily examine heat capacity, heat of transition, kinetic data, and glass transition and purity of a sample. Moreover, DSC curves can be examined to identify substances, to construct a phase diagram and to measure degree of crystallinity [63].

There are two different types of DSC with the same use: the heat flux DSC and the power compensation DSC. The DSC, which is utilized in this study, is a heat flux calorimeter. In this type of DSC, while both reference and sample cells are heated by the same furnace at a given heating rate, the differential heat flow between two cells is measured, which is proportional to the difference in temperature of the cells. [64]. In the power compensation DSC, the reference and sample cells seat on two separated furnaces located inside a single heat sink. The power input to these two furnaces is controlled so that temperatures of the cells are kept identical throughout a given temperature program. The measured signal is the power input difference between the furnaces. [65].

### 3.1.2 Principle

The measured signals in the DSC are the temperature difference and the differential heat flow rate. The relation between temperature difference and differential heat flow rate in real DSCs can be derived based on some assumptions. Steady-state and non-steady-state processes occur in the heat flux DSC. In the steady-state process, it was postulated that there is a constant heat flow rate, only one thermal resistance is applied with no interaction between sample and reference cells, only the heat capacities of the sample and reference cells are considered, and there is no heat loss to the surrounding [66].

The Biot-Fourier equation for heat conduction (steady-state), together with the formulation in absolute values is as follows [63]:

$$\frac{|\bar{\phi}|}{A} = -\lambda \cdot |\text{grad } T| \quad 3-1$$

where  $\phi$ ,  $A$ ,  $\lambda$ , and  $T$  are heat flow rate, cross section area, thermal conductivity, and temperature, respectively. In other words, the heat flux value,  $\phi/A$ , is proportional to the gradient of the temperature while the thermal conductivity,  $\lambda$ , is the proportionality factor.

Equation 3-1 can be rewritten as follows for the sample,  $S$ , and reference,  $R$ :

$$\frac{\phi_{FS}}{A} = -\lambda \cdot \frac{(T_F - T_S)}{\Delta l} \quad 3-2$$

$$\frac{\phi_{FR}}{A} = -\lambda \cdot \frac{(T_F - T_R)}{\Delta l} \quad 3-3$$

where subscript  $F$  stands for furnace and  $\Delta l$  is distance between the measured temperature point and the furnace. In the case of absolute thermal symmetry,  $T_S=T_R$  and  $A$  is identical so that  $\phi_{FS}=\phi_{FR}$ .

$T_S$  increases by  $\Delta T_S$  if a constant heat flow rate  $\phi_r < 0$  is produced in the sample, accordingly, temperature differences,  $T_F - T_S$ , and the heat flow rate  $\phi_{FS}$  decreases. Due to the balance, the steady state will reach again, so the change of  $\phi_{FS}$  ( $\Delta\phi_{FS}$ ) must be equal to  $\phi_r$ :

$$\Delta\phi_{FS} = \phi_r = -\frac{A\lambda}{\Delta l} \cdot \Delta T_S \quad 3-4$$

Since there is no change on the reference side, we have:

$$\Delta T_S = \Delta T_R = T_S - T_R \quad 3-5$$

And,

$$\phi_r = \Delta\phi_{SR} = \phi_{FS} - \phi_{FR} \quad 3-6$$

Consequently:

$$\phi_r = -\frac{A\lambda}{\Delta l} (T_S - T_R) = -\frac{A\lambda}{\Delta l} \Delta T_{SR} = -K \cdot \Delta T \quad 3-7$$

In this model,  $K$  is one of the properties given by the DSC manufacture and belongs to heat conduction path between the furnace and the samples. It leads to a direct proportionality between the measured  $\phi$  and the measurement signal  $\Delta T$ . The constant heat consumption conditions can be obtained in monitoring operations when the sample and the reference sample have different “heat capacities”. A higher amount of heat will always go into the sample whose heat capacity is greater, in order that the steady-state heating rate is constant. With the heat capacity of the sample higher than heat capacity of the reference ( $C_{p,S} > C_{p,R}$ ) the following equation is applied for the difference between the heat flow rates to the sample and reference:

$$\Delta\phi_{SR} = -K \cdot \Delta T_{SR} \quad 3-8$$

The above approximation could not be assumed if there is no steady state during sample transitions or reactions, moreover, the  $\Delta\phi_{SR}$  might change with temperature, but these changes

are in many cases quite slow and do not affect the steady-state condition considerably, i.e. so called quasi steady state. In this case, the following equation can be used:

$$\Delta\phi_{SR} = \beta (C_{p,S} - C_{p,R}) \quad 3-9$$

Or,

$$\beta (C_{p,S} - C_{p,R}) = -K \cdot \Delta T \quad 3-10$$

Where  $\beta$  is the average heating rate.

Equation 3-10 is the basic equation to measure the sample heat capacity ( $C_{p,S}$ ) using a heat flux DSC. Practically, the asymmetry of the device should be checked first by a zero line  $\Delta T_0$  that is recorded with both crucibles empty and subtracted from the measured curves.

In the non-steady state process, except for the  $\Delta T$  which is not constant in time, other assumptions can be used as for the steady-state process. In this case, the equation for the sample heat capacity is as follows:

$$C_{p,S} \frac{dT_S}{dt} = \phi_{FS} - \phi_r \quad 3-11$$

Where  $\phi_{FS}$  is the heat flow rate from the furnace to the sample,  $\phi_r(t)$  is the time dependent heat flow rate produced inside the sample (reaction, transition).

With  $\Delta T = T_S - T_R$ , Equation 3-10 becomes:

$$C_{p,S} \frac{dT_R}{dt} + C_{p,S} \frac{d\Delta T}{dt} = \phi_{FS} - \phi_r \quad 3-12$$

For the reference sample we have ( $\phi_r = 0$  by definition):

$$C_{p,R} \frac{dT_R}{dt} = \phi_{FR} \quad 3-13$$

By subtracting two balance equations, the following is obtained:

$$\Phi_{FS} - \Phi_{FR} = (C_{p,S} - C_{p,R}) \frac{dT_R}{dt} + C_{p,S} \frac{d\Delta T}{dt} + \Phi_r \quad 3-14$$

We have the following expressions for the heat flow rates  $\Phi_{FS}$  and  $\Phi_{FR}$ :

$$\Phi_{FS} = \frac{T_F - T_S}{R_{FS}} \quad 3-15$$

$$\Phi_{FR} = \frac{T_F - T_R}{R_{FR}} \quad 3-16$$

Where  $R_{FS}$  and  $R_{FR}$  are the global heat resistances between the furnace and the samples and the furnace and the reference, respectively. If there is a thermal symmetry  $R_{FS} = R_{FR} = R$ , thus, Equation 3-14 becomes:

$$\Phi_r = -\frac{\Delta T}{R} - (C_{p,S} - C_{p,R}) \frac{dT_R}{dt} - C_{p,S} \frac{d\Delta T}{dt} \quad 3-17$$

The asymmetry of the measuring system is taken into account by the second term as the difference between heat capacities of the sample and the reference cells. The contribution of the thermal inertia of the system is considered in the third term when a measured signal  $\Delta T(t)$  appears. Similarly to the charging or discharging a capacitor of capacity  $C_p$ , a time constant  $\tau$  can be defined for the heat flow rates in the same way:

$$\tau = C_{p,S} \cdot R \quad 3-18$$

Where  $R$  is the effective thermal resistance to the charging or discharging the heat capacity ( $C_{p,S}$ ). With  $\Delta T$  changing in time, with this resistance and with the heating rate defined as  $dT_R/dt = \beta$ , as the reference sample is usually in a steady-state heating mode, the following equation results from Equation 3-17:

$$\Phi_r(t) = -\frac{\Delta T(t)}{R} - (C_{p,S} - C_{p,R}) \cdot \beta - \frac{\tau}{R} \cdot \frac{d\Delta T(t)}{dt} \quad 3-19$$

The temperature dependence of thermal resistance  $R$  and heat capacities ( $C_{p,S}$  and  $C_{p,R}$ ) is reflected by the second term. This causes the temperature dependence of the measured curve



even without any thermal effect in the sample. The third term in Equation 3-19 should be considered when the signal  $\Delta T$  measured in time is to be assigned to the heat flow rate by which it is created. The time constant ( $\tau$ ) and thermal resistance ( $R$ ) must be measured by calibration.

The following equation can be used for the overall heat of reaction or transition ( $Q_r$ ) which are produced or consumed in the sample.

$$Q_r = \int_{t_1}^{t_2} \phi_r(t) dt \quad 3-20$$

Where  $t_1$  and  $t_2$  are the beginning and end of the peak, respectively. Inserting Equation 3-19 into 3-20, we have:

$$Q_r = -\frac{1}{R} \left[ \int_{t_1}^{t_2} \Delta T(t) dt - \int_{t_1}^{t_2} (-R \cdot \Delta C \cdot \beta) dt \right] - \int_{t_1}^{t_2} \left( \frac{\tau}{R} \cdot \frac{d\Delta T}{dt} \right) dt \quad 3-21$$

For the partial integration of the peak between  $t_1$  and  $t^*$ :

$$Q_r = -\frac{1}{R} \left[ \int_{t_1}^{t^*} \Delta T(t) dt - \int_{t_1}^{t^*} (-R \cdot \Delta C \cdot \beta) dt \right] - \int_{t_1}^{t^*} \left( \frac{\tau}{R} \cdot \frac{d\Delta T}{dt} \right) dt \quad 3-22$$

The partial integration of peaks is important for kinetic evaluation and to specify the purity of a sample [63].

### 3.1.3 Application

Differential Scanning Calorimetry is the most widely applied thermal technique which can be used in the study of oxidative stability, liquid crystals, food science, drug analysis, polymers, etc. With the use of DSC output signals, the heat flow rate as a function of temperature and any other derived quantity, such as the heat of reaction or transformation, or changes in heat capacity of a sample can be studied to figure out the properties of a substance.

Different types of DSC and thermal analysis instrumentation are offered by different instrument manufactures, such as Setaram Instrumentation, Netzsch Instruments, TA

Instruments, PerkinElmer Instruments and Mettler Toledo, depending on the research to be conducted. The DSC utilized in this study is Setaram TG – DSC 111, which is a heat-flux DSC [67 – 71].

### **3.2 Setaram TG-DSC 111 Description**

The TG – DSC 111 thermo-analyzer from Setaram is made up of the CS 32 processing unit and the assembly coupling the B111 microbalance to the DSC 111 calorimeter as seen in Figure 3-1.

The CS 32 controller mainly includes a power supply card, a CPU card, an amplification card for calorimetric signal, a balance card, and a temperature acquisition card for temperature regulation, a temperature acquisition card for temperature measurement.

The B111 electronic microbalance is a beam balance connected to a torsion band located between two springs under load. Variation in mass is measured by current variation which has a proportional relationship to the force of electromagnetic equilibrium. A potential difference proportional to the equilibrium current is magnified and is ready for digital use in the CS 32 controller.



a) Calorimeter



b) Electronic Microbalance



c) Processing unit



Figure 3-1: TG-DSC 11 apparatus

The DSC 111 calorimeter includes a junction box, a calorimetric transducer, and pre amplification and amplification cards for the DSC signal. The calorimetric transducer has two sintered alumina tubes with the inner diameter of 7 mm parallel to each other. The tubes are open from both sides and only the centre has sensitive elements. On the front panel of the working chamber, there are two pipes (inlet and outlet) for the refrigerant cooling the calorimeter. A schematic view of the DSC instrument is shown in Figure 3-2.

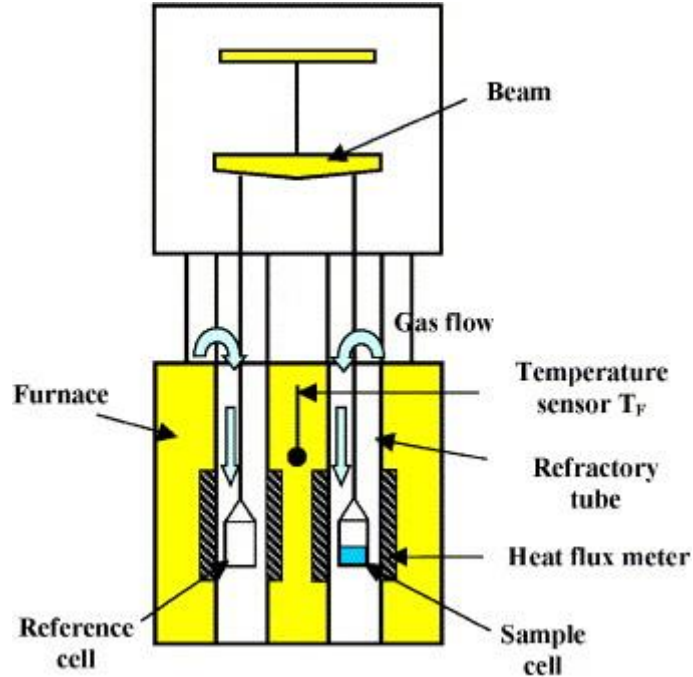


Figure 3-2: Schematic view of the experimental set-up: TG-DSC 111 Setaram [69].

The central area in each tube is a sensitive part of the calorimeter. The center of the calorimetric block includes two cavities in which thermocouple-carrying heat-flux transducers are positioned around the central part of the tubes. The heat exchange between the furnace and the cell takes place only through the thermocouple-carrying heat-flux transducers and, as a result, can be monitored precisely. The close-to-symmetrical arrangement of the transducer almost cancels the signal coming from the two heat-flux transducers being located in oppositions, when the two samples are themselves alike irrespective of the thermal state of the calorimeter. Thus, the working signal includes the passive component in the properties of an active sample set in one of the tubes and of a “reference” placed in the other tubes plus a small asymmetry correction determined by proper calibration.

### 3.3 Calibration

In contrast to adiabatic calorimetry, DSC/DTA instruments are not absolute measuring instruments; heat and heat flow rate are measured dynamically and always yield relative values that must be turned to the absolute values. The setting of the instrument parameter and the sample studied have a strong influence in DSC measurements. It is essential to examine all experimental parameter and to calibrate the device before conducting experiments. The definition of calibration is the measuring of a quantitatively defined relationship between a value of a quantity indicated by the measuring device and the actual value. In DSC the quantities of interest are temperature and heat flow rate. Two calibrations must be carefully conducted: one is the calibration verified by the manufacture and the other is necessary to check the reproducibility, accuracy and precision of the measurement and conducted before each single experiment. Device asymmetry and any other non-linearities should be fixed by this calibration [73].

The group “Calibration of Scanning Calorimeters” of the German Society of Thermal Analysis (GEFTA) [71-73] suggested several types of calibration for a DSC calorimeter. According to their recommendation, the temperature calibration to ITS 90 was conducted using indium (NIST standard reference material 2232), tin (NIST SRM 2220), lead (NIST standard reference material 1059c) and aluminum (NIST standard reference material 854). Energy calibration was carried out in the factory applying the Joule effect method and examined by measuring the heat of fusion of naphthalene, which was a basic reference material for the heat of fusion measurements suggested by International Confederation for Thermal Analysis and Calorimetry ICTAC [75]. The accuracy was within 2% if compared with the literature value [74,76,77]. Heat capacity  $C_p$  (heat flux) calibration was conducted using synthetic sapphire,

which was a basic reference material according to NIST (SRM 720) and ICTAC, and naphthalene, a secondary reference material for  $C_p$  measurements suggested by ICTAC [75]. The uncertainty of  $C_p$  measurements was obtained to be less than 2 % (0.02 J/g/K) in the temperature range from 300 K to 560 K.

Another calibration, different from the general calibration mentioned above, was performed before each experiment for checking the consistency and accuracy of the measured data and for correcting it by adding an offset value. For this calibration, two different masses of synthetic sapphire were used; one as a reference material and one as a sample, and the final heat capacity of the second sample was compared to the literature values [78].

### 3.4 Heat Capacity Calculation

The three-step procedure was applied for DSC measurements to measure heat capacity values. The measuring cell was empty in the first run (run 1), then filled with the reference material (run 2) (synthetic sapphire) and the measured sample (run 3) in the second and third runs, respectively. The reference cell was empty during all runs and each runs was repeated three times for the higher accuracy. As mentioned above, run 2 was divided into two sub-runs with sapphire of two different masses to ensure the accuracy of the measurements.

The heat capacity of a sample is calculated using the procedure utilizing the following output results from DSC: heat flow in J/s, temperature in Kelvin, and time in seconds. The equation can be expressed as [79]:

$$C_{p_{Sample}}(T) = \frac{HF_{sample} - HF_{Blank}}{HF_{Sapphire} - HF_{Blank}} * \frac{Mass_{Sapphire}}{Mass_{Sample}} * C_{p_{Sapphire}}(T) \quad 3-23$$

Where  $HF_{\text{blank}}$  is the heat flow from run 1 (empty sample cell),  $HF_{\text{sapphire}}$  is the heat flow from run 2 (sample cell with sapphire of  $Mass_{\text{Sapphire}}$ ), and  $HF_{\text{sample}}$  is the heat flow from run 3 (sample cell with a sample to be studied with  $Mass_{\text{Sample}}$ ). The value for  $Cp_{\text{sapphire}}$  is obtained from Equation 3-24 named the Archer equation [78]:

$$Cp_{\text{sapphire}} = aT^6 + bT^5 + cT^4 + dT^3 + eT^2 + fT + g \quad 3-24$$

For temperatures higher than 20 °C, the Archer values are listed in Table 3-1.

Table 3-1: Coefficient for the Archer equation at temperatures higher than 20 °C

Coefficient	Value
<i>a</i>	$1.197441280319 \times 10^{-17}$
<i>b</i>	$-2.5923466515291 \times 10^{-14}$
<i>c</i>	$1.3104884522373 \times 10^{-11}$
<i>d</i>	$1.1963323706663 \times 10^{-8}$
<i>e</i>	$-1.8121828407681 \times 10^{-5}$
<i>f</i>	$9.2237456478216 \times 10^{-3}$
<i>g</i>	-0.73178005598711

## **4 On Transferring New Constant Pressure Heat Capacity Computation**

### **Methods to Engineering Practice**

#### **4.1 Introduction**

Constant pressure heat capacity was measured experimentally for a number of model hydrocarbon mixtures and then compared with the estimated liquid heat capacity predicted by Dadgostar-Shaw correlation [7], Lee-Kesler correlation [6], and ideal gas heat capacity based correlations plus equation of state based departure function [11].

#### **4.2 Experimental Procedure and Set up Condition**

Experimental isobaric liquid heat capacity data were measured using a differential scanning calorimeter, TG-DSC 111. The measurements were carried out with a heating rate of 20 K/min, appeared to be the best scanning rate with smaller noises at the final signal. The isothermal period was 60 minutes at the beginning and the end of each experiment as presented in Figure 4-1. It was proven that for liquids the difference between isobaric heat capacity and saturation heat capacity is negligible as long as the upper temperature limit of the measurements is less than boiling point of the mixture (about  $0.9 T_b$ ) [79]. Thus, the temperature range in this work was from 293 K to  $0.9 T_b$  in order to avoid artifacts introduced by sample vaporization. The sample material was enclosed in a recyclable crucible made of stainless steel with volume of  $100 \text{ mm}^3$ . The crucible was sealed with a nickel ring and a stainless steel lid. This sealed crucible withstands an internal pressure up to 20 bar.



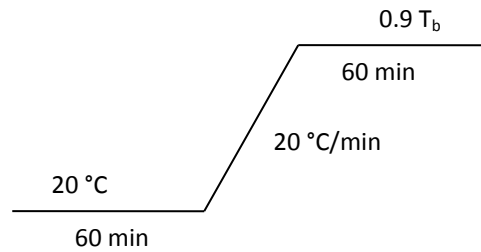


Figure 4-1: Experimental step example

A possible impact of sample vaporization was further reduced by performing a trial experiment with water to find the volume of the crucible which should be filled with the sample. The result, illustrated in Figure 4-2, showed that depending on the sample density, just 10 mm<sup>3</sup> of total volume should be left empty and filling less or more than that might cause error in the measurement. Crucibles were weighed prior to and after each experiment. No mass loss of the samples occurred. Another trial experiment with the reference material was performed to see if there is a need of having constant flow of inert gas during an experiment or not. The result showed that having purge gas flow causes more error in the final results as it is shown in Figure 4-3. Accordingly, although it is suggested to have a constant flow of purge gas during an experiment in Setaram TG – DSC 111, it is better not to have any inert gas flow through the calorimeter tubes during the experiment.

According to the DSC device instruction manual, the systematic error with DSC measurement is around +/- 0.05 J.K<sup>-1</sup>.g<sup>-1</sup>, so that this amount of error with the experimental result is anticipated.

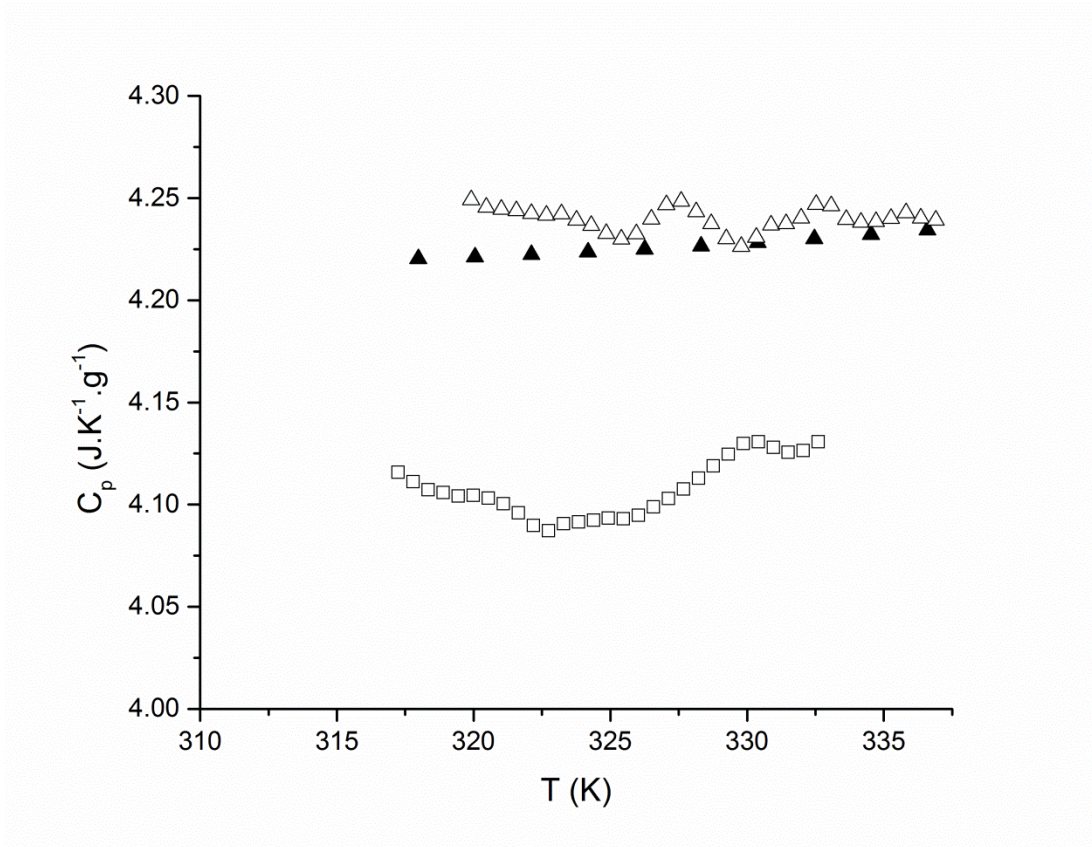


Figure 4-2: Water isobaric liquid heat capacity: ▲, Data from literature [46]; △, Experimental data for the sample mass of 89.30 mg; □, Experimental data for the sample mass of 54 mg; with 0.05 J.K<sup>-1</sup>.g<sup>-1</sup> experimental error

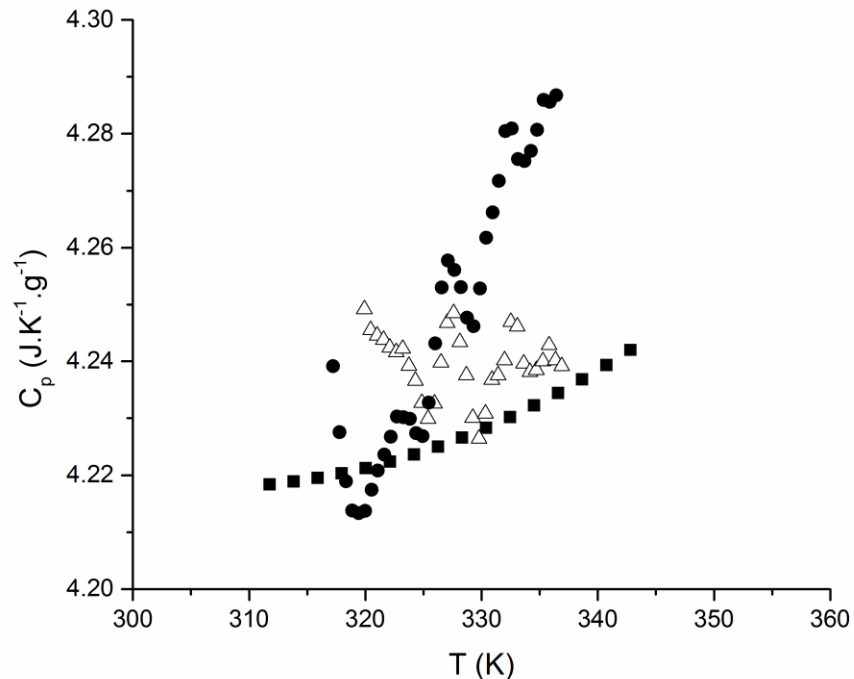


Figure 4-3: Water isobaric liquid heat capacity: ■, data from literature [46]; ●, experimental data for the sample mass 89.30 mg with inert gas; △, experimental data for the sample mass 89.30 mg without inert gas; with  $0.05 \text{ J.K}^{-1}.\text{g}^{-1}$  experimental error

### 4.3 Sample Preparation

Four liquid mixtures were prepared: n-alkanes only (Mixture 1), a mixture of aromatic and n-alkane constituents (Mixture 2), a mixture of naphthenic and n-alkane constituents (Mixture 3), and a mixture of naphthenic and aromatic constituents (Mixture 4). The compositions of these mixtures are listed in Table 4-1. SARTORIUS CP225D balance with an accuracy of 0.01 mg was utilized to prepare the samples. Some properties of the mixtures are listed in Table 4-2. Table shows the mass and the temperature range of DSC experiments for each studied sample.

Table 4-1: The composition of model hydrocarbon mixtures

Mixture 1			Mixture 2			Mixture 3			Mixture 4		
Composition	Weight <sup>1</sup>	Purity	Composition	Weight	Purity	Composition	Weight	Purity	Composition	Weight	Purity
Nonane	2.430	99%	1,2,4-TMB <sup>2</sup>	2.646	98%	Trans-decalin	2.602	99%	Trans-decalin	3.498	99%
Decane	2.431	99%	Decane	2.615	99%	Decane	2.617	99%	Durene	1.749	98.5%
Undecane	2.434	99%	Undecane	2.610	99%	Undecane	2.609	99%	1,2,4-TMB	3.499	98%

1- Units are in gram

2- 1,2,4-Trimethylbenzene

Table 4-2: Some properties of the model mixtures

Sample	Density <sup>1</sup> at 15 °C kg.m <sup>-3</sup>	T <sub>b</sub> <sup>1</sup> K	MW	Similarity variable
Mixture 1	734.05	440-445	141.35	0.2252
Mixture 2	775.31	450-455	137.95	0.1874
Mixture 3	779.64	455-460	145.21	0.2173
Mixture 4	880.20	450-455	129.67	0.1868

1. Values obtained by simulating the mixtures using the mixing rulls method in VMGSim[46].

Table 4-3: The temperature range and the mass of sample for each mixture used in DSC experiments

Sample	Sample mass	Temperatures range
	mg	°C
Mixture 1	71.40	20 – 150
Mixture 2	73.39	20 – 150
Mixture 3	70.97	20 - 160
Mixture 4	79.31	20 - 160

#### 4.4 Available Methods to Predict Similarity Variable

Both the liquid and ideal gas element base correlations [7, 10] are simple and predictive, and, hence, suitable for inclusion in process simulators. Implementation was expected to be straightforward. For compounds or mixtures comprising constituents defined on a molecular basis, elemental compositions of streams are readily calculated. For mixtures defined on other bases, the API method [15] described in Section 2.3 can be applied to obtain elemental composition, or elemental analysis can be conducted experimentally and included in the input data set. In practice, Virtual Material Group took an approach based on the API method to estimate similarity variable ( $\alpha$ ) [47]. The experimental  $\alpha$  data for the broad range of compounds comprising n-alkenes, n-alkynes, naphthenics, aromatics, and C<sub>10</sub>H<sub>10</sub> isomers, shown in Table 4-4, are compared to the API-based calculated alpha in this work, and the results are shown in Figure 4-4. It is obvious that the deviation of the API method is large and positive leading to Cp values predicted by the elemental based correlations to be overestimated.

Table 4-4 : The compounds used for calculating the API method deviation in  $\alpha$  estimation

<b>Compound</b>	<b>Chemical Formula</b>	<b>Density(kg.m<sup>-3</sup>)</b>	<b>MW</b>	<b>Actual <math>\alpha</math> (mol.g<sup>-1</sup>)</b>	<b>API method <math>\alpha</math>(mol.g<sup>-1</sup>)</b>
1-heptene	C <sub>7</sub> H <sub>14</sub>	701.3	98.187	0.2141	0.2300
1-octene	C <sub>8</sub> H <sub>16</sub>	719.1	112.2	0.2141	0.2280
1-nonene	C <sub>9</sub> H <sub>18</sub>	733.3	126.2	0.2141	0.2264
1-decene	C <sub>10</sub> H <sub>20</sub>	744.4	140.3	0.2141	0.2252
1-undecene	C <sub>11</sub> H <sub>22</sub>	753.7	154.3	0.2141	0.2242
1-dodecene	C <sub>12</sub> H <sub>24</sub>	762.5	168.3	0.2141	0.2234
1-tridecene	C <sub>13</sub> H <sub>26</sub>	769.4	182.3	0.2141	0.2227
1-tetradecene	C <sub>14</sub> H <sub>28</sub>	774.4	196.4	0.2141	0.2220
1-pentadecene	C <sub>15</sub> H <sub>30</sub>	779.7	210.4	0.2141	0.2215
1-ethylcyclopentene	C <sub>7</sub> H <sub>12</sub>	802.2	96.17	0.1977	0.2300
1,2-dimethylcyclohexene	C <sub>8</sub> H <sub>14</sub>	829.1	110.2	0.1998	0.2280
1-butylcyclopentene	C <sub>9</sub> H <sub>16</sub>	811.1	124.2	0.2014	0.2264
1-butylcyclohexene	C <sub>10</sub> H <sub>18</sub>	828.2	138.2	0.2027	0.2252
1-hexylcyclopentene	C <sub>11</sub> H <sub>20</sub>	819.5	152.3	0.2037	0.2242

Compound	Chemical Formula	Density(kg.m <sup>-3</sup> )	MW	Actual $\alpha$ (mol.g <sup>-1</sup> )	API method $\alpha$ (mol.g <sup>-1</sup> )
1-heptylcyclopentene	C <sub>12</sub> H <sub>22</sub>	822.8	166.3	0.2046	0.2234
1-heptylcyclohexene	C <sub>13</sub> H <sub>24</sub>	884.7	180.3	0.2054	0.1970
1-nonylcyclopentene	C <sub>14</sub> H <sub>26</sub>	827.7	194.4	0.2060	0.2221
1-decylcyclopentene	C <sub>15</sub> H <sub>28</sub>	829.7	208.4	0.2065	0.2215
1-heptyne	C <sub>7</sub> H <sub>12</sub>	736.2	96.17	0.1977	0.2299
1-octyne	C <sub>8</sub> H <sub>14</sub>	750.9	110.2	0.1998	0.2280
1-nonyne	C <sub>9</sub> H <sub>16</sub>	759.9	124.2	0.2014	0.2264
1-decyne	C <sub>10</sub> H <sub>18</sub>	768.8	138.2	0.2027	0.2252
1-undecyne	C <sub>11</sub> H <sub>20</sub>	775.9	152.3	0.2037	0.2242
1-dodecyne	C <sub>12</sub> H <sub>22</sub>	781.9	166.3	0.2046	0.2233
1-tridecyne	C <sub>13</sub> H <sub>24</sub>	787.6	180.3	0.2054	0.2227
1-tetradecyne	C <sub>14</sub> H <sub>26</sub>	793.8	194.4	0.2060	0.2221
1-pentadecyne	C <sub>15</sub> H <sub>28</sub>	795.9	208.4	0.2065	0.2215
ethylcyclopentane	C <sub>7</sub> H <sub>14</sub>	770.9	98.19	0.2141	0.2300
ethylcyclohexane	C <sub>8</sub> H <sub>16</sub>	791.8	112.2	0.2141	0.2280
butylcyclopentane	C <sub>9</sub> H <sub>18</sub>	788.2	126.2	0.2141	0.2264
cyclodecane	C <sub>10</sub> H <sub>20</sub>	860.4	140.3	0.2141	0.2022
cycloundecane	C <sub>11</sub> H <sub>22</sub>	865.3	154.3	0.2141	0.2013
cyclododecane	C <sub>12</sub> H <sub>24</sub>	866.5	168.3	0.2141	0.2013
cyclotridecane	C <sub>13</sub> H <sub>26</sub>	864.8	182.3	0.2141	0.2013
cyclotetradecane	C <sub>14</sub> H <sub>28</sub>	863.6	196.4	0.2141	0.2013
cyclopentadecane	C <sub>15</sub> H <sub>30</sub>	870.0	210.4	0.2141	0.2003
cyclodecane	C <sub>10</sub> H <sub>20</sub>	860.4	140.3	0.2141	0.2252
1-cyclopentylpentane	C <sub>10</sub> H <sub>20</sub>	794.8	140.3	0.2141	0.2252
2-methyl-2-nonene	C <sub>10</sub> H <sub>20</sub>	748.5	140.3	0.2141	0.2252
isobutylcyclohexane	C <sub>10</sub> H <sub>20</sub>	798.8	140.3	0.2141	0.2252
cis-1,2-diethylcyclohexane	C <sub>10</sub> H <sub>20</sub>	814.8	140.3	0.2141	0.2252
2-octene,2,6-dimethyl	C <sub>10</sub> H <sub>20</sub>	756.3	140.3	0.2141	0.2252
ethylcyclooctane	C <sub>10</sub> H <sub>20</sub>	841.6	140.3	0.2141	0.2252
2,3,4,4-tetramethyl-1-hexene	C <sub>10</sub> H <sub>20</sub>	800.4	140.3	0.2141	0.2252

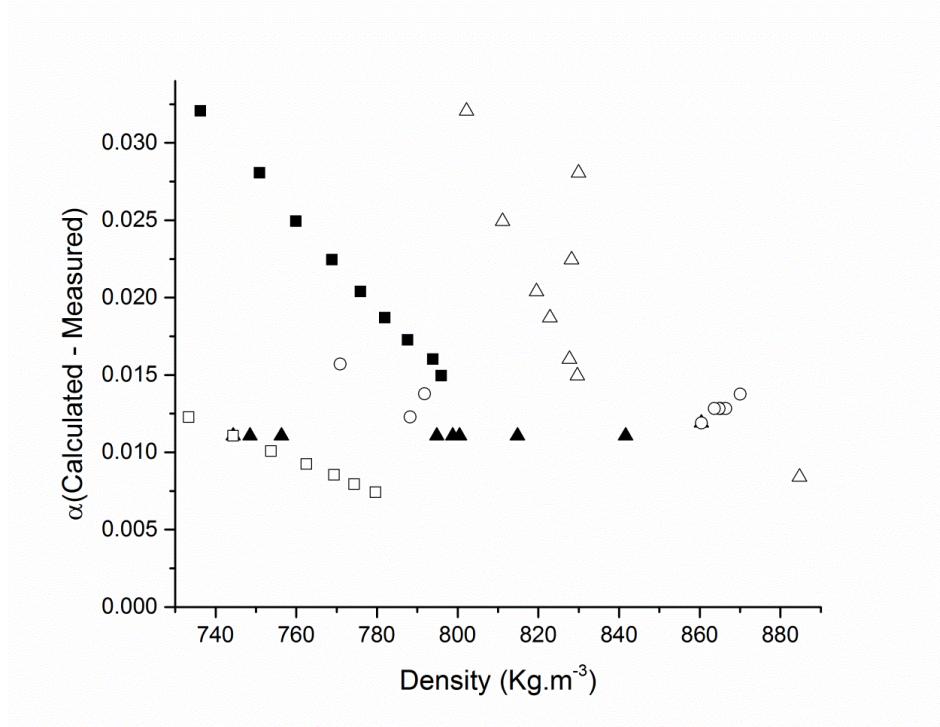


Figure 4-4: The deviation of the API procedure in estimating  $\alpha$  for different families: ■ , n-alkyne; ▲ , C<sub>10</sub>H<sub>20</sub> isomers; Δ , Aromatic; □ , n-alkene; ○ , Naphthenic

#### 4.5 Experimental Heat Capacity for the Model Mixtures

The data from the second type of calibration with sapphire described in Section 3.3 were used to calculate the offset values using “fminunc” syntax in Matlab R2012 [80] to obtain the optimized heat capacity value. Fminunc finds a minimum of a problem specified by  $\min_x f(x)$ .

Equation 4-1 is the optimization equation which should be solved to find the offset value.

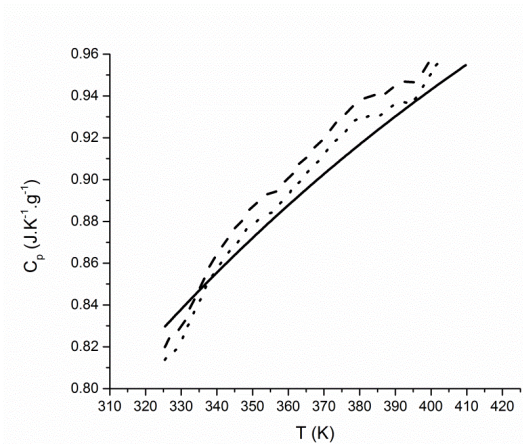
$$f_x(x) = \sum_{i=1}^N [y_{Real}(i) - y_{Exp}(i) + x] \quad 4-1$$

where  $y_{Real}$  is the actual value,  $y_{Exp}$  is the experimental value,  $i$  is the number of data points, and  $x$  is the offset value. By using “fminunc”, the  $x$  value can be calculated which will be further used to optimized the experimental data obtained for heat capacity of the hydrocarbon mixtures.

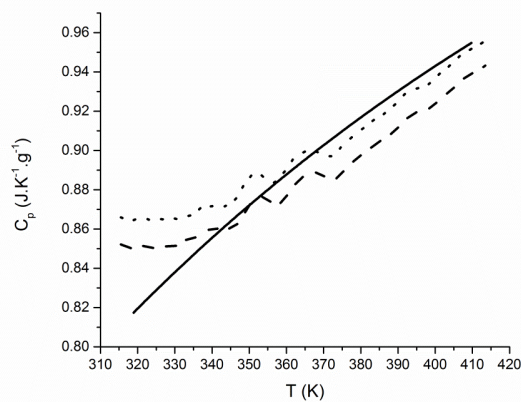
The calibration results for each mixture are shown in

(a)

(b)



(c)



(d)

Figure . Corrected experimental data for heat capacity of each studied mixture are illustrated in Figure 4-6 to 4-9 along with ideal mixture heat capacity calculated by Equation 2-2 utilizing compound's liquid heat capacity obtained from literature [82]. Consequently, since there is a systematic error with DSC measurement around  $\pm 0.05 \text{ J.K}^{-1}.\text{g}^{-1}$ , the experimental data for the mixture has high compatibility with ideal mixture heat capacity in which the liquid  $C_p$  for the single component obtained from literature. In the Figure 4-9 the data for the ideal mixture obtained from the literature is illustrated above 350 K, since under this temperature, durene is solid.



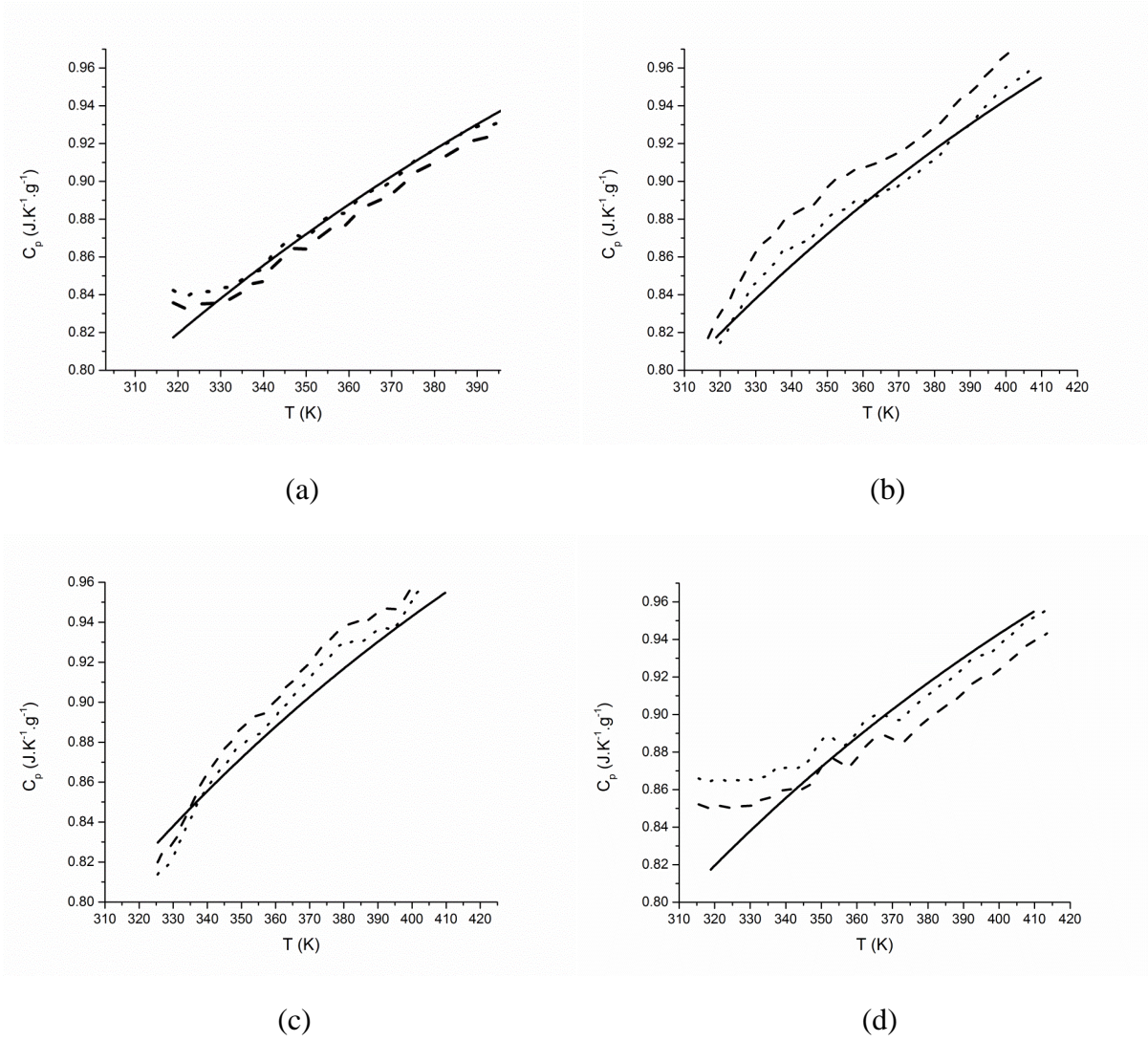


Figure 4-5: Sapphire heat capacity obtained from: — , literature data using the Archer equation; - - , Experimental data without offset ; . . . , Optimized data using offset for a) Mixture 1, b) Mixture 2, c) Mixture 3, and d) Mixture 4.

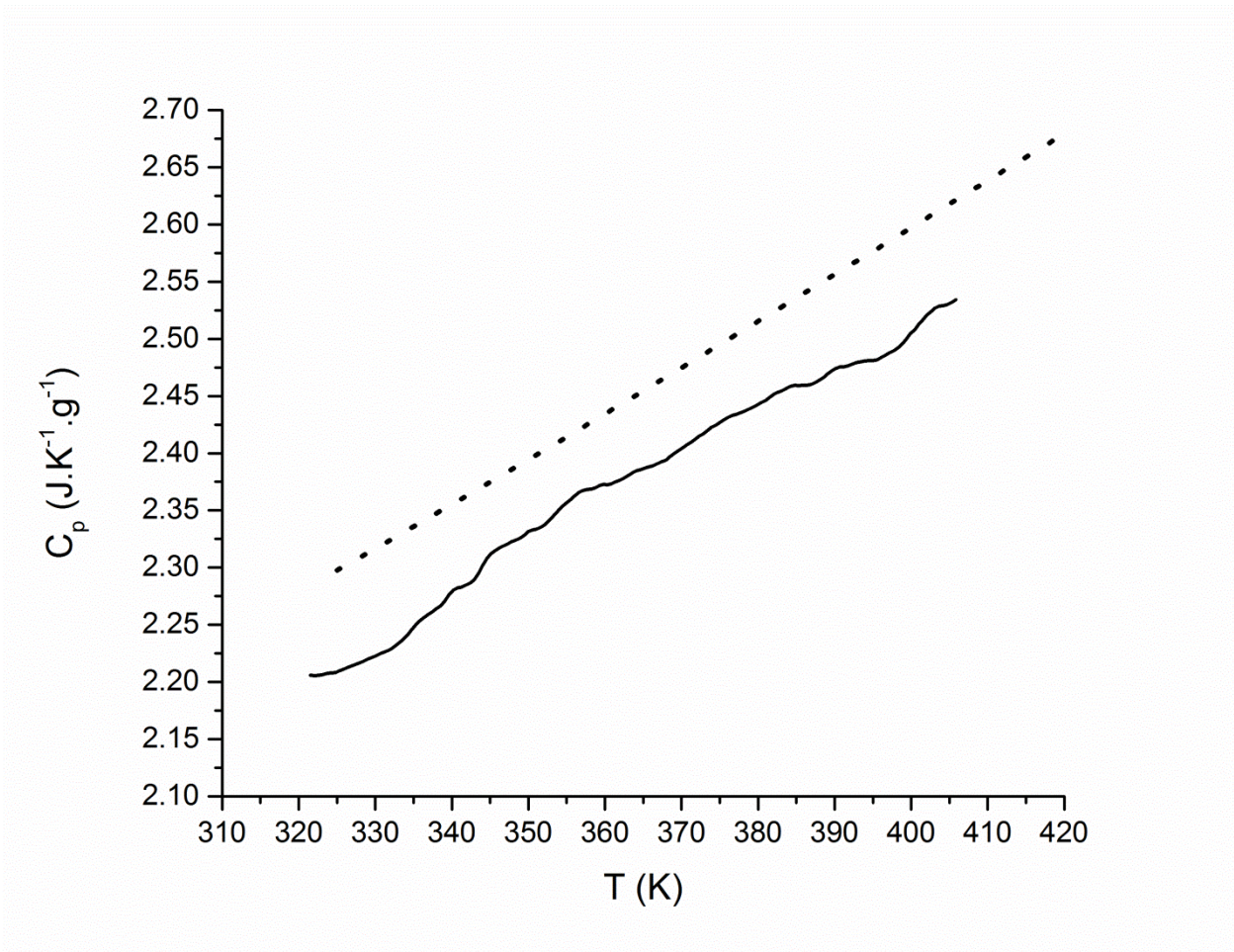


Figure 4-6: Liquid heat capacity for Mixture 1: — , ..., Literature data calculated for ideal mixture (component liquid  $C_p$  obtained from NIST [82])

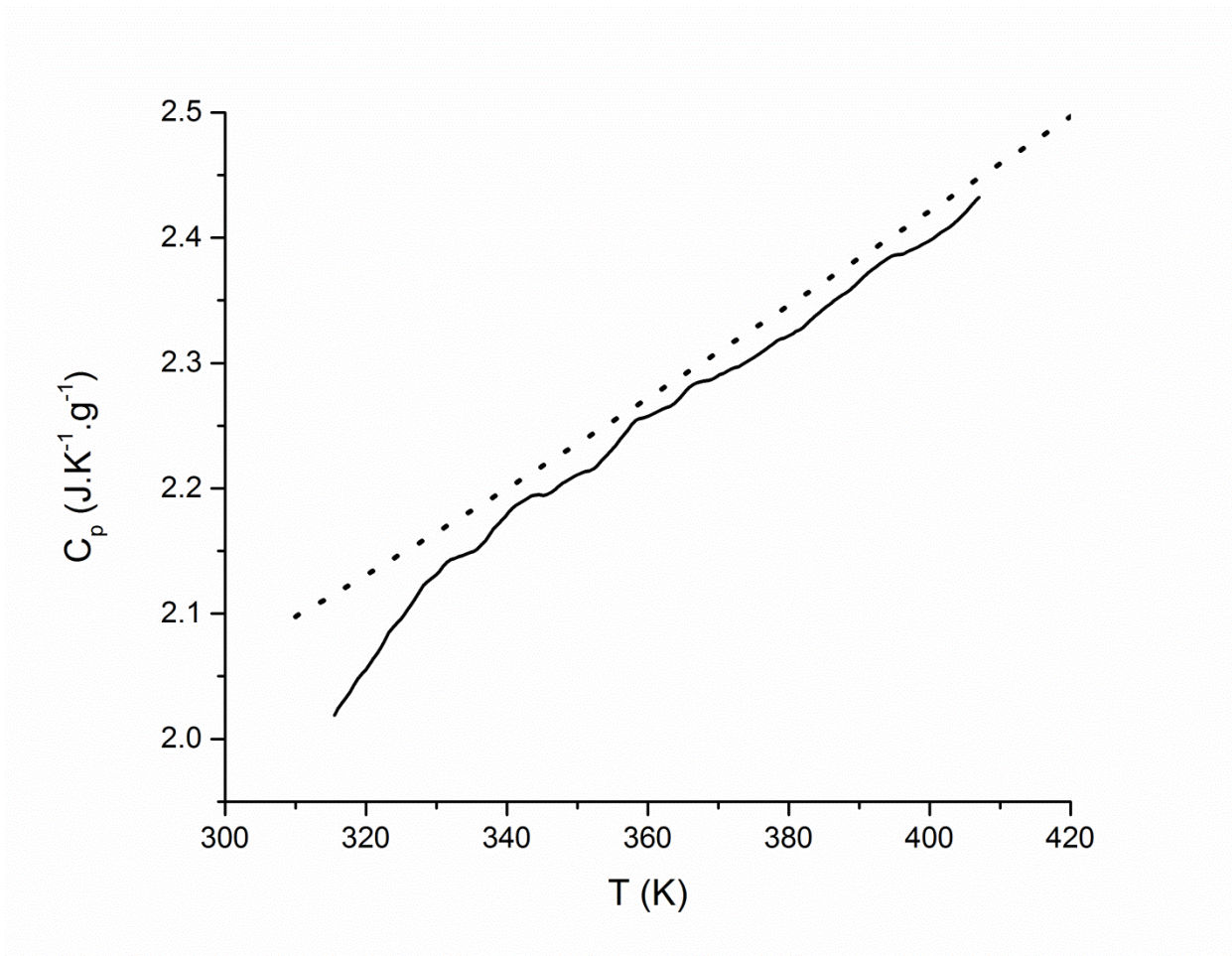


Figure 4-7: Liquid heat capacity for Mixture 2: — , ..., Literature data calculated for ideal mixture (component liquid  $C_p$  obtained from NIST [82])

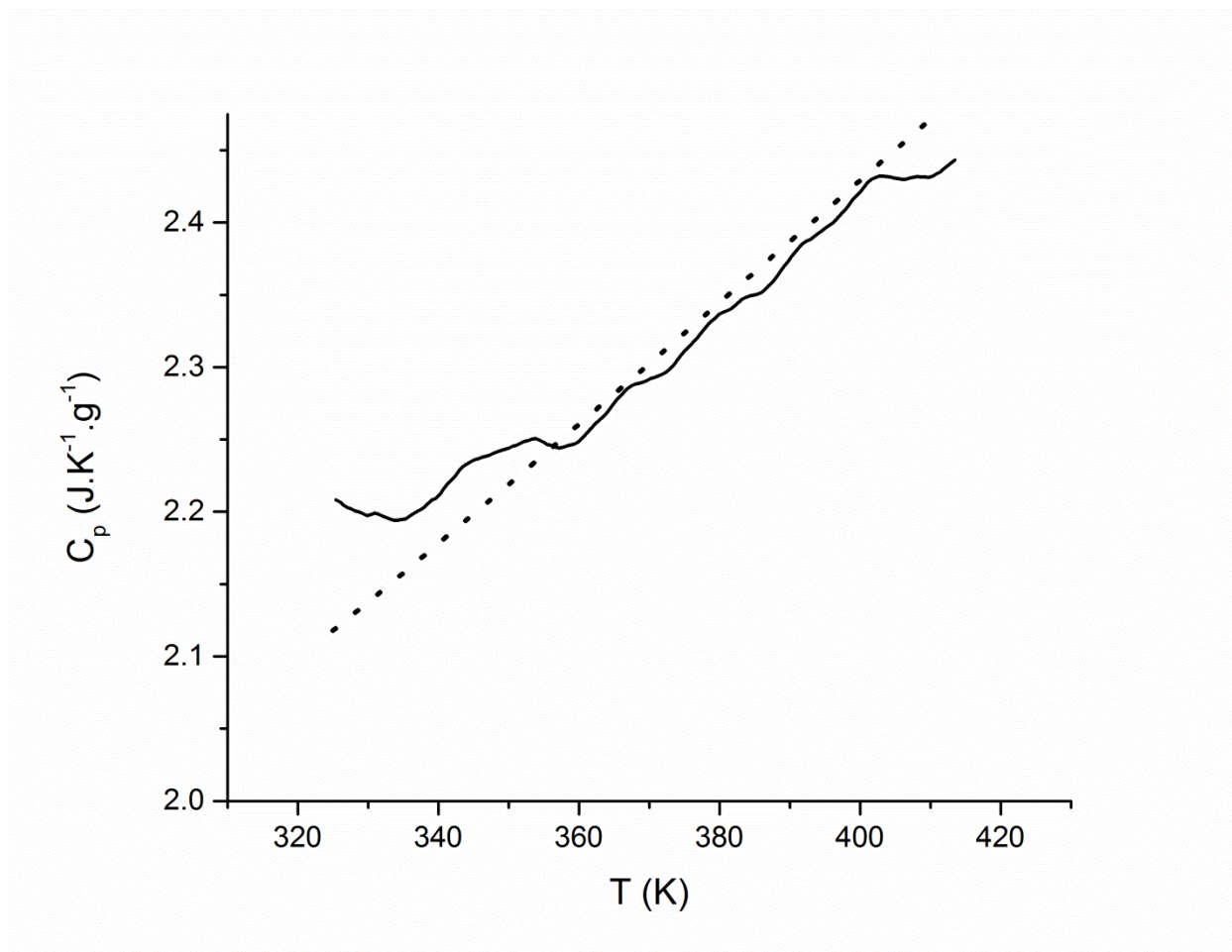


Figure 4-8: Liquid heat capacity for Mixture 3: — , Optimized experimental data..., Literature data calculated for ideal mixture (component liquid Cp obtained from NIST [82])

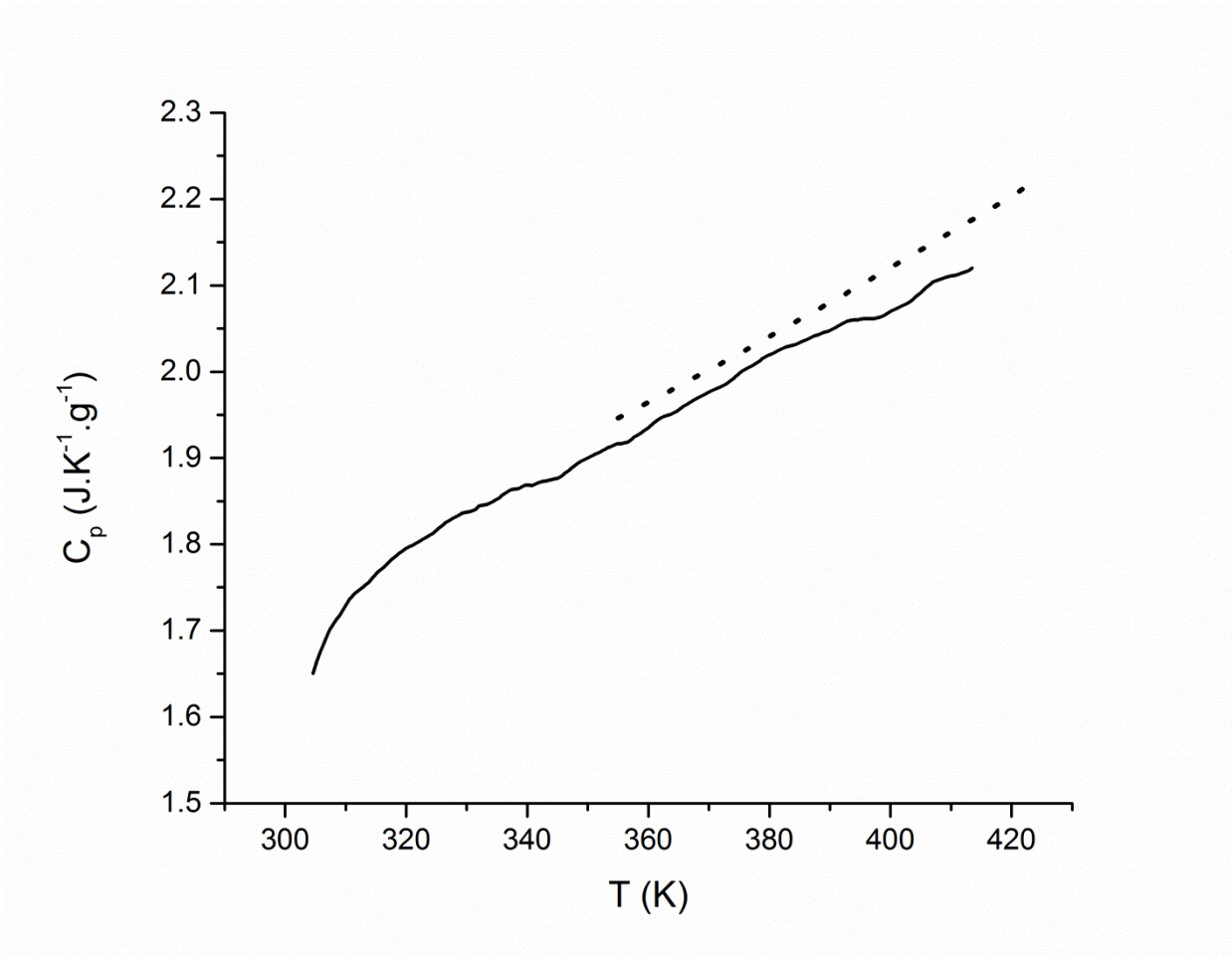


Figure 4-9: Liquid heat capacity for Mixture 4: — , ..., Literature data calculated for ideal mixture (component liquid Cp values were obtained from NIST [82])

#### 4.6 Comparisons Among Available Methods for Predicting Isobaric Liquid Heat Capacity

The diversity of methods to calculate isobaric liquid heat capacity is challenging as shown in Chapter 2. To identify the hierarchies and the best niches for the combination of these methods, a comparison with experimental data has been done for the prepared mixtures.

Hydrocarbon fluids can be characterized on a molecular or refinery basis, so-called “known compound” and “petroleum cut” terms in simulator software, i.e. VMGSim. “Known compounds” are those known in terms of thermophysical properties, i.e critical properties and structures. In this work their ideal gas heat capacity value is predicted by means of a group contribution method [32]. “Petroleum cuts” are those which are not known and should be defined by their boiling temperature range, density, or other easy-to-measure properties and their ideal gas heat capacity value is estimated by Lee-Kesler [6] correlation.

With these two methods of composition identification, the element based correlations described in Chapter 2.2.4 add two direct calculation options and four indirect calculation options for liquid phase  $C_p$  calculation per equation of state. Each of these approaches along with the conventional approaches for predicting liquid heat capacity described in Chapter 2 has advantages and disadvantages, and possesses different input data requirements as illustrated in Figure 4-10, where D-S is the Dadgostar-Shaw correlation [7], L-S IG  $C_p$  is the Lastovka-Shaw ideal gas heat capacity correlation [10], IG is the ideal gas, and EOS is any equation of state.

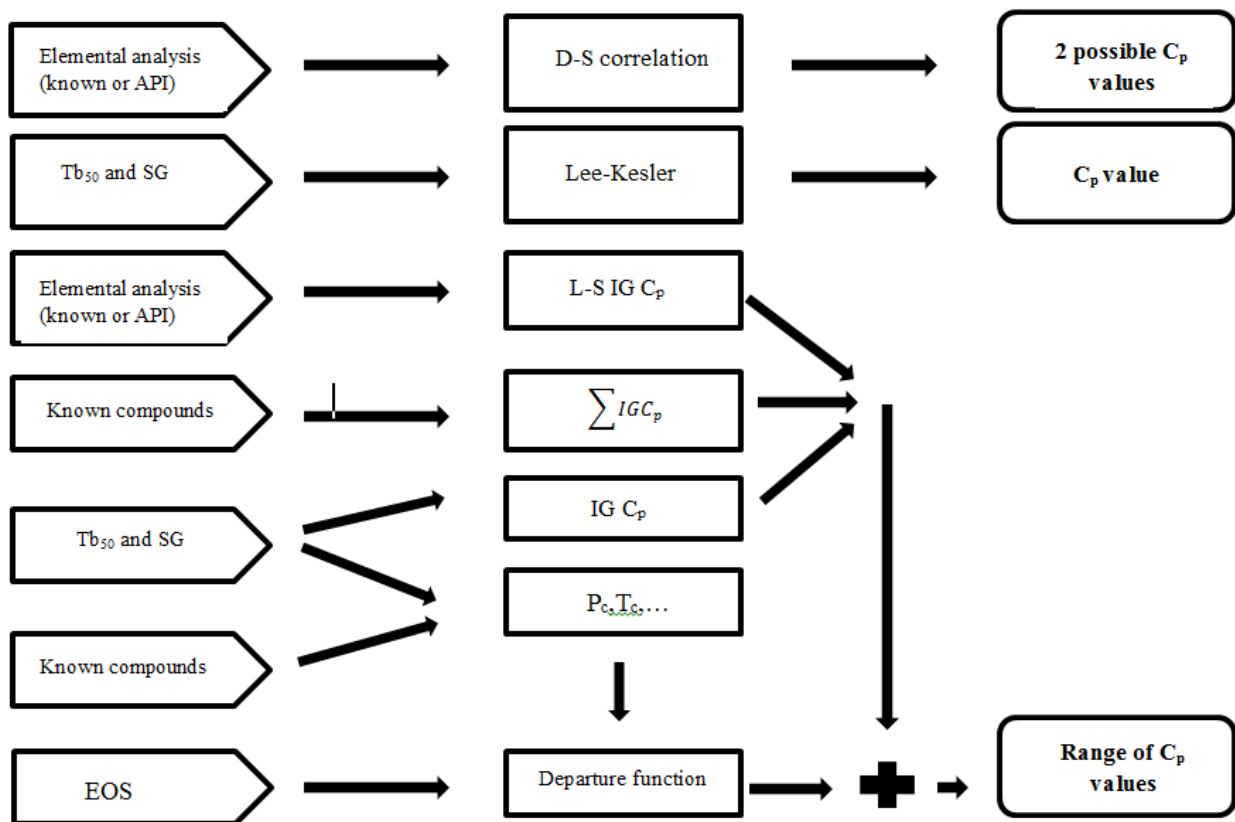


Figure 4-10: Computational matrix for the possible direct and indirect approaches for calculating isobaric liquid heat capacity of liquids

The widely used methods chosen to be compared are described as follows:

1)  $IGC_p$  +APR based departure function (known compounds)

According to the corresponding state theory [25], the difference between ideal gas heat capacity and liquid heat capacity can be calculated by a departure function based on an equation of state (EOS). In this work, ideal gas heat capacity ( $IGC_p$ ) is estimated by means of the group contribution method [32] developed for structurally known compounds. The Advanced Peng-

Robinson (APR), Equation 4-2, is chosen to derive the departure function defined in Equation 4-5.

$$P = \frac{RT}{V-b} - \frac{a \cdot \alpha}{V(V+b) + b(V-b)} \quad 4-2$$

$$b = 0.07780RT_c/P_c \quad 4-3$$

$$a = 0.45724 * R^2 * T_c^2/P_c \quad 4-4$$

$$\alpha = ((1 + \kappa(1 - T_r)^{0.5}))^2 \quad 4-4$$

$$\frac{C_p}{R} = \frac{C_p^\circ}{R} - 1 - T * \int_V^\infty \left[ \left( \frac{\partial^2 P}{\partial T^2} \right)_V \right] * dV - T * \left( \frac{\partial P}{\partial T} \right)_V^2 / \left( \frac{\partial P}{\partial V} \right)_T \quad 4-5$$

where R is the universal gas constant,  $\kappa$  is a term related to the acentric factor and  $T_c$  and  $P_c$  are the critical temperature and pressure respectively.

## 2) IGC<sub>p</sub> + APR based departure function (Petroleum cut)

As it is mentioned above, petroleum cut refers to those compounds which are not known in terms of either structure or critical properties. The correlation used to estimate the ideal gas heat capacity is based on the method developed by Lee-Kesler [28]. The sample is considered as an unknown compound, and the critical properties used to calculate the APR departure function are also estimated from the correlations.

## 3) L-S + APR based departure function (Known compound and actual $\alpha$ )



In this method, the  $IGC_p$  is estimated by Lastovka-Shaw correlation (L-S) [10] and the sample is treated as a known compound, similarly to method 1.

4) L-S + APR based departure function (Petroleum cut)

It is the same as method 3, but the sample is treated as an unknown compound similarly to method 2.

5) D-S (Actual  $\alpha$ )

In this method, the heat capacity of a sample is estimated by Dadgostar-Shaw (D-S) correlation [7] and the actual value is used for  $\alpha$ .

6) D-S (API base  $\alpha$ )

This is the same as method 5, but the  $\alpha$  is estimated by the API approach [15].

7) L-S +APR based departure function (Known compound and API base  $\alpha$ )

This is the same as method 3, but the  $\alpha$  is estimated by the API approach [15].

8) L-K  $C_p$

In this method, the liquid heat capacity is predicted by Lee-Kesler (L-K) correlation [6].

The experimental and computed results are reported in Figures 4-11 to 4-14 for all the prepared samples. The mean absolute percentage error (MAPE) between the experimental and computed values is shown in Table 4-5.

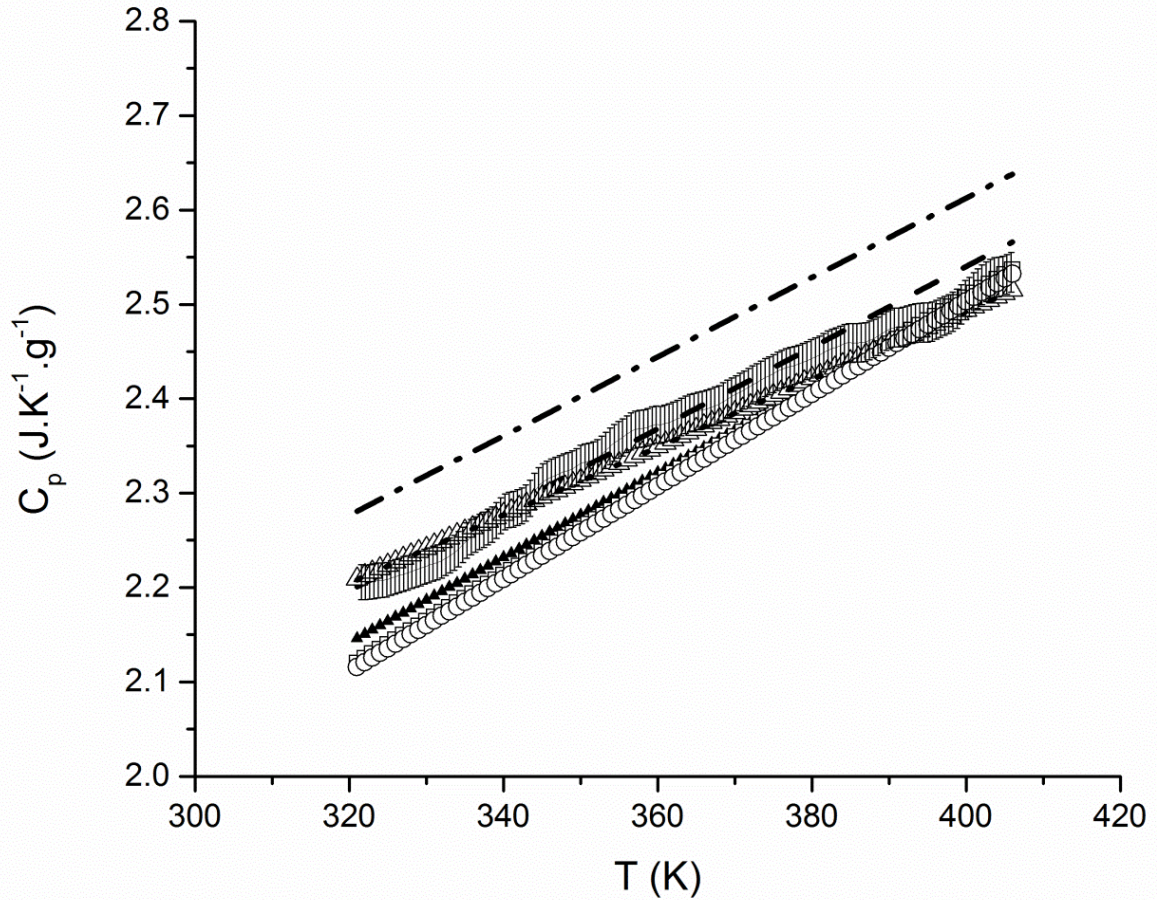


Figure 4-11: Isobaric liquid heat capacity of Mixture 1 calculated by various methods: —, Experimental data with the error of  $0.02 \text{ J.K}^{-1}.\text{g}^{-1}$  shown as a shaded area; - - ,  $\text{IGC}_p+\text{APR}$  base departure function (known compound);  $\blacktriangle$  ,  $\text{IGC}_p+\text{APR}$  base departure function (petroleum cut);  $\square$  , Lastovka-Shaw  $\text{IGC}_p+\text{APR}$  base departure function (known compound);  $\text{—}$  , Lastovka-Shaw  $\text{IGC}_p+\text{APR}$  base departure function (petroleum cut);  $\Delta$  , Dadgostar-Shaw  $C_p$  (Actual  $\alpha$ ); . . . , Dadgostar-Shaw  $C_p$  (API base  $\alpha$ );  $\circ$  , Lastovka-Shaw  $\text{IGC}_p$  (API base  $\alpha$ )+APR base departure function; - . - , Lee-Kesler  $C_p$

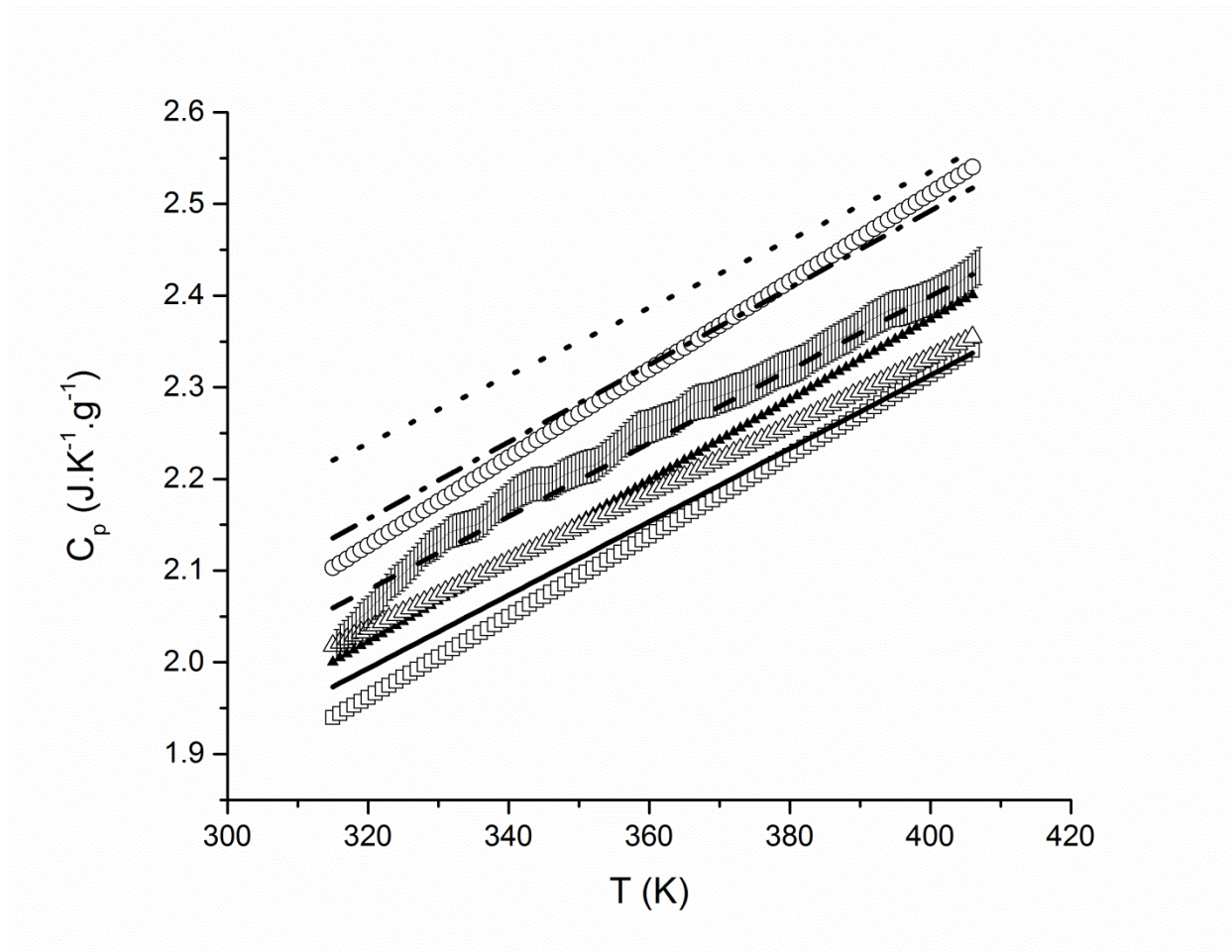


Figure 4-12: Isobaric liquid heat capacity of Mixture 2 calculated by various methods: —, Experimental data with the error of  $0.02 \text{ J.K}^{-1}.\text{g}^{-1}$  shown as a shaded area; - - , IGC<sub>p</sub>+APR base departure function (known compound); ▲ , IGC<sub>p</sub>+APR base departure function (petroleum cut); □, Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (known compound); —, Lastovka-Shaw IGC<sub>p</sub>+APR base departure function (petroleum cut); Δ, Dadgostar-Shaw C<sub>p</sub> (Actual  $\alpha$ ); . . . , Dadgostar\_Shaw C<sub>p</sub> (API base  $\alpha$ ); ○ , Lastovka-Shaw IGC<sub>p</sub> (API base  $\alpha$ ) +APR base departure function; \_ . \_ , Lee-Kesler C<sub>p</sub>

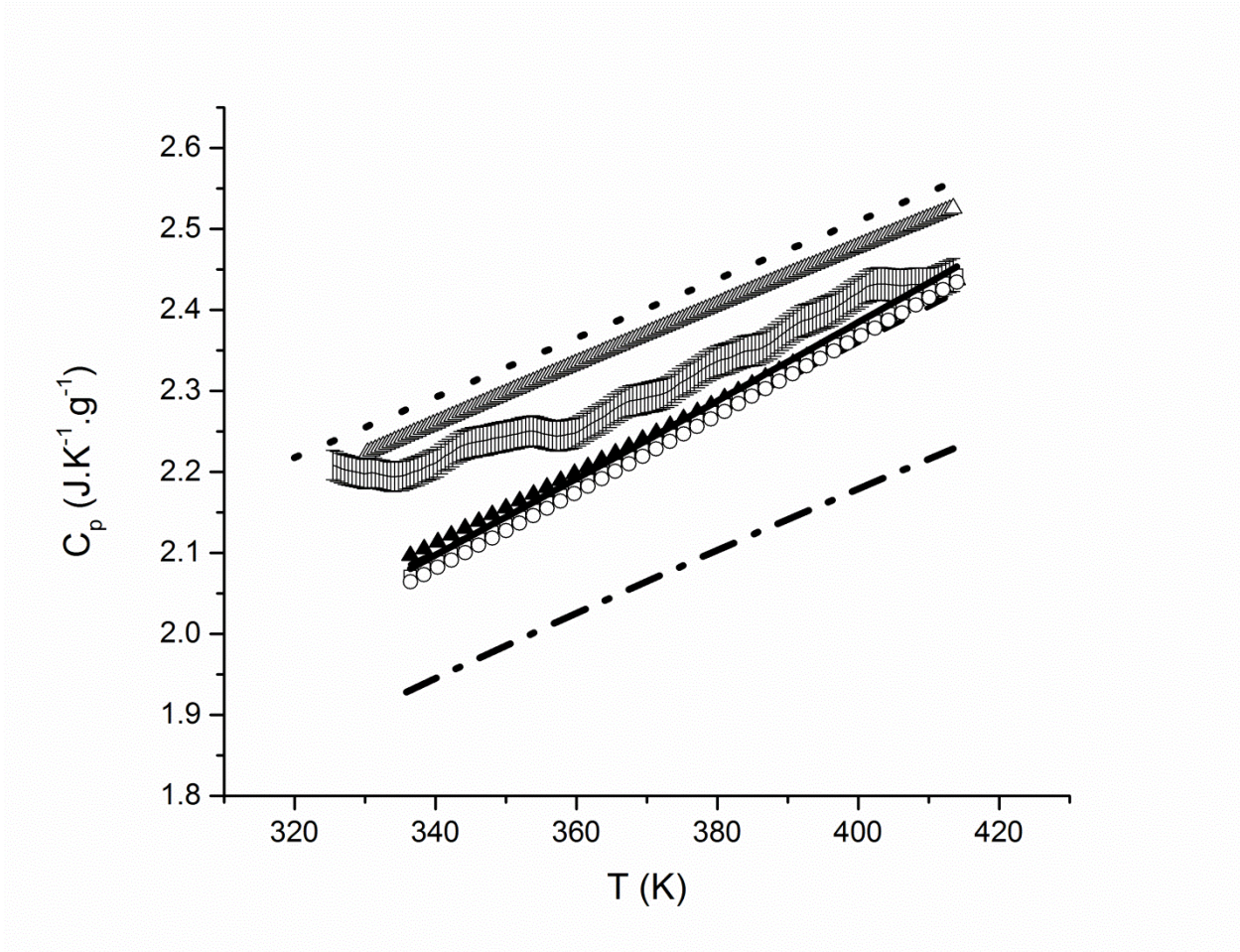


Figure 4-13: Isobaric liquid heat capacity of Mixture 3 calculated by various methods: —, Experimental data with the error of  $0.02 \text{ J.K}^{-1}.\text{g}^{-1}$  shown as a shaded area; - - ,  $\text{IGC}_p+\text{APR}$  base departure function (known compound);  $\blacktriangle$  ,  $\text{IGC}_p+\text{APR}$  base departure function (petroleum cut);  $\square$  , Lastovka-Shaw  $\text{IGC}_p+\text{APR}$  base departure function (known compound); —, Lastovka-Shaw  $\text{IGC}_p+\text{APR}$  base departure function (petroleum cut);  $\Delta$  , Dadgostar-Shaw  $C_p$  (Actual  $\alpha$ ); . . . , Dadgostar\_Shaw  $C_p$  (API base alpha);  $\circ$  , Lastovka-Shaw  $\text{IGC}_p$  (API base  $\alpha$ ) +APR base departure function; - . - , Lee-Kesler  $C_p$

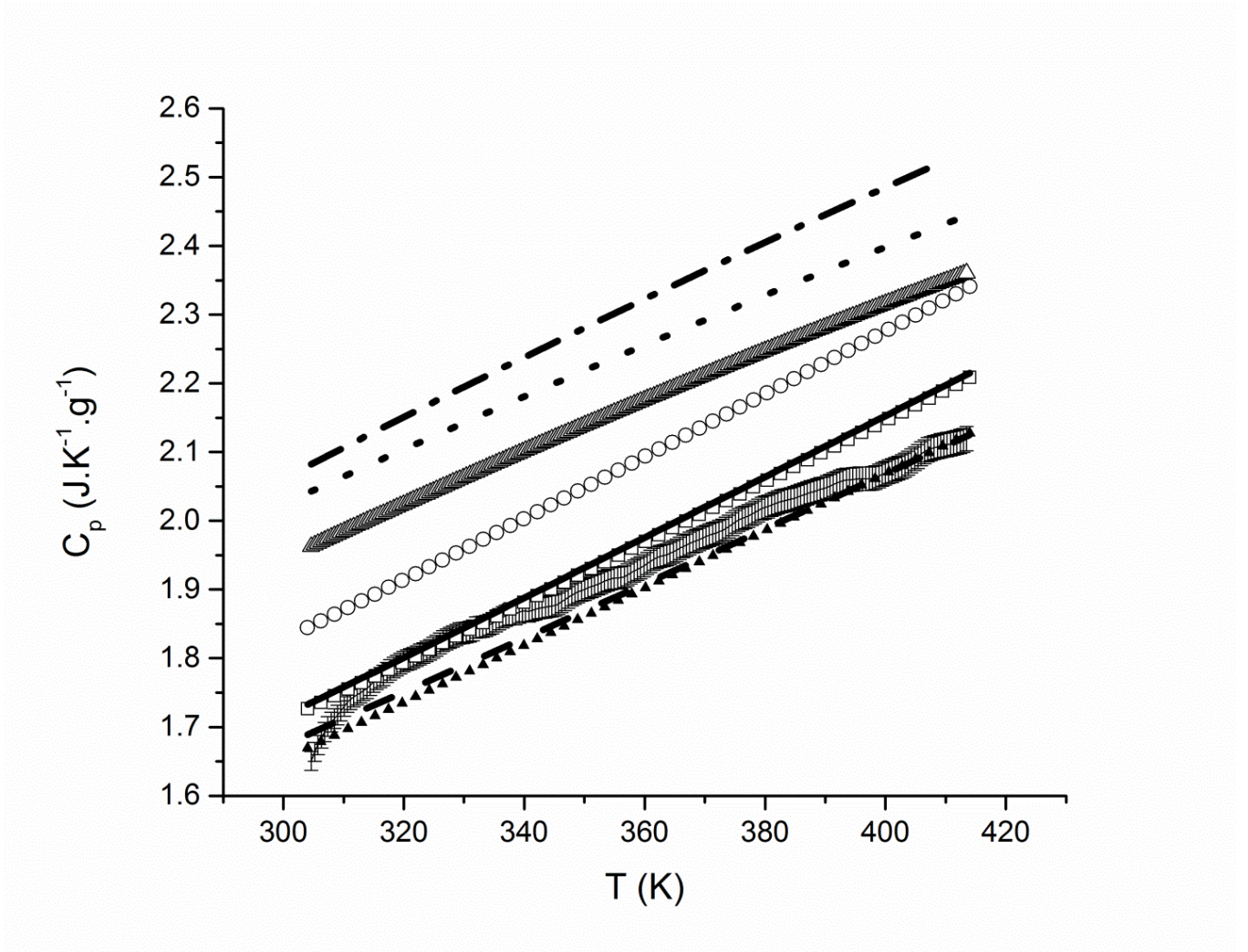


Figure 4-14: Isobaric liquid heat capacity of Mixture 4 calculated by various methods: —, Experimental data with the error of  $0.02 \text{ J.K}^{-1}.\text{g}^{-1}$  shown as a shaded area; - - ,  $\text{IGC}_p+\text{APR}$  base departure function (known compound);  $\blacktriangle$  ,  $\text{IGC}_p+\text{APR}$  base departure function (petroleum cut);  $\square$ , Lastovka-Shaw  $\text{IGC}_p+\text{APR}$  base departure function (known compound); —, Lastovka-Shaw  $\text{IGC}_p+\text{APR}$  base departure function (petroleum cut);  $\Delta$ , Dadgostar-Shaw  $C_p$  (Actual  $\alpha$ ); . . . , Dadgostar-Shaw  $C_p$  (API base  $\alpha$ );  $\circ$  , Lastovka-Shaw  $\text{IGC}_p$  (API base  $\alpha$ ) +APR base departure function; - . - , Lee-Kesler  $C_p$

Table 4-5: Deviation of liquid phase constant pressure heat capacity computational approaches from experimental data for mixtures 1-4

	MAPE							
	L-S IG C <sub>p</sub> (API based $\alpha$ )+APR based DF (known compounds)	D-S C <sub>p</sub> (API based $\alpha$ )	L-K C <sub>p</sub>	IG C <sub>p</sub> + APR based DF (known compounds)	L-S IG C <sub>p</sub> +APR based DF (petroleum cut)	L-S IG C <sub>p</sub> +APR based DF (known compounds)	IG C <sub>p</sub> + APR based DF (petroleum cut)	D-S C <sub>p</sub> (know n $\alpha$ )
Mixture 1	1.91	0.43	3.67	0.50	2.46	1.91	1.49	0.43
Mixture 2	8.57	6.31	3.35	0.49	1.25	0.76	2.07	2.69
Mixture 3	4.47	3.71	10.89	3.35	2.62	2.89	2.53	4.10
Mixture 4	7.86	16.89	20.1	0.92	2.22	2.24	1.27	13.24

## 4.7 Conclusion and Recommendation

The element base correlations add four indirect computational options for liquid phase C<sub>p</sub> calculation per equation of state, and two direct calculation options. Each of these computational variants has advantages and disadvantages and different input data requirements. Figure 4-11 to 4-14 confirms that selecting a wrong method for predicting liquid heat capacity can lead to 0.2 J.K<sup>-1</sup>g<sup>-1</sup> or higher deviations from the actual values, so, defining a potential application for each method is a necessity. For known compounds with known critical properties, estimated ideal gas heat capacity by correlation (group contribution method), and APR based departure functions shows the most precise and accurate result in predicting constant-pressure heat capacity for liquids as it is illustrated for four prepared mixtures; however, for ill-defined material the

Dadgostar-Shaw correlation has the highest accuracy and precision compared to existing correlations.

Estimated  $C_p$  by the Dadgostar-Shaw correlation is sensitive to similarity variable, since having a deviation of 0.001 in estimating similarity variable would cause a deviation of 0.004 ( $J.K^{-1}.g^{-1}$ ) in predicted  $C_p$  value. It is shown in Figure 4-4 that the deviation in estimated values of  $\alpha$  by the API procedure is large; consequently, element based heat capacity computational approaches must currently have the experimental elemental analysis as a required input and developing a precise and accurate correlation for estimating similarity variable is in the scope of the next chapter.

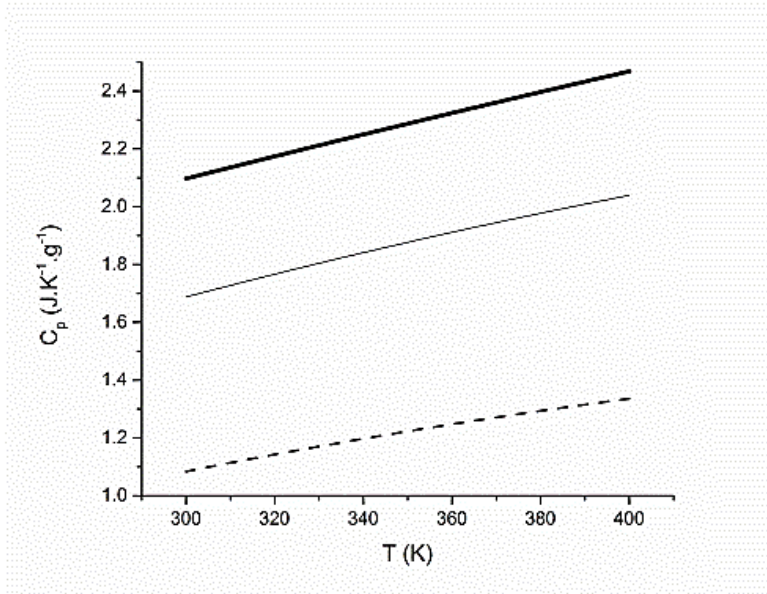
## 5 Development of a Predictive Correlation for the Composition Similarity Variable for Organic Compounds

### 5.1 Introduction

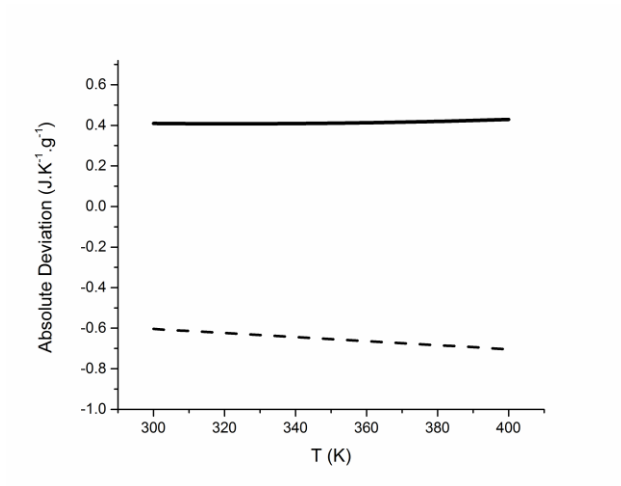
In the previous chapters it was mentioned that in order to implement the element based heat capacity correlations in commercial chemical engineering process simulator software, similarity variables must be either computed from experimental elemental analysis measurements, or estimated from available property data with little deviation. Elemental compositions of ill-defined hydrocarbons are frequently estimated using the API approach [15]. It was shown in Chapter 4 that this approach systematically overestimates the values of the similarity variable and hence overestimates heat capacity values irrespective of the phase state; the deviation can exceed 0.2 J/g/K or 20 % for representative cases.

The sensitivity of the Dadgostar-Shaw correlation to  $\alpha$  values is shown in Figure 5.1 (a-c) for the range of anticipated  $\alpha$  values. For large n-alkanes, with the empirical formula  $\text{CH}_2$ ,  $\alpha = 0.215 \text{ mol}\cdot\text{g}^{-1}$ , for large aromatic compounds, the empirical formula approaches C, and  $\alpha = 0.085 \text{ mol}\cdot\text{g}^{-1}$ . So, based on the heat capacity of midpoint in this range,  $\alpha = 0.15 \text{ mol}\cdot\text{g}^{-1}$ , relative and absolute deviations for  $C_p$  values at fixed temperature range from +25 to -35% and +0.4 to -0.6  $\text{J}\cdot\text{K}^{-1}\cdot\text{g}^{-1}$  respectively. This range approximates the maximum uncertainty for this correlation.

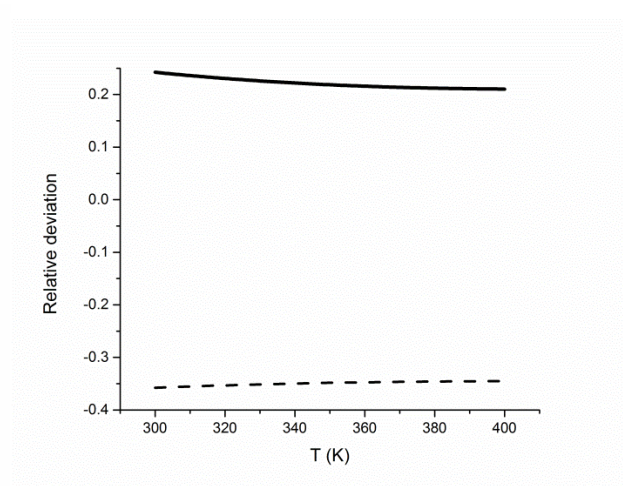




(a)



(b)



(c)

Figure 5-1: a)  $C_p$  predicted by the D-S correlation for: — ,  $\alpha=0.15 \text{ mol}\cdot\text{g}^{-1}$ ; - - -,  $\alpha=0.085 \text{ mol}\cdot\text{g}^{-1}$ ; — ,  $\alpha=0.215 \text{ mol}\cdot\text{g}^{-1}$ . Absolute (b) and relative (c) deviations from  $C_p$  values predicted using  $\alpha = 0.15 \text{ mol}\cdot\text{g}^{-1}$ : - - -, for  $\alpha=0.085 \text{ mol}\cdot\text{g}^{-1}$  and —  $\alpha=0.215 \text{ mol}\cdot\text{g}^{-1}$ .

The objective in this chapter is to reduce the uncertainty of heat capacity calculations for compounds or mixtures where the elemental analysis is not available, by correlating other known or calculated thermophysical properties to  $\alpha$ .

## **5.2 Correlation Development to Predict Similarity Variable**

### **5.2.1 Dependence of Similarity Variable on Physical Properties**

As a starting point for the development of a correlation for  $\alpha$ , as a function of physical properties, the relationships between  $\alpha$  and physical properties [density at 25 °C, molar mass and boiling temperature at 1.01 bar] are shown in Figure 5-2 (a-c) respectively for 154 organic liquid compounds, shown in Table 5-1, comprising n-alkane, n-alkene, n-alkyne, naphthenic, aromatic, and heteroatom containing compounds. Data are obtained from the NIST chemistry web-book [81]. Density, molar mass and boiling temperature are selected among other physical properties, since they are the most available characterization factors for the oil cuts and other ill-defined hydrocarbon compounds.

Table 5-1: Property database used for similarity variable correlation <sup>1</sup>

Compound	Formula	Data Base	Molar mass <sup>1</sup> , g/mol	$\alpha$ , mol·g <sup>-1</sup>	T <sub>b</sub> <sup>1</sup> at 1 atm / K	Density <sup>1</sup> at 25°C, kg·m <sup>-3</sup>	Family	Sample No.
Indene	C <sub>9</sub> H <sub>8</sub>	Training Set	116.2	0.1464	454.0	991.8	Aromatic	1
Indane	C <sub>9</sub> H <sub>10</sub>	Training Set	118.2	0.1609	450.0	958.2	Aromatic	2
1,2,3-Trimethylbenzene	C <sub>9</sub> H <sub>12</sub>	Training Set	120.2	0.1749	449.2	890.5	Aromatic	3
1,2,3,4-Tetramethylbenzene	C <sub>10</sub> H <sub>14</sub>	Training Set	134.2	0.1790	478.1	900.9	Aromatic	4
1-Methylnaphthalene	C <sub>11</sub> H <sub>10</sub>	Training Set	142.2	0.1478	515.0	1016.4	Aromatic	5
Ethylene, 1,1-diphenyl-	C <sub>14</sub> H <sub>12</sub>	Training Set	180.2	0.1443	543.7	1019.5	Aromatic	6
1,2-Dihydroanthracene	C <sub>14</sub> H <sub>12</sub>	Training Set	180.2	0.1443	606.0	1138	Aromatic	7
1,2,3,5-Tetraethylbenzene	C <sub>14</sub> H <sub>22</sub>	Training Set	190.3	0.1893	521.7	876.5	Aromatic	8
Cyclohexene, 1-octyl-	C <sub>14</sub> H <sub>26</sub>	Training Set	194.3	0.2060	530.5	838.6	Aromatic	9
1-ethylcyclopentene	C <sub>7</sub> H <sub>12</sub>	Training Set	96.17	0.1978	379.3	793.24	Aromatic	10
1,2-dimethylcyclohexene	C <sub>8</sub> H <sub>14</sub>	Training Set	110.2	0.1998	409.7	820.1	Aromatic	11
dihydro-1,6-dimethyl-4-(1-methylethyl)naphthalene	C <sub>15</sub> H <sub>20</sub>	Training Set	200.3	0.1749	562.0	936.8	Aromatic	12
cyclopentene, 4-butyl-	C <sub>9</sub> H <sub>16</sub>	Training Set	124.2	0.2015	427.0	836.0	Aromatic	13
Cyclohexene, 3-methyl-6-(1-methylethyl)-	C <sub>10</sub> H <sub>18</sub>	Training Set	138.2	0.2027	438.0	820.4	Aromatic	14
1-butylcyclohexene	C <sub>10</sub> H <sub>18</sub>	Training Set	138.2	0.2027	453.7	820.1	Aromatic	15
Benzene, (1-methylundecyl)-	C <sub>18</sub> H <sub>30</sub>	Training Set	246.4	0.1950	576.7	851.6	Aromatic	16
1-hexylcyclopentene	C <sub>11</sub> H <sub>20</sub>	Training Set	152.3	0.2038	478.0	808.4	Aromatic	17
Benzene, (1-pentylhexyl)-	C <sub>17</sub> H <sub>28</sub>	Training Set	232.4	0.1938	580.0	962.0	Aromatic	18
Benzene, m-bis(1-methylbutyl)	C <sub>16</sub> H <sub>26</sub>	Training Set	218.4	0.1925	553.0	945.0	Aromatic	19
n-decylbenzene	C <sub>16</sub> H <sub>26</sub>	Training Set	218.4	0.1925	571.0	852.1	Aromatic	20
Phenanthrene, 2-dodecyl-	C <sub>26</sub> H <sub>34</sub>	Training Set	346.6	0.1733	746.7	962.0	Aromatic	21
Benzene, (3-octylundecyl)-	C <sub>25</sub> H <sub>44</sub>	Training Set	344.6	0.2004	671.6	852.6	Aromatic	22
1H-Indene, 2-hexadecyl-2,3-dihydro-	C <sub>25</sub> H <sub>42</sub>	Training Set	342.6	0.1958	674.0	879.7	Aromatic	23
1,1-Diphenyldodecane	C <sub>24</sub> H <sub>34</sub>	Training Set	322.5	0.1800	672.0	924.6	Aromatic	24
Phenanthrene, 9-nonyl-	C <sub>23</sub> H <sub>28</sub>	Training Set	304.5	0.1676	708.0	1109	Aromatic	25
Pentadecane, 2-methyl-2-phenyl-	C <sub>22</sub> H <sub>38</sub>	Training Set	302.5	0.1985	638.0	858.4	Aromatic	26
Naphthalene, 2-butyl-3-hexyl-	C <sub>20</sub> H <sub>28</sub>	Training Set	268.4	0.1790	642.1	930.0	Aromatic	27
2-dodecyl-naphthalene	C <sub>22</sub> H <sub>32</sub>	Training Set	296.5	0.1823	706.0	912.4	Aromatic	28
Naphthalene, 1,2,3,4-tetrahydro-1-nonyl-	C <sub>19</sub> H <sub>30</sub>	Training Set	258.5	0.1898	633.0	991.0	Aromatic	29
1H-Indene, 2-butyl-1-hexyl-2,3-dihydro-	C <sub>19</sub> H <sub>30</sub>	Training Set	258.5	0.1898	610.7	893.0	Aromatic	30
1-heptylcyclopentene	C <sub>12</sub> H <sub>22</sub>	Training Set	166.3	0.2047	491.9	816.4	Aromatic	31
1-heptylcyclohexene	C <sub>13</sub> H <sub>24</sub>	Training Set	180.3	0.2054	507.0	875.0	Aromatic	32
1-decylcyclopentene	C <sub>15</sub> H <sub>28</sub>	Training Set	208.4	0.2066	536.0	825.9	Aromatic	33
1,2-diphenyl-1-butene	C <sub>16</sub> H <sub>16</sub>	Training Set	208.3	0.1537	590.0	1008	Aromatic	34
2,6-Diisopropyl-naphthalene	C <sub>16</sub> H <sub>20</sub>	Training Set	212.3	0.1697	597.0	1048	Aromatic	35

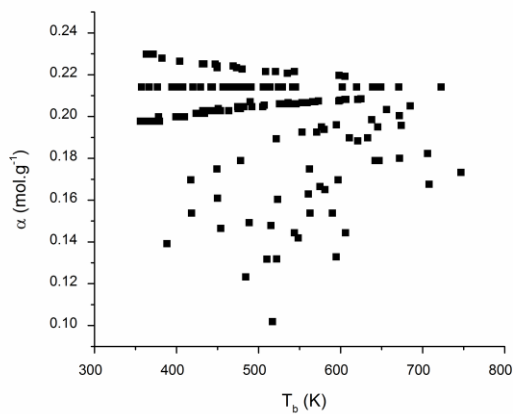
Compound	Formula	Data Base	Molar mass, g/mol	$\alpha$ , mol·g <sup>-1</sup>	T <sub>b</sub> at 1 atm / K	Density at 25°C, kg·m <sup>-3</sup>	Family	Sample No.
Benzene, 1,1'-pentyldienebis-	C <sub>17</sub> H <sub>20</sub>	Training Set	224.3	0.1650	597.0	1048	Aromatic	36
Naphthalene, 1,2,3,4-tetrahydro-1-octyl-	C <sub>18</sub> H <sub>28</sub>	Training Set	244.4	0.1884	580.9	962.7	Aromatic	37
Benzene, (1-hexylheptyl)-	C <sub>19</sub> H <sub>32</sub>	Training Set	260.5	0.1960	621.0	911.8	Aromatic	38
Naphthalene, 1,4-dimethyl-5-octyl-	C <sub>20</sub> H <sub>28</sub>	Training Set	268.4	0.1790	594.8	850.5	Aromatic	39
1,6-Heptadiene	C <sub>7</sub> H <sub>12</sub>	Training Set	96.17	0.1978	362.6	705.4	n-alkene	40
1,5-Hexadiene, 2-methyl-	C <sub>7</sub> H <sub>12</sub>	Training Set	96.17	0.1978	361.2	716.9	n-alkene	41
2,3-Pentadiene, 2,4-dimethyl-	C <sub>7</sub> H <sub>12</sub>	Training Set	110.2	0.1978	356.1	701.3	n-alkene	42
1,3-Pentadiene, 2,4-dimethyl-	C <sub>7</sub> H <sub>12</sub>	Training Set	96.17	0.1978	366.7	732.5	n-alkene	43
1-heptene	C <sub>7</sub> H <sub>14</sub>	Training Set	98.19	0.2141	366.8	692.8	n-alkene	44
1,4-Heptadiene, 3-methyl-	C <sub>8</sub> H <sub>14</sub>	Training Set	182.3	0.1998	378.2	725.1	n-alkene	45
2,4-Hexadiene, 2,5-dimethyl-	C <sub>8</sub> H <sub>14</sub>	Training Set	110.2	0.1998	408.4	757.8	n-alkene	46
1-octene	C <sub>8</sub> H <sub>16</sub>	Training Set	112.2	0.2141	394.4	710.9	n-alkene	47
1-nonene	C <sub>9</sub> H <sub>18</sub>	Training Set	126.2	0.2141	420.0	725.4	n-alkene	48
1,9-Decadiene	C <sub>10</sub> H <sub>18</sub>	Training Set	138.2	0.2027	437.6	749.8	n-alkene	49
1-decene	C <sub>10</sub> H <sub>20</sub>	Training Set	140.3	0.2141	443.7	737.0	n-alkene	50
1-undecene	C <sub>11</sub> H <sub>22</sub>	Training Set	154.3	0.2141	465.8	746.6	n-alkene	51
1-dodecene	C <sub>12</sub> H <sub>24</sub>	Training Set	168.3	0.2141	486.5	754.8	n-alkene	52
1-tridecene	C <sub>13</sub> H <sub>26</sub>	Training Set	182.3	0.2141	505.9	761.9	n-alkene	53
2-methyl-2-tridecene	C <sub>14</sub> H <sub>28</sub>	Training Set	196.4	0.2141	513.0	769.4	n-alkene	54
1-tetradecene	C <sub>14</sub> H <sub>28</sub>	Training Set	196.4	0.2141	524.3	768.2	n-alkene	55
3-methyl-1-tetradecene	C <sub>15</sub> H <sub>30</sub>	Training Set	210.4	0.2141	529.0	770.4	n-alkene	56
1-Pentadecene	C <sub>15</sub> H <sub>30</sub>	Training Set	210.4	0.2141	541.6	772.7	n-alkene	57
1-Pentadecene, 2-methyl-	C <sub>16</sub> H <sub>32</sub>	Training Set	224.4	0.2141	545.0	780.9	n-alkene	58
2-Methyl-1-octadecene	C <sub>19</sub> H <sub>38</sub>	Training Set	266.5	0.2141	602.0	908.0	n-alkene	59
2-methyl-2-nonadecene	C <sub>20</sub> H <sub>40</sub>	Training Set	280.5	0.2141	619.5	796.1	n-alkene	60
7-hexyl-7-pentadecene	C <sub>21</sub> H <sub>42</sub>	Training Set	294.6	0.2141	639.0	937.0	n-alkene	61
8-heptyl-7-pentadecene	C <sub>22</sub> H <sub>44</sub>	Training Set	308.6	0.2141	639.0	801.0	n-alkene	62
9-octyl-8-heptadecene	C <sub>25</sub> H <sub>50</sub>	Training Set	350.7	0.2141	649.0	805.6	n-alkene	63
1-hexacosene	C <sub>26</sub> H <sub>52</sub>	Training Set	364.7	0.2141	671.0	982.0	n-alkene	64
10-nonyl-7-nonadecene	C <sub>28</sub> H <sub>56</sub>	Training Set	392.8	0.2141	722.5	975.0	n-alkene	65
1-Undecene, 2-methyl-	C <sub>12</sub> H <sub>24</sub>	Training Set	168.3	0.2141	478.0	758.5	n-alkene	66
1-Nonene, 2,4,6,8-tetramethyl	C <sub>13</sub> H <sub>26</sub>	Training Set	168.3	0.2141	470.0	764.0	n-alkene	67
1-Decene, 2-methyl-	C <sub>11</sub> H <sub>22</sub>	Training Set	96.17	0.2141	458.0	750.7	n-alkene	68
1-heptyne	C <sub>7</sub> H <sub>12</sub>	Training Set	96.17	0.1978	372.9	728.7	n-alkyne	69
1-octyne	C <sub>8</sub> H <sub>14</sub>	Training Set	110.2	0.1910	399.4	742.0	n-alkyne	70
1-nonyne	C <sub>9</sub> H <sub>16</sub>	Training Set	124.2	0.2015	424.0	753.3	n-alkyne	71
1-decyne	C <sub>10</sub> H <sub>18</sub>	Training Set	138.2	0.2027	446.8	764.6	n-alkyne	72
1-undecyne	C <sub>11</sub> H <sub>20</sub>	Training Set	152.3	0.2038	477.0	770.3	n-alkyne	73
2,9-Dimethyl-5-decyne	C <sub>12</sub> H <sub>21</sub>	Training Set	166.3	0.2047	482.0	778.2	n-alkyne	74
1-dodecyne	C <sub>12</sub> H <sub>22</sub>	Training Set	166.3	0.2047	505.0	777.3	n-alkyne	75
1-tetradecyne	C <sub>14</sub> H <sub>26</sub>	Training Set	194.4	0.2060	525.6	849.0	n-alkyne	76
1-pentadecyne	C <sub>15</sub> H <sub>28</sub>	Training Set	208.4	0.20662	553.1	828.0	n-alkyne	77
2,6,6-trimethyl-3,3-diisopropyl-4-heptyne	C <sub>16</sub> H <sub>30</sub>	Training Set	222.4	0.2070	489.8	818.3	n-alkyne	78
3-heptadecyne	C <sub>17</sub> H <sub>32</sub>	Training Set	236.4	0.2075	573.0	881.0	n-alkyne	79

Compound	Formula	Data Base	Molar mass, g/mol	$\alpha$ , mol·g <sup>-1</sup>	T <sub>b</sub> at 1 atm / K	Density at 25°C, kg·m <sup>-3</sup>	Family	Sample No.
3-octadecyne	C <sub>18</sub> H <sub>34</sub>	Training Set	250.5	0.2078	599.2	801.6	n-alkyne	80
1-nonadecyne	C <sub>19</sub> H <sub>36</sub>	Training Set	264.5	0.2082	605.4	909.0	n-alkyne	81
Heptane	C <sub>7</sub> H <sub>16</sub>	Training Set	100.2	0.2298	371.5	679.6	n-alkane	82
Hexane, 2-methyl-	C <sub>7</sub> H <sub>16</sub>	Training Set	100.2	0.2298	363.1	674.5	n-alkane	83
n-Decane	C <sub>10</sub> H <sub>22</sub>	Training Set	142.3	0.2251	447.3	726.6	n-alkane	84
Octane, 2,6-dimethyl-	C <sub>10</sub> H <sub>22</sub>	Training Set	142.3	0.2252	431.5	725.3	n-alkane	85
n-Undecane	C <sub>11</sub> H <sub>24</sub>	Training Set	156.3	0.2242	469.0	736.8	n-alkane	86
Nonane, 3,7-dimethyl-	C <sub>11</sub> H <sub>24</sub>	Training Set	156.3	0.2233	449.4	769.0	n-alkane	87
Tridecane	C <sub>13</sub> H <sub>28</sub>	Training Set	184.4	0.2215	508.6	752.9	n-alkane	88
2,4-dimethyl-4-(1,1-dimethylethyl)heptane	C <sub>13</sub> H <sub>28</sub>	Training Set	184.4	0.2227	479.9	793.1	n-alkane	89
Pentadecane	C <sub>15</sub> H <sub>32</sub>	Training Set	212.4	0.2215	543.8	765.1	n-alkane	90
Dodecane, 2,6,11-trimethyl-	C <sub>15</sub> H <sub>32</sub>	Training Set	212.4	0.2215	520.6	826.0	n-alkane	91
5,5-Dibutylnonane	C <sub>17</sub> H <sub>36</sub>	Training Set	240.5	0.2207	535.0	777.7	n-alkane	92
Octadecane, 2,6-dimethyl-	C <sub>20</sub> H <sub>42</sub>	Training Set	282.6	0.2110	598.2	909.0	n-alkane	93
Eicosane, 2,4-dimethyl-	C <sub>22</sub> H <sub>46</sub>	Training Set	310.6	0.2192	605.3	789.8	n-alkane	94
Decane, 2,4-dimethyl-	C <sub>12</sub> H <sub>26</sub>	Training Set	170.3	0.2233	473.2	744.6	n-alkane	95
ethylcyclopentane	C <sub>7</sub> H <sub>14</sub>	Training Set	98.19	0.2141	376.6	762.3	naphthenic	96
ethylcyclohexane	C <sub>8</sub> H <sub>16</sub>	Training Set	112.2	0.2141	404.9	784.3	naphthenic	97
butylcyclopentane	C <sub>9</sub> H <sub>18</sub>	Training Set	126.2	0.2141	429.3	780.9	naphthenic	98
1,1'-Bicyclopentyl	C <sub>10</sub> H <sub>18</sub>	Training Set	138.2	0.2027	463.6	861.0	naphthenic	99
Naphthalene, decahydro-2-methyl-	C <sub>11</sub> H <sub>20</sub>	Training Set	152.3	0.2038	475.0	886.0	naphthenic	100
cycloundecane	C <sub>11</sub> H <sub>22</sub>	Training Set	154.3	0.2141	456.8	804.0	naphthenic	101
1,1,3-	C <sub>21</sub> H <sub>38</sub>	Training Set	290.5	0.2033	656.0	935.0	naphthenic	102
Tricyclohexylpropane								
Cyclopentane, 1,1'-[4-(3-cyclopentylpropyl)-1,7-heptanediyl]bis-	C <sub>25</sub> H <sub>46</sub>	Training Set	346.6	0.2050	684.9	885.4	naphthenic	103
1-butyl-2,2,6-trimethylcyclohexane	C <sub>13</sub> H <sub>26</sub>	Training Set	182.3	0.2141	491.0	846.0	naphthenic	104
Butyldecalin	C <sub>14</sub> H <sub>26</sub>	Training Set	194.4	0.2060	537.1	872.7	naphthenic	105
Cyclohexane, (3-cyclopentylpropyl)-	C <sub>14</sub> H <sub>26</sub>	Training Set	194.4	0.2060	543.6	864.3	naphthenic	106
Octane, 2-cyclohexyl-	C <sub>14</sub> H <sub>28</sub>	Training Set	196.4	0.2141	528.6	820.1	naphthenic	107
2-Isopropylbicyclohexyl	C <sub>15</sub> H <sub>28</sub>	Training Set	208.4	0.2066	553.3	894.5	naphthenic	108
Cyclohexane, 1,1'-(1-methylethylidene)bis-	C <sub>15</sub> H <sub>28</sub>	Training Set	208.4	0.2066	559.4	903.8	naphthenic	109
2,6,6,9-tetramethylcycloundecane	C <sub>15</sub> H <sub>30</sub>	Training Set	210.4	0.2141	514.0	820.0	naphthenic	110
1,1'-Bicyclohexyl, 2-butyl-	C <sub>16</sub> H <sub>30</sub>	Training Set	222.4	0.2070	568.5	882.1	naphthenic	111
1,1-dicyclohexylbutane	C <sub>16</sub> H <sub>30</sub>	Training Set	222.4	0.2070	566.0	885.9	naphthenic	112
1,5-dicyclohexylpentane	C <sub>17</sub> H <sub>32</sub>	Training Set	236.4	0.2075	598.0	866.3	naphthenic	113
Cyclohexane, 1,1'-(1,2-ethanediyl)bis-	C <sub>14</sub> H <sub>26</sub>	Training Set	194.4	0.2060	545.8	872.4	naphthenic	114
1,1-dicyclohexylheptane	C <sub>19</sub> H <sub>36</sub>	Training Set	264.5	0.2082	620.9	885.5	naphthenic	115
Naphthalene, decahydro-	C <sub>20</sub> H <sub>38</sub>	Training Set	278.5	0.2085	624.8	866.0	naphthenic	116
2,6-dimethyl-3-octyl-								
Heptane, 1,1-dicyclohexyl-	C <sub>19</sub> H <sub>36</sub>	Training Set	264.5	0.2082	620.9	885.5	naphthenic	117

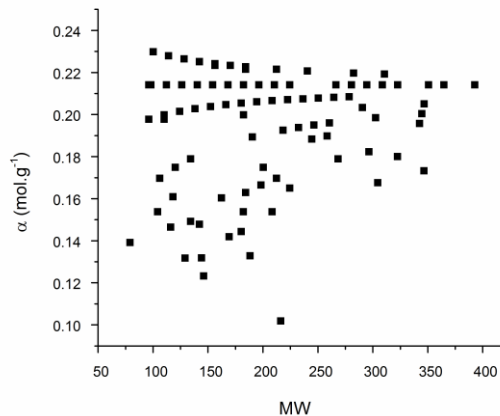
Compound	Formula	Data Base	Molar mass, g/mol	$\alpha$ , mol·g <sup>-1</sup>	T <sub>b</sub> at 1 atm / K	Density at 25°C, kg·m <sup>-3</sup>	Family	Sample No.
1H-Indene, 5-butyl-6-hexyloctahydro-	C <sub>19</sub> H <sub>36</sub>	Training Set	264.5	0.2082	606.5	866.0	naphthenic	118
Chrysene, octadecahydro-	C <sub>18</sub> H <sub>30</sub>	Training Set	246.45	0.1950	645.4	977.7	naphthenic	119
Cyclohexane, (1,2-dimethylbutyl)-	C <sub>12</sub> H <sub>24</sub>	Training Set	168.35	0.2141	479.7	831.1	naphthenic	120
Cyclohexane, (1-octylonyl)-	C <sub>23</sub> H <sub>46</sub>	Training Set	322.65	0.2141	645.3	830.9	naphthenic	121
1-Fluoronaphthalene	C <sub>10</sub> H <sub>7</sub> F	Training Set	146.2	0.1232	484.5	1131.5	heteroatom	122
1,3-butanedione, 4,4,4-trifluoro-1-phenyl-	C <sub>10</sub> H <sub>7</sub> F <sub>3</sub>	Training Set	216.2	0.1018	517.0	1540	heteroatom	123
4-Phenyl-3-butyn-2-one	C <sub>10</sub> H <sub>8</sub> O	Training Set	144.2	0.1319	522.0	1023.3	heteroatom	124
Dibenzothiophene, 1,2,3,4-tetrahydro-	C <sub>12</sub> H <sub>12</sub> S	Training Set	188.3	0.1329	594.7	1143	heteroatom	125
pyridine	C <sub>5</sub> H <sub>5</sub> N	Training Set	79.10	0.1391	388.4	978.2	heteroatom	126
2-methyl-6-phenylpyridine	C <sub>12</sub> H <sub>11</sub> N	Training Set	169.2	0.1419	548.3	1085	heteroatom	127
(1-methylethylidene)cyclohexane	C <sub>9</sub> H <sub>16</sub>	Training Set	124.2	0.2015	434.1	832	heteroatom	128
Styrene	C <sub>8</sub> H <sub>8</sub>	Test/Comp. Set	104.1	0.1537	418.4	901.7	Aromatic	129
1,2-dimethylbenzene	C <sub>8</sub> H <sub>10</sub>	Test/Comp. Set	106.2	0.1697	417.5	875.7	Aromatic	130
3,3'-Dimethylbiphenyl	C <sub>14</sub> H <sub>14</sub>	Test/Comp. Set	182.3	0.1537	562.7	994.9	Aromatic	131
Naphthalene, 1-butyl-	C <sub>14</sub> H <sub>16</sub>	Test/Comp. Set	184.3	0.1629	560.7	971.5	Aromatic	132
Naphthalene, 1,6-dimethyl-4-(1-methylethyl)-	C <sub>15</sub> H <sub>18</sub>	Test/Comp. Set	198.3	0.1665	575.0	974.2	Aromatic	133
1,2-diphenyl-1-butene	C <sub>16</sub> H <sub>16</sub>	Test/Comp. Set	208.3	0.1537	590.0	1008	Aromatic	134
2-Heptene, 4-methyl-, (E)-	C <sub>8</sub> H <sub>16</sub>	Test/Comp. Set	112.2	0.2141	398.2	740	n-alkene	135
1-Pentene, 2,3-dimethyl-	C <sub>7</sub> H <sub>14</sub>	Test/Comp. Set	98.19	0.2141	357.4	700.7	n-alkene	136
1-Hexene, 3,5,5-trimethyl-	C <sub>9</sub> H <sub>18</sub>	Test/Comp. Set	126.2	0.2141	394.5	719.6	n-alkene	137
2,3-Dimethyl-2-octene	C <sub>10</sub> H <sub>20</sub>	Test/Comp. Set	140.3	0.2141	442.0	757.3	n-alkene	138
1-Decene, 2-methyl-	C <sub>11</sub> H <sub>22</sub>	Test/Comp. Set	154.3	0.2141	458.0	750.7	n-alkene	139
4-Octyne	C <sub>8</sub> H <sub>14</sub>	Test/Comp. Set	110.2	0.1998	406.6	747.3	n-alkyne	140
2,7-dimethyl-4-octyne	C <sub>10</sub> H <sub>18</sub>	Test/Comp. Set	138.2	0.2027	432.0	758.3	n-alkyne	141
3,3-dimethyl-4-nonyne	C <sub>11</sub> H <sub>20</sub>	Test/Comp. Set	152.3	0.2038	451.0	762.5	n-alkyne	142
3,3-dimethyl-4-decyne	C <sub>12</sub> H <sub>22</sub>	Test/Comp. Set	166.3	0.2047	474.6	769.9	n-alkyne	143
Hexane, 2,5-dimethyl-	C <sub>8</sub> H <sub>18</sub>	Test/Comp. Set	114.2	0.2280	382.2	690.01	n-alkane	144

Compound	Formula	Data Base	Molar mass, g/mol	$\alpha$ , mol·g <sup>-1</sup>	T <sub>b</sub> at 1 atm / K	Density at 25°C, kg·m <sup>-3</sup>	Family	Sample No.
Hexane, 2,3,5-trimethyl-	C <sub>9</sub> H <sub>20</sub>	Test/Comp. Set	128.3	0.2264	404.0	718.0	n-alkane	145
Octane, 3,3-dimethyl-	C <sub>10</sub> H <sub>22</sub>	Test/Comp. Set	142.3	0.2252	433.5	735.2	n-alkane	146
Nonane, 2,7-dimethyl	C <sub>11</sub> H <sub>24</sub>	Test/Comp. Set	156.3	0.2242	449.5	774	n-alkane	147
Cyclopentane, propyl-	C <sub>8</sub> H <sub>16</sub>	Test/Comp. Set	112.2	0.2141	404.0	772.4	naphthenic	148
1,1,4-Trimethylcyclohexane	C <sub>9</sub> H <sub>18</sub>	Test/Comp. Set	126.2	0.2141	407.6	767.6	naphthenic	149
Cyclooctane, ethyl-	C <sub>10</sub> H <sub>20</sub>	Test/Comp. Set	140.3	0.2141	463.1	833.5	naphthenic	150
1-butyl-1-methyl-cyclohexane	C <sub>11</sub> H <sub>22</sub>	Test/Comp. Set	154.3	0.2141	464.7	809.9	naphthenic	151
quinoline	C <sub>9</sub> H <sub>7</sub> N	Test/Comp. Set	129.2	0.1317	510.2	1090	heteroatom	152
nicotine	C <sub>10</sub> H <sub>14</sub> N <sub>2</sub>	Test/Comp. Set	162.2	0.1604	523.2	1005	heteroatom	153
chroman	C <sub>9</sub> H <sub>10</sub> O	Test/Comp. Set	134.2	0.1492	488.7	1060	heteroatom	154

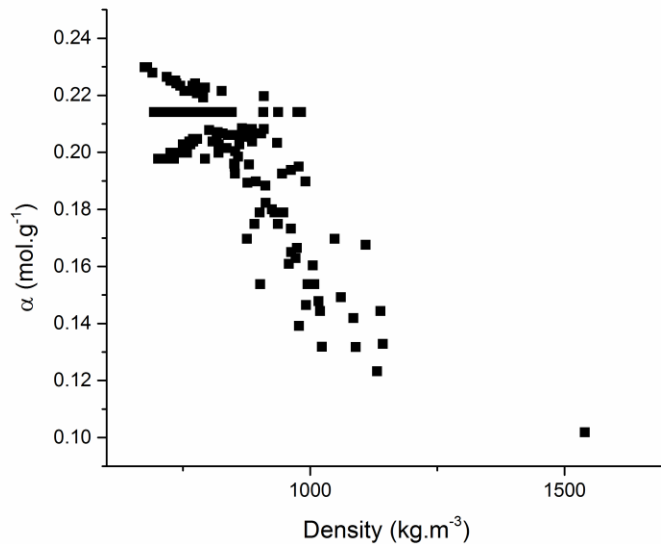
<sup>1</sup> The property values are taken from the NIST Webbook [82]



(a)



(b)



(c)

Figure 5-2: Dependence of  $\alpha$  on: a)  $T_b$ , b) MW, c) density .

Figure 5-2 confirms that  $\alpha$  has more specific trend with density compared to boiling temperature and molecular weight. Consequently, density appears to be a robust basis for the development of a correlation but not adequate, since  $\alpha$  does not have a well understood relation



to the density. Molecular weight and boiling temperature showed the same trend, so having both as independent variables is not necessary. Least square method is needed to be applied in order to find a linear correlation as a function of density and either molecular weight or boiling temperature.

### 5.2.2 Similarity Variable Prediction by Correlation

As it is mentioned before, one of the categories for the least square method is multiple linear regressions (MLR). It can be applied to develop a correlation when independent variables are few in the number, are not collinear, and have a well understood relationship to the dependent variable; otherwise, partial least square can be a better method to develop a model.

It is illustrated in Figure 5-2 that  $\alpha$  does not have any specific trend with molecular weight and boiling temperature, while the trend between  $\alpha$  and density is obvious; however, the scatter is large. In order to find any collinearity between density and either molecular weight or boiling temperature, the r value [82], linear correlation coefficient, is calculated using Equation 5-1 and the results are 0.5311, 0.3436, 0.9389; for {Density at 25°C and MW}, {Density at 25°C and T<sub>b</sub>}, and {MW and T<sub>b</sub>}, respectively.

$$r = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2} \sqrt{n(\sum y^2) - (\sum y)^2}} \quad 5-1$$

where x and y are two variables and n is the number of available data. The range of values for r is between -1 and +1 which relates to negative and positive linear correlation, respectively [82]. The r value for the independent variable shows how close these variables are collinear. Independent variables are called collinear when the r value is higher than 0.05 [83].

The trend of changes for density, MW, and  $T_b$  values for 154 organic liquid compounds,

Table 5-1, is illustrated in Figure 5-3 versus sample number which is 1 to 154:

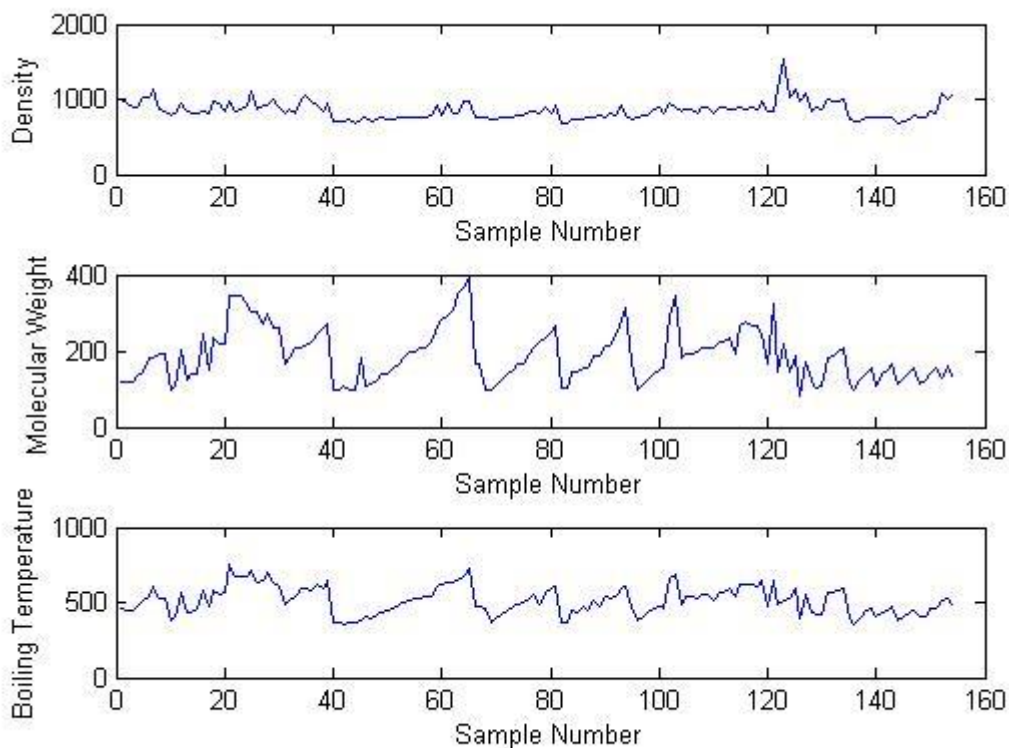


Figure 5-3: Changes of MW,  $T_b$ , and density versus sample number

It can be seen from Figure 5-3 and the  $r$  value that MW and  $T_b$  are strongly collinear, and there is a collinearity between density and MW/ $T_b$ . Accordingly, PLS would be the best method to develop a linear model to estimate similarity variable as a function of density at 25 C and either MW or  $T_b$ .

### 5.2.3 Similarity Variable Database

Liquids from Table 5-1 used to develop the similarity variable correlation were distributed into three data sets: a training data set comprising 128 organic liquid compounds (including paraffins, naphthenes, aromatics, sulphur/oxygen/nitrogen derivatives) used to regress correlation coefficients, a test data set comprising 26 organic compounds (including paraffins, naphthenes, aromatics, sulphur/oxygen/nitrogen derivatives) used to evaluate the predictive character of the correlation, and a comparison data set comprises 4 similarity variable values for four prepared mixtures in chapter 4, used to evaluate the relative performance of the correlation vis-à-vis the API procedure. In total, 154 compound with density ranging from 670 to 1200 kg.m<sup>-3</sup> and molar mass ranging from 79 to 400 g/mol were used to cover the wide range of properties. The training data set density range was from 701 to 1134 kg.m<sup>-3</sup> and the molar mass was from 97 to 393 g/mol. The test data set density range was from 690 to 1090 kg.m<sup>-3</sup> and the molar mass was from 98 to 209 g/mol.

### 5.2.4 Results

To obtain the correlation, the MATLAB R2012a [80] was used for programming a PLS method to find the linear predictive correlation over the training data set presented in Table 5-1. The possible combinations of inputs were {Density and MW}, {Density and T<sub>b</sub>}, and {Density and T<sub>b</sub> and MW}. All the combinations were examined and the root mean square error (RMSE) and the bias of the models are shown in Table 5-2.

Table 5-2: The RMSE for different combinations of inputs

	RMSE	Bias
MW and Density	0.011	$3.77*10^{-17}$
MW and $T_b$	0.020	$1.73*10^{-17}$
Density and $T_b$	0.012	$1.14*10^{-16}$
MW, $T_b$ , and Density	0.011	$4.37*10^{-17}$

Among two-variable correlations, the combination of {Density and MW} was shown to work better than {Density and  $T_b$ }. Although the same deviation occurred for {Density, MW, and  $T_b$ } and {Density and MW} as the inputs, having the input as just {Density and MW} makes the model less complicated. Moreover, since there is not much difference between {Density and MW} and {Density and  $T_b$ } as the inputs, both models have been developed so that if the MW is not available one can estimate the  $\alpha$  value by having  $T_b$ . The correlation with different input required are as follows:

$$\alpha = \beta_0 + \beta_1 * p + \beta_2 * MW \quad 5-2$$

$$\alpha = \delta_0 + \delta_1 * p + \delta_2 * T_b \quad 5-3$$

Where  $\alpha$  is similarity variable,  $p$  is density at 25 °C ( $\text{kg.m}^{-3}$ ), MW is the molecular weight,  $T_b$  is boiling temperature for the single compound and average boiling temperature for the mixture (K),  $\beta_i$  and  $\delta_i$  are the universal coefficients with values shown in Table 5-3:

Table 5-3: Universal coefficient for the Equation 5-2 and 5-3

$\beta_0$	0.3412	$\gamma_0$	0.3192
$\beta_1$	$-1.8586 \cdot 10^{-4}$	$\gamma_1$	$-1.9645 \cdot 10^{-4}$
$\beta_2$	$1.2283 \cdot 10^{-4}$	$\gamma_2$	$9.0678 \cdot 10^{-5}$

The quality of models is illustrated in Figure 5-4 to 5-7 using training and test data sets. In this way, the uncertainty of the estimated heat capacities is reduced from the maximum noted in section 5.1 (+20 to -35 %; +0.4 to -0.6 J.K<sup>-1</sup>.g<sup>-1</sup>). to +0.025 to -0.025 J.K<sup>-1</sup>.g<sup>-1</sup> using a simple and simple to implement correlation.

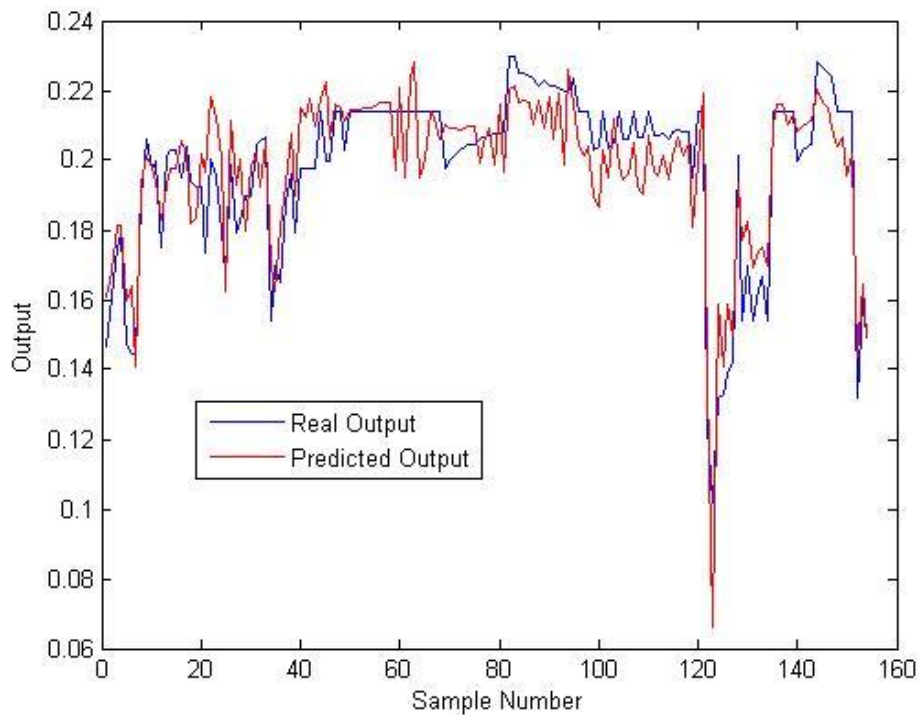


Figure 5-4: The quality of the model (Equation 5-2) over the training data set

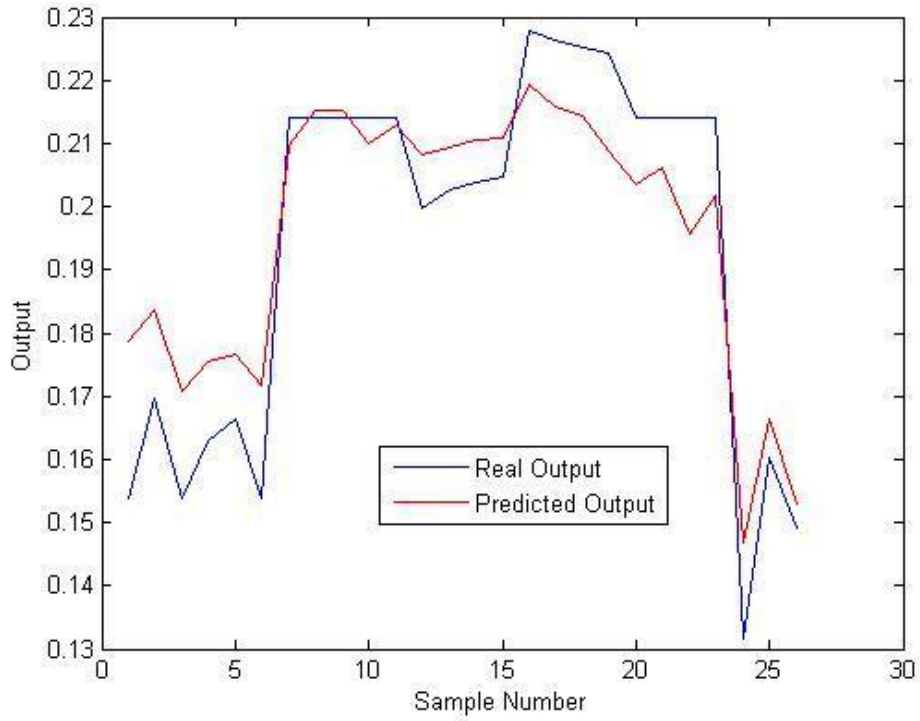


Figure 5-5: The quality of the model (Equation 5-2) over the test data set

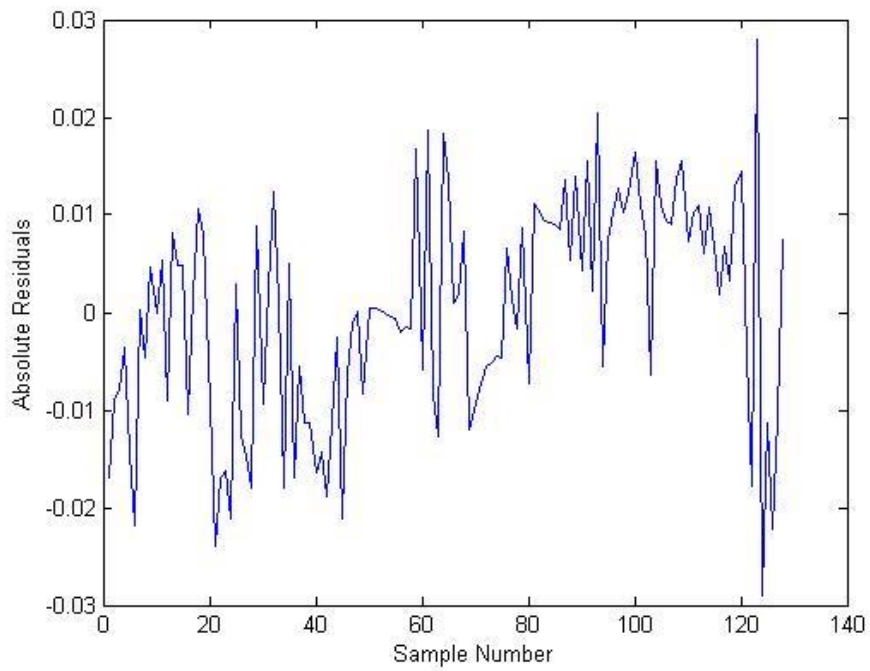


Figure 5-6: The absolute residual of the model (Equation 5-2) over the training data set

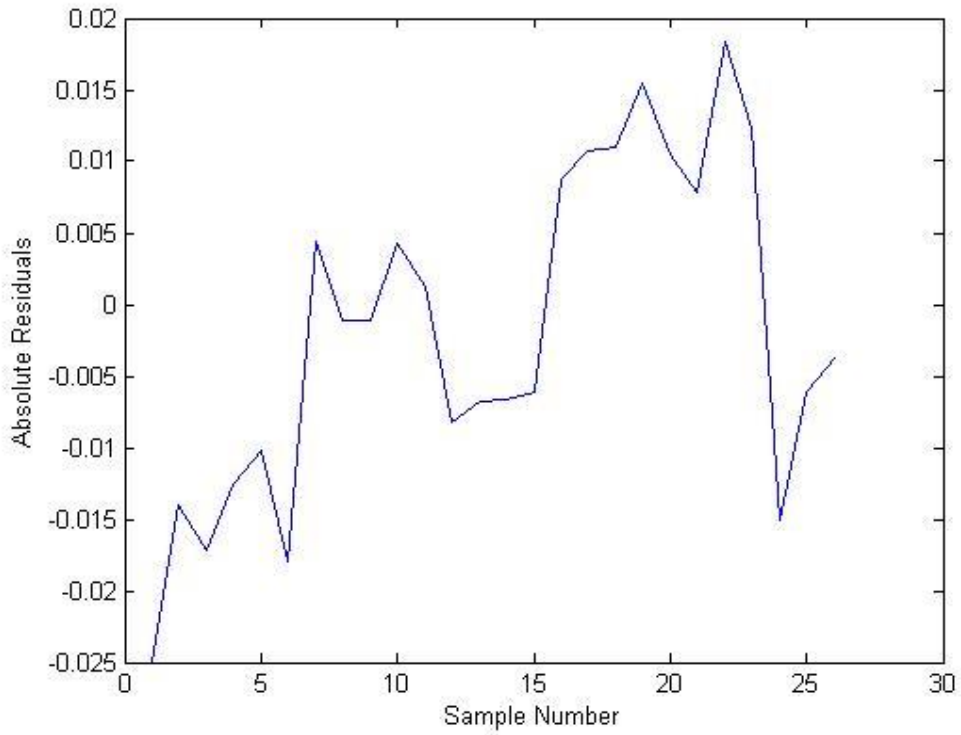


Figure 5-7: The absolute residual of the model (Equation 5-2) over the test data set

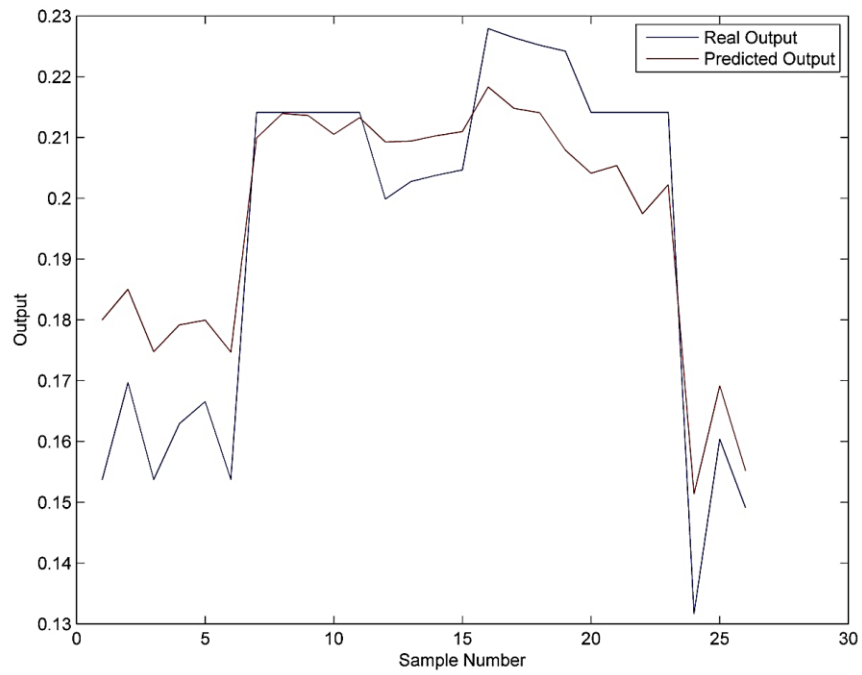


Figure 5-8: The quality of the model (Equation 5-3) over the test data set

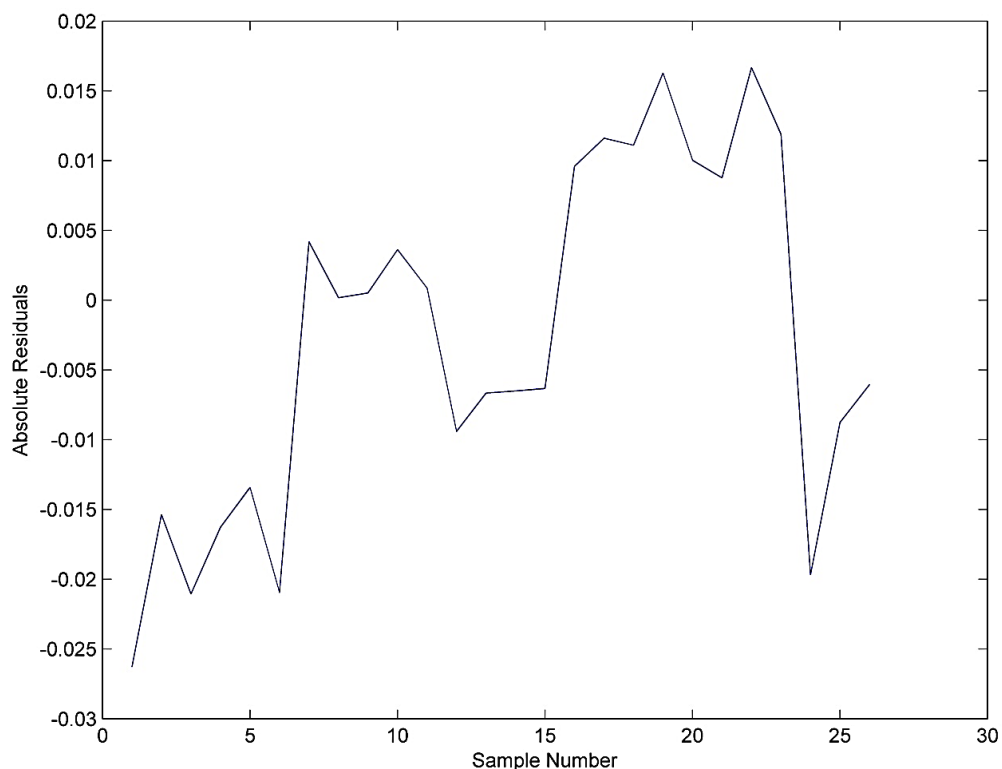


Figure 5-9: The absolute residual of the model (Equation 5-3) over the test data set

## 5.3 Artificial Neural Network Model

### 5.3.1 Network design

Another possible method to develop a model to predict  $\alpha$  is Artificial Neural Network, since there is no well-defined relationship between  $\alpha$  and the properties mentioned above. The ANN is applied to see if a model can be developed with higher precision and accuracy. The possible input variables are molecular weight, average boiling temperature, and density at 25 °C. According to the complexity of the problem, the back propagation learning algorithm has been used in a network with one hidden layer and feed forward algorithm. Variants of the algorithm, which should be specified, are the required inputs, number of nodes in the hidden layer, training



(learning) function and the transfer function. In order to train a neural network the data set, Table 5-1, comprising 154 organic liquid compounds are used, where 70, 15, and 15 % of the data are used as the training, test, and validation data sets by the network, respectively. The error is calculated as the mean square error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n (t_i - o_i)^2 \quad 5-4$$

Where  $t_i$  is the target value, and  $o_i$  is the desired output value.

For studying required inputs to train the neural network, five configurations of inputs were considered and the results are compared. Levenberg-Marquardt back propagation (trainlm) [84,85] is used as the training algorithm, and Verhulst logistic sigmoid [86] (logsig) is used as the transfer function, while the hidden layer size is selected to be 8. The results for MSE for each configuration are shown in Table 5-4. According to the results, having all three terms as the required inputs (Tb, MW, and density) yields the least error; however, there is not much difference between a and d. Since fewer required inputs make the process less complex, density and MW are taken into account as the inputs.

Table 5-4: Error evaluation for different input configurations

Inputs	MSE	Bias
a) Density and MW	9.07E-05	-4.2E-05
b) Density and $T_b$	9.52E-05	-5.9E10-4
c) MW and $T_b$	2.66E-04	3.6E-04
d) Density, MW and $T_b$	7.35E-05	-6.4E-05
e) Density	1.91e-04	1.9E04

The size of hidden layer is defined by the number of nodes used in that layer. For finding an optimum size of the hidden layer, a neural network is performed for four different sizes and the performance is evaluated. Levenberg-Marquardt back propagation (trainlm) and logistic sigmoid (logsig) are used as the training algorithm and the transfer function, respectively. The inputs of the ANN are density and MW. The results for MSE for each size are presented in Table 5-5.

Table 5-5: Error evaluation for different hidden layer sizes

size	MSE	Bias
a) 4	1.2526e-04	-5.9E-05
b) 5	8.3918e-05	-1.7E-05
c) 6	7.9552e-05	2.3E-04
d) 7	8.1637e-05	9.3E-05
e) 8	8.3107e-05	1.0E-03
f) 9	9.4921e-05	1.8E-04
g) 10	8.9158e-05	-2.0E-03

The number of hidden neurons can be selected as five. Increasing the number of neurons to more than five will result in an over trained network as it is shown in the Figure 5-10 and 5-11. Therefore, results of the predicted values from the neural network will deviate from the real values as it is shown in Figure 5-10 and 5-11 for the training, validation, test, and all data sets.

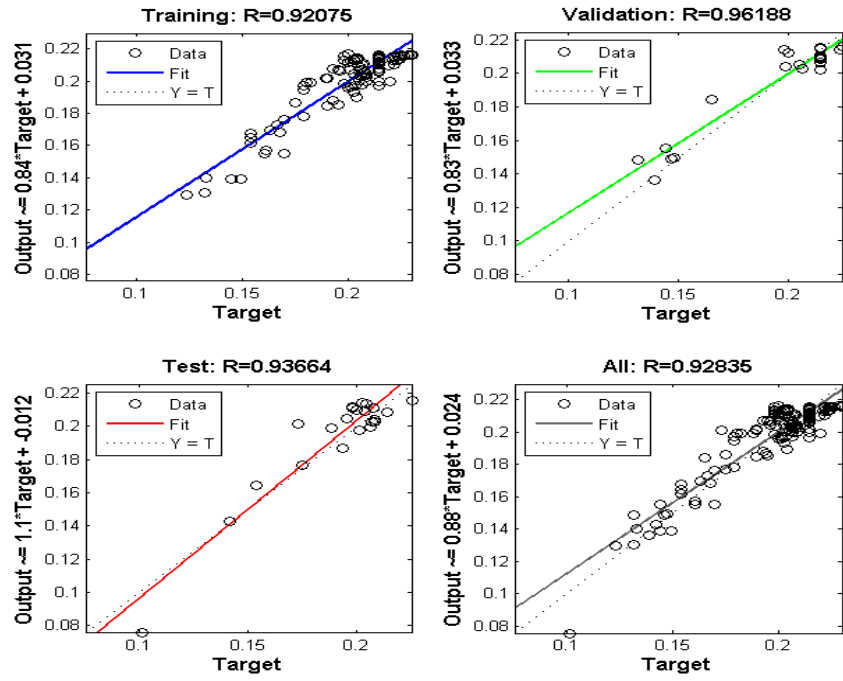


Figure 5-10: ANN regression graph with hidden layer size of 5

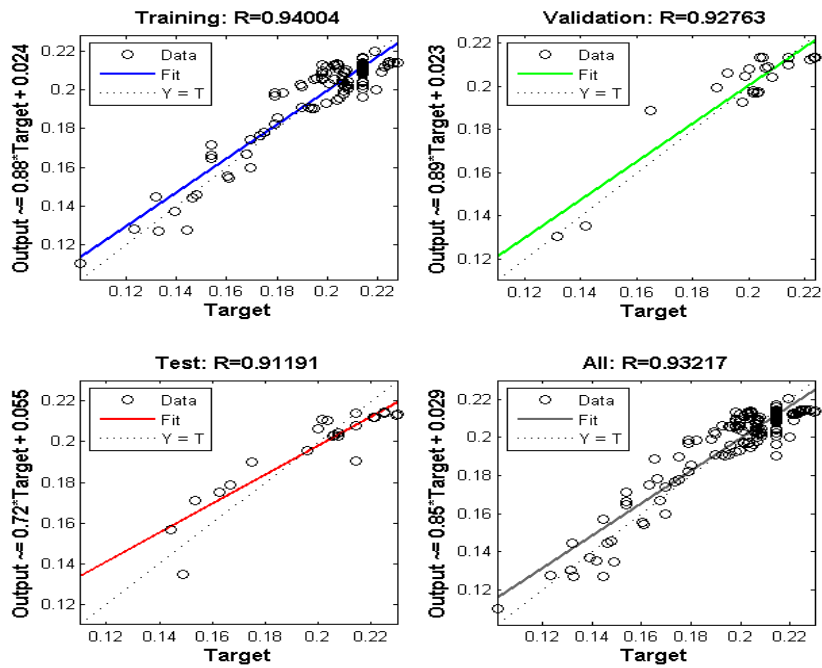


Figure 5-11: ANN regression graph with a hidden layer size of 6

Variants of the training (learning) algorithms used in the study are scaled conjugate gradient (SCG) [87], Polak–Ribiere conjugate gradient (CGP)[88], and Levenberg–Marquardt (LM) [84]. The conjugate gradient is an iterative method for the numerical solution of specific systems of nonlinear equations; it can be used to sparse systems that are too large to be examined by direct methods. In mathematics and computing, the LM algorithm is a numerical solution to the minimization problems, mostly nonlinear, over the parameters of the function. Minimizing a function arises especially in the least squares curve fitting and nonlinear programming. The same procedure as above is performed for the evaluation of each algorithm and the results are compared. The best algorithm which has minimum errors is the LM algorithm, as it is noted in Table 5-6.

Table 5-6: Error evaluation for trained ANN with different training algorithms

algorithm	MSE	Bias
a) SCG	1.27E-04	1.0E-03
b) CGP	1.26E-04	-1.0E-03
c) LM	8.37E-05	2.8E-5

Variants of the transfer function used in the study are Linear transfer function (purelin) [89], Log-Sigmoid transfer function (logsig) [90], Radial basis function (radbas) [91], and Triangular basis function (tribas) [92]. Transfer functions compute a layer's output from its net input. The purelin has a linear form while the logsig function is a mathematical function having an "S" shape. The radbas is a real valued function whose value relates only on the distance from some

other point. The tribas is a function which has a triangular form. The minimum error is resulted from tribas transfer function, as it is presented in Table 5-7.

Table 5-7: Error evaluation for trained ANN with different transfer functions

transfer function	MSE	Bias
a) purelin	1.27E-04	8.2E-05
b) logsig	8.94E-05	1.2E-05
c) radbas	8.82E-05	-4.4E-05
d) tribas	8.22E-05	-5.7E-05

Based on the above mentioned results, the following parameters of the ANN developed for estimating the similarity variable were selected. The ANN is trained based on the back propagation learning algorithm in feed-forward single hidden layer network. As it is mentioned, there are two (number of variant), five and one (number of output) neurons on input layer, hidden layer and output layer, respectively. One selected hidden layer, the corresponding nodes, Levenberg–Marquardt (LM) training algorithm and the Triangular basis (tribas) transfer function have been investigated through the trial and error mechanism. For all the procedures mentioned above a computer program has been performed under MATLAB and the associated code is presented in Appendix A.

The model evaluation is illustrated in Figure 5-12 which shows the performance of the trained ANN for the training, validation, test, and all data sets by plotting the predicted values (output) versus real values (target). As it is shown, the overall regression coefficient (R) is 0.9315 which proves the accuracy of the ANN.

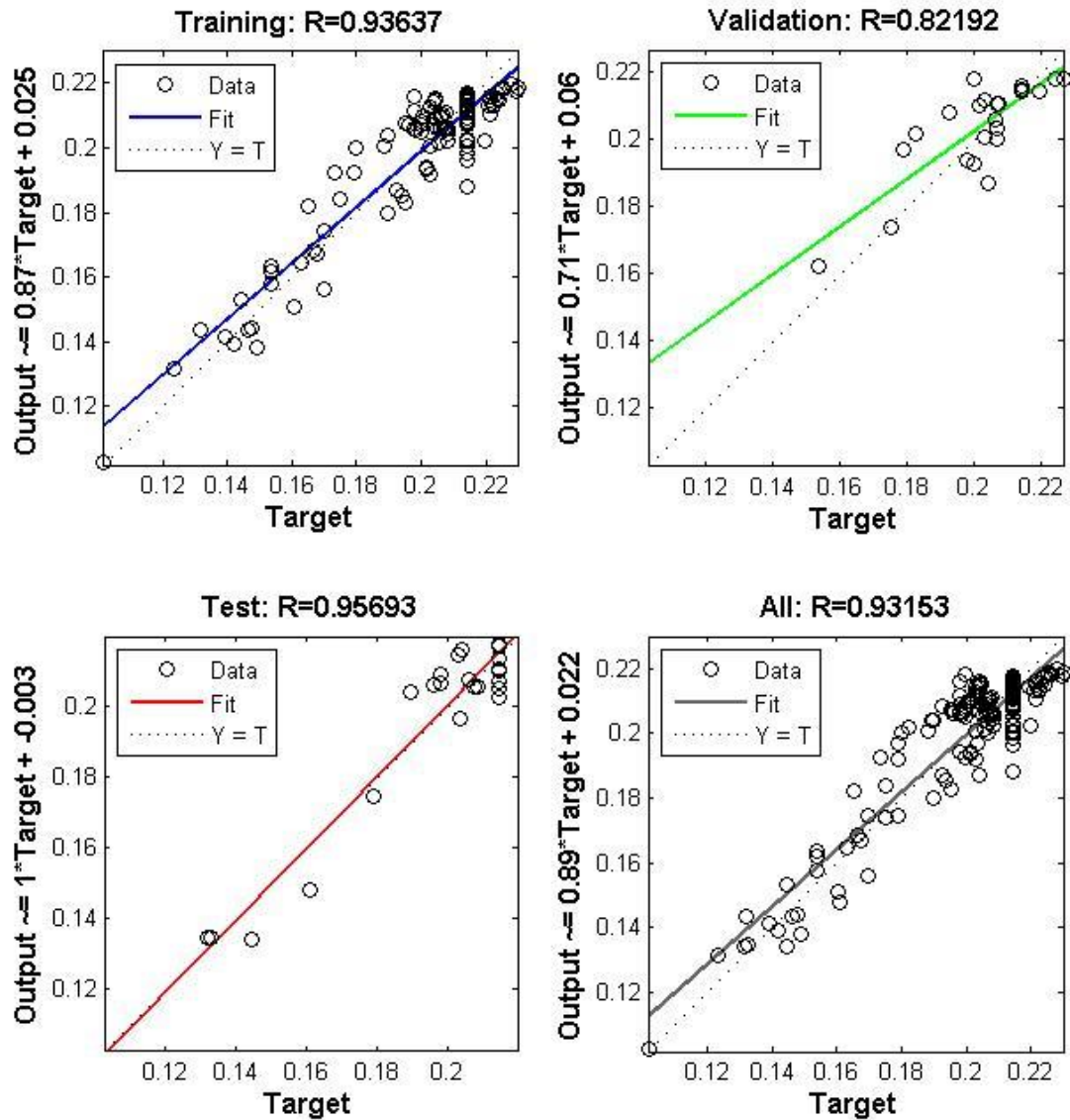


Figure 5-12: The evaluation of the ANN developed to predict  $\alpha$

## 5.4 Comparison

The Root Mean Square Deviation (RMSD) is a widely used method for measuring the difference between predicted values by a model and real values. The difference between real and predicted values are called residual when the RMSD is applied over the training data set that was used for developing a model, and are called prediction error when it is applied over the test data set. RMSD, Equation 5-5, is a good method for comparing predicting deviation of different models [93,94]. Also, to show the model tendency to over or under estimation the bias value is calculated based on Equation 5-6:

$$RMSD = \sqrt{\frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{n}} \quad 5-5$$

$$Bias = \left[ \sum (y_t - \hat{y}_t) \right] / n \quad 5-6$$

Where  $\hat{y}_t$  is the predicted value, and  $y_t$  is the real value for n different points from the training data set. Also, to show the model tendency to over or under estimation the bias value is calculated by Equation 5-6:

The RMSD values for the model developed with the use of PLS and ANN are listed in Table 5-8.

Table 5-8: RMSD and Bias value for Equation 5-2, 5-3 and ANN models over the training data set

	RMSD	Bias
Equation 5-2	0.011	$3.77 \cdot 10^{-17}$
Equation 5-3	0.012	$1.14 \cdot 10^{-16}$
ANN	0.011	$5.84 \cdot 10^{-10}$

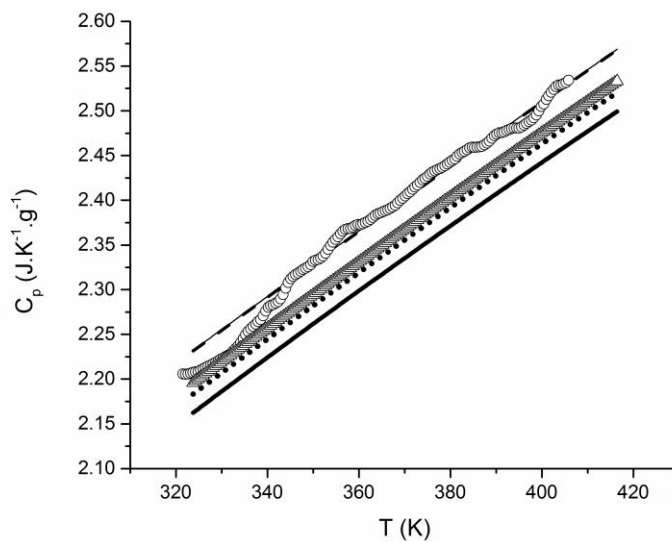


The result for the test data set comprising 26 organic liquid compounds, Table 5-1, (including paraffins, naphthenes, aromatics, sulphur/oxygen/nitrogen derivatives) are compared and the RMSD value as the prediction error is shown in Table 5-9.

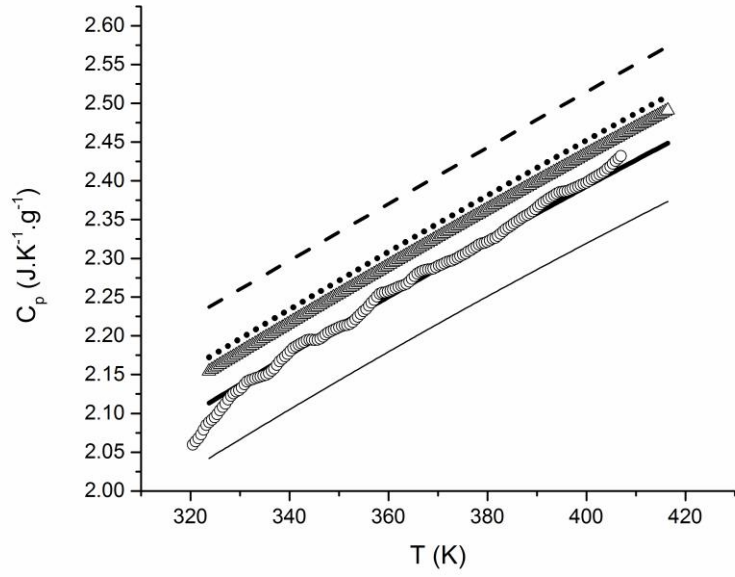
Table 5-9: RMSD and Bias value for PLS, ANN, and API models for the test data set

	RMSD	Bias
Equation 5-2	0.014	-0.0018
Equation 5-3	0.013	-0.0027
ANN	0.011	0.004
API	0.020	-0.016

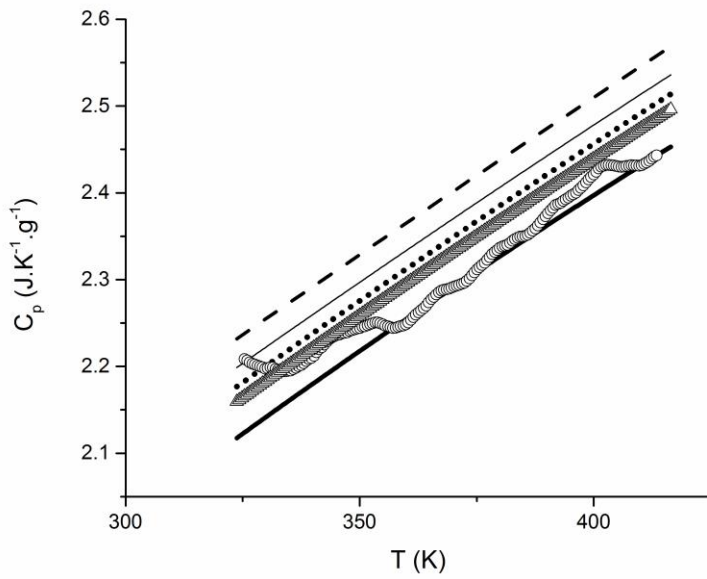
The models developed by PLS, ANN and API are applied to predict the similarity variable of the sample mixtures (Table 4-1) prepared in Chapter 4 for the experiments. In this comparison the heat capacity values predicted by Dadgostar-Shaw equation by the similarity variable estimated by API, PLS and ANN are illustrated in Figure 5-13.



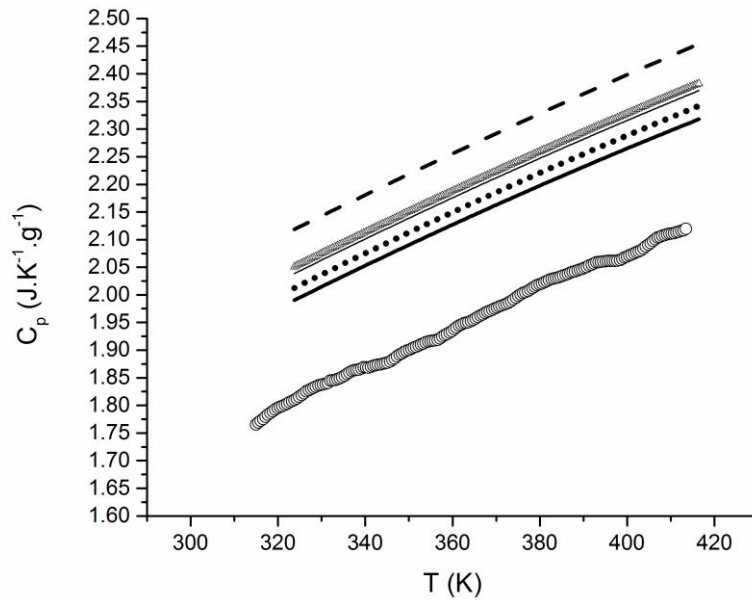
(a)



(b)



(c)



(d)

Figure 5-13: Experimental  $C_p$ ,  $\alpha$ , and  $C_p$  predicted by the Dadgostar-Shaw correlation with similarity variable estimated by; - -, API procedure; . . ., ANN; —, PLS (Equation 5-2);  $\Delta$ , PLS (Equation 5-3) and —, actual  $\alpha$  for mixtures a) 1, b) 2, c) 3, d) 4 (see Table 4-1 for compositions).

The mean absolute percentage error (MAPE) between the Dadgostar-Shaw correlation final results using experimental and computed values for  $\alpha$  is shown in Table 5-8.

Table 5-8: The mean absolute percentage error (MAPE) between the predicted and experimental heat capacities based on different methods for calculating  $\alpha$

	Dadgostar-Shaw correlation				
	API $\alpha$	Equation 5-2	Equation 5-3	ANN	Actual $\alpha$
Mixture 1	0.43	0.51	1.28	1.24	0.43
Mixture 2	6.31	3.25	1.88	2.19	2.69
Mixture 3	3.71	2.26	1.11	1.36	4.10
Mixture 4	16.89	12.12	14.75	12.89	13.24
Lastovka-Shaw correlation + APR based departure function for petroleum cut					
Mixture 1	2.61	2.43	3.49	3.49	2.61
Mixture 2	3.75	3.10	1.61	3.21	4.11
Mixture 3	4.12	4.00	1.79	5.14	4.11
Mixture 4	10.00	7.97	15.50	6.93	2.21

## 5.5 Conclusion

Two approaches are developed for estimation of similarity variable. One is correlative and based on partial least square fits to functions of density and MW, and density and boiling point. The other approach is based on an Artificial Neural Net with density and MW as the required inputs. Both approaches provide better performance than the API based method described in Chapter 2 with respect to the absolute error and the bias of estimates. The estimated heat capacity values of the mixtures based on various methods for estimating  $\alpha$  illustrate that both ANN, and PLS approaches yield nearly equivalent outcomes. The advantage of a correlation over an ANN is, it is also practical for hand calculation.

Since the difference between the performance of PLS and ANN with {density and  $T_b$ } and {density and MW} as the required inputs is not great, in cases where the MW value is not available, one can calculate the  $\alpha$  value based on PLS and ANN with  $T_b$  and density as the required inputs.

## 6 Overall Conclusion and Future Work

### 6.1 Conclusion and Recommendation

The following conclusion can be drawn:

1. There are many available methods for calculating isobaric liquid heat capacity, and each has some advantages and disadvantages and different required inputs. It was shown in this work that choosing a wrong method can result in a deviation of  $\pm 0.4 \text{ J.K}^{-1}.\text{g}^{-1}$  in estimated liquid  $C_p$  which can have both capital and operating cost consequences for processes designed using them.
2. For known compounds and molecularly defined mixtures, the best method for calculating isobaric  $C_p$  of liquids is the ideal gas heat capacity plus an equation of state based departure function. For petroleum cuts with known critical properties, the ideal gas  $C_p$  estimation by Lastovka-Shaw correlation is more accurate than the Lee-Kesler correlation. These methods cannot be used for ill-defined fluids, since the critical properties are not known. For ill-defined fluids, two methods are available to estimate liquid  $C_p$ : the Lee-Kesler correlation and the Dadgostar-Shaw correlation. The Dadgostar-Shaw correlation is more accurate and precise, and substantially so in almost all cases.
3. The similarity variable  $\alpha$  must be estimated in order to implement the element based correlations (both the Dadgostar-Shaw and Lastovka-Shaw correlations) into process simulation software. The API based  $\alpha$  estimation technique is shown to be poor. Large and biased errors arise, leading to significant over estimation of liquid phase heat capacities.

4. Two approaches were developed for estimating  $\alpha$  value. One is a correlation based on partial least square and the other one is a trained neural network. Both methods show better performance than the API based method, from a deviation and a bias perspective. The outcomes from the PLS and ANN approaches are nearly equivalent.

## **6.2 Future Work**

1. Modifying the code in VMGSIM to accommodate the correlations for  $\alpha$ .
2. providing a decision tree to users of the element based and other correlations for liquid phase heat capacity to ensure the best option is selected.

## Reference

- [1] “Unconventional oil” [Online]. Available: [http://en.wikipedia.org/wiki/Unconventional\\_oil](http://en.wikipedia.org/wiki/Unconventional_oil); Accessed: July 2013
- [2] S. A. Holditch, “The Increasing Role of Unconventional Reservoirs”. *Management*, no. November, pp. 34–38, 2003.
- [3] “OPEC share of world crude oil reservoir 2011” [Online]. Available: [http://www.opec.org/opec\\_web/en/data\\_graphs/330.htm](http://www.opec.org/opec_web/en/data_graphs/330.htm); Accessed: July 2013
- [4] "[Canadian Energy Overview 2011](https://www.neb-one.gc.ca/clf-nsi/rnrgynfmtn/nrgyrprt/nrgyvrvw/cndnrgyvrvw2011/cndnrgyvrvw2011-eng.pdf)" [Online]. Available: <https://www.neb-one.gc.ca/clf-nsi/rnrgynfmtn/nrgyrprt/nrgyvrvw/cndnrgyvrvw2011/cndnrgyvrvw2011-eng.pdf>. Accessed: July 2013
- [5] J. H. Masliyah, J. Czarnecki, Z. Xu, *Handbook on Theory and Practice of Bitumen Recovery from Athabasca Oil Sands*. Canada: Kingsley Knowledge, pp. 1–18, 2011.
- [6] B. I. Lee and M. G. Kesler, “Improve prediction of enthalpy of fractions” *Hydrocarbon Processing*, no. March, pp. 153–158, 1976.
- [7] N. Dadgosta and J. M. Shaw, “A predictive correlation for the constant-pressure specific heat capacity of pure and ill-defined liquid hydrocarbons,” *Fluid Phase Equilibria*, vol. 313, pp. 211–226, 2011.
- [8] N. Dadgostar and J. M. Shaw, “A predictive correlation for the constant-pressure specific heat capacity of pure and ill-defined liquid hydrocarbons –including the critical point,” *Fluid Phase Equilibria*, vol. 344, pp. 139 – 151, 2013.
- [9] V. Lastovka, N. Sallamie, and J. M. Shaw, *Fluid Phase Equilibria*, vol. 268, pp. 51–60, 2008.
- [10] V. Lastovka and J. M. Shaw, “Predictive correlations for ideal gas heat capacities of pure hydrocarbons and petroleum fractions.” *Fluid Phase Equilibria*, Accepted manuscript, July 2013.
- [11] Chung-Nin Mak, P. "Thermodynamic Properties from Cubic Equation of State". Master Thesis at UBC, 1988.
- [12] J.M. Smith, H.C. Van Ness, M.M. Abbott “Introduction to Chemical Engineering Thermodynamics”, sixth edition. McGraw-Hill, 2001
- [13] M. Záborský, V. Růžička, and E. S. Domalski, “Heat Capacity of Liquids: Critical Review and Recommended Values. Supplement I *J. Phys. Chem. Ref. Data* 30, 1199 ,2001.



- [14] D. Bessières and J. Daridon, "Measurement and calculation of heat capacity of heavy distillation cuts under pressure up to 40 MPa" *Thermal Analysis and Calorimetry*, vol. 58, pp. 39–49, 1999.
- [15] *Technical data book-petroleum refining*. American Petroleum Institute, p. 7D1\_7D3, 1997.
- [16] B. E. Poling, J. M. Prausnitz, and J. P. O Connell, *The properties of gases and liquids*, Fifth. McGRAW HILL, 2001.
- [17] K. M. Watson, "Thermodynamics of the Liquid State," *Industrial & Engineering Chemistry*, vol. 35, no. 4, pp. 398–406, 1943.
- [18] A. I. Johnson and C. J. Huang, "Estimation of heat capacity of organic liquids", *Canadian J. of Technology*, vol. 33, p. 421 1955.
- [19] C. F. Chueh and A. C. Swanson, "Estimation of liquid heat capacity," *Canadian Journal of Chemical Engineering*, vol. 5, no. 51, pp. 596–600, 1973.
- [20] S. W. Benson and J. H. Buss, "Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties" *J. Chem. Phys.*, 29, p. 546, 1958.
- [21] S. W. Benson, M. N. Cruickshank et al, "Additivity rules for the estimation of thermochemical properties" *Chem. Rev.* 69, p. 279, 1969.
- [22] V Růžička, E. S. Domalski, "Estimation of the Heat Capacities of Organic Liquids as a Function of Temperature using Group Additivity. I. Hydrocarbon Compounds" *J. Phys. Chem. Ref. Data* 22, p. 597, 1993.
- [23] V Růžička, E. S. Domalski, "Estimation of the heat capacities of organic liquids as a function of temperature using group additivity. II. Compounds of Carbon, Hydrogen, Halogens, Nitrogen, Oxygen, and Sulfur" *J. Phys. Chem. Ref. Data* 22, p. 619, 1993.
- [24] M. Zábanský and V. Ruzicka, "Estimation of the Heat Capacities of Organic Liquids as a Function of Temperature Using Group Additivity: An Amendment," *Journal of Physical and Chemical Reference Data*, vol. 33, no. 4, p. 1071, 2004.
- [25] K. P. Tyagi, "Estimation of Saturated Liquid Heat Capacity," *Ind. Eng. Chem. Process. Des.*, vol. 14, no. 4, pp 484-488, 1975.
- [26] J. Rowlinson, *Liquids and Liquid Mixtures*. Landan: Butterworth, 1969.
- [27] A. Bondy, "Estimation of the heat capacity of liquids," *Industrial & Engineering Chemistry Fundamentals*, vol. 4, pp. 442–449, 1966.

- [28] B. I. Lee and M. G. Kesler, "A Generalized Thermodynamic Correlation Based on Three-Parameter Corresponding States," *AIChE Journal*, vol. 21, no. 3, pp 510-527, 1975.
- [29] R. C. Reid and J. E. Sobel, "Estimation of Saturated Liquid Heat Capacities above Boiling Point," *Industrial & Engineering Chemistry Fundamentals*, vol. 4, no. 3, pp. 328–331, 1965.
- [30] B. Lee and W. C. Edmister, "Fugacity Coefficients and Isothermal Enthalpy Differences for Pure Hydrocarbon Liquids," *Ind. Eng. Chem. Fundamen.*, 10 (2), pp 229–236 1971
- [31] W. F. Stevens and G. Thodos, "Estimation of Enthalpies: Multicomponent Hydrocarbon Mixtures at Their Saturated Vapor and Liquid States," *AIChE Journal*, vol. 9, no. 3, pp. 293–296, 2004.
- [32] D. N. Rihani and L. K. Doraiswamy, "Estimation of heat capacity of organic compounds from group contributions," *Ind. Eng. Chem. Fundamen.*, vol. 4, no. 1, pp. 17–21, 1965.
- [33] D.-Y. Peng and D. B. Robinson, "A New Two-Constant Equation of State," *Industrial & Engineering Chemistry Fundamentals*, vol. 15, no. 1, pp. 59–64, 1976.
- [34] A. Peneloux and E. Rauzy, "A consistent correlation for Redlich-Kwong-Soave volumes," *Fluid Phase Equilibria*, vol. 1, pp. 7–23, 1952.
- [35] G. Soave, "Equilibrium constants from a modified Redlich-Kwong equation of state," *Chemical Engineering Science*, vol. 27, no. 6, pp. 1197–1203, 1972.
- [36] S. Jullian, A. Barreau, E. Behar, and J. Vidal, "Application of the sbr equation of state of high molecular weight hydrocarbons," *Chemical Engineering Science*, vol. 4, pp. 1001–1004, 1989.
- [37] M. Benedict, "An Empirical Equation for Thermodynamic Properties of Light Hydrocarbons and Their Mixtures II. Mixtures of Methane, Ethane, Propane, and n-Butane," *The Journal of Chemical Physics*, vol. 10, no. 12, p. 747, 1942.
- [38] H. Nishiumi and S. Saito, "An improved generalized bwr equation of state applicable to low reduced," *Journal Of Chemical Engineering Of Japan*, no. 1967, pp. 356–360, 1975.
- [39] R. W. Hankinson and G. H. Thomson, "A new correlation for saturated densities of liquids and their mixtures," *AIChE Journal*, vol. 25, no. 4, pp. 653–663, 1979.
- [40] K. S. Pedersen, P. L. Christensen, and A. Fredenslund, "Viscosity of crude oils," *Chemical Engineering Science*, vol. 6, no. 39, pp. 1011–1016, 1984.
- [41] U. Plocker, H. Knapp, and J. Prausnitz, "Calculation of High-pressure Vapor-Liquid Equilibria from a Corresponding-States Correlation with Emphasis on Asymmetric Mixtures," *Journal of Ind. Eng. Chem. Process Des*, vol. 17, no. 3, pp. 324–332, 2008.

- [42] C. F. Spencer and R. P. Danner, "Improved equation for prediction of saturated liquid density," *Journal of Chemical & Engineering Data*, vol. 17, no. 2, pp. 236–241, 1972.
- [43] A. S. Teja, "A corresponding states equation for saturated liquid densities. I. Application to lng," *AIChE Journal*, vol. 3, no. 26, pp. 337–341, 1980.
- [44] J. Joffe, "Compressibilities of Gas Mixtures," *Industrial & Engineering Chemistry*, no. 39, p. 837, 1947.
- [45] S. Stamataki, D. Tassios, and E. H. Pressions, "Performance of cubic EOS at high pressures," *Journal of Oil & Gas Science and Technology* vol. 53, pp. 367–377, 1998.
- [46] <http://virtualmaterials.com/vmgsim>, VMGSim [computer software], Calgary: Alberta, Virtual Materials Group Inc.
- [47] V. M. G. Inc, "New Ideal Gas Heat Capacity Estimation Methods for Oil Fractions Lastovka-Shaw IG Cp Estimation Method Dadgostar-Shaw IG Cp Estimation Method Implementation in VMGSim," pp. 1–8, 2011.
- [48] W. S. McCulloch and W. H. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, Volume 5, Issue 4, pp 115-133 1943.
- [49] S. A. Kalogirou, *Artificial intelligence for the modeling and control of combustion processes: a review*, *Progress in Energy and Comb. Science*, vol. 29, no. 6. 2003, pp. 515–566.
- [50] A. Sözen, E. Arcakliog˘lu, and M. Özalp, "Formulation based on artificial neural network of thermodynamic properties of ozone friendly refrigerant/absorbent couples," *Applied Thermal Engineering*, vol. 25, no. 11–12, pp. 1808–1820, 2005.
- [51] S. A. Kalogirou, "Applications of artificial neural-networks for energy systems," *Applied Energy Journal*, vol. 67, pp. 17–35, 2000.
- [52] S. Hykin, *Neural networks: a comprehensive foundation*. New York: Macmillan, 1994.
- [53] C. Neocleous, "A neural network architecture composed of adaptively defined dissimilar single-neurons," Brunel University, UK, 1998.
- [54] Pj. Werbos, "New tools for prediction and analysis in the behavioral science," Harvard University, Cambridge, 1974.
- [55] P. R. C. Avron Barr, Edward A. Feigenbaum, *The handbook of artificial intelligence*. Los Altos, CA: Morgan Kaufmann, 1982, p. volume 1.

- [56] A. Chouai, S. Laugier, and D. Richon, "Modeling of thermodynamic properties using neural networks Application to refrigerants," *Fluid Phase Equilibria*, vol. 199, pp. 53–62, 2002.
- [57] M. Mandischer, "Comparison of Neural Networks , Evolutionary Techniques and Thermodynamic Group Contribution Methods for the Prediction of Heats of Vaporization," Universität Dortmund, pp. 1–19, 2001.
- [58] M. Mohandes, S. Rehman, T.O. Halawani, "Estimation of global solar radiation using artificial neural networks," *Renewable energy*, vol. 14, pp. 179–184, 1998.
- [59] S. Sheikh and M. Bagherpour, "Estimating the Saturation Thermodynamic Properties of Propene Using a Feed Forward Neural Network," *World Applied Science Journal*, vol. 4, no. 2, pp. 169–173, 2008.
- [60] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM: Philadelphia, p.p 1-34, 1996.
- [61] C. R. Rao, H. Toutenburg, Shalabh, C. Heumann, and M. Schomaker, *Linear Models: Least Squares and Alternative*. Second Edition, Springer Series in Statistics, Springer-Verlag: New York, 1999.
- [62] R. D. Tobias, "An Introduction to Partial Least Squares Regression Example : Spectrometric Calibration." SAS Institute Inc., Cary, NC, pp 1-8: <http://support.sas.com/techsup/technote/ts509.pdf>. Accessed: July 2013.
- [63] G. W. H. Höhne, W.F. Hemminger, H.J.Flammersheim, *Differential Scanning Calorimetry*, Springer, Second Ed.,pp. 1-50, 2003.
- [64] K. Q. Tran, "Reversing and non-reversing phase transitions in athabasca bitumen," MSc thesis, University of Alberta, Canada, 2009.
- [65] G. W. H. Höhne, *Problems with the calibration of differential-temperature scanning-calorimeters.*, *Acta: Thermochemica*, 69 (1-2), pp. 175–197, 1983.
- [66] R. Marini, A. Berbenni, V. Flor, G. Massarotti, and V. Riccardi, "An analysis of the factors affecting the peak shape and the quantitative reliability of a heat flux DSC cell," *Thermochemica Acta*, vol. 95, no. 2, pp. 419–424.
- [67] X. Wu, J. Chen, and X. Zeng, "The application of DSC in identification of LDPE / LLDPE blends mulching film," *Angew. Makromol. Chem.*, vol. 189, no. 3197, pp. 183–193, 1991.
- [68] C. Farkas, J. Mohacsi-Farkas, "Application of differential scanning calorimetry in food research and food quality assurance," *Journal of Thermal Analysis*, vol. 47, pp. 1787–1803, 1996.

- [69] J.H. Ferrasse and D. Lecomte, “Simultaneous heat-flow differential calorimetry and thermogravimetry for fast determination of sorption isotherms and heat of sorption in environmental or food engineering,” *Chemical Engineering Science*, vol. 59, no. 6, pp. 1365–1376, 2004.
- [70] D. M. Price, “Temperature calibration of differential scanning calorimeters,” *Journal of Thermal Analysis*, vol. 45, no. 6, pp. 1285–1296, 1995.
- [71] S. M. Sarge, E. Gmelin, G. W. H. Höhne, H. K. Cammenga, W. Hemminger, and W. Eysel, “The caloric calibration of scanning calorimeters,” *Thermochimica Acta*, vol. 247, pp. 129–168, 1994.
- [72] E. Gmelin and St. M. Sarge, “Calibration of differential scanning calorimetry,” *Pure & Appl. Chem.*, Vol. 67, No. 11, pp. 1789-1800, 1995.
- [73] H. K. Cammenga, S. M. Sarge, W. Eysel, and E. Gmelin, “Temperature , heat and heat flow rate calibration of scanning calorimeters in the cooling mode ,” *Thermochimica Acta*, vol. 361, 2000.
- [74] C. Sabbah ,R., A. Xu-wu, J. S. Chickos, M. L. P. Leitão, M. V. Roux , and L. a. Torres, “Reference materials for calorimetry and differential thermal analysis,” *Thermochimica Acta*, vol. 331, no. 2, pp. 93–204, 1999.
- [75] “International Confederation for Thermal Analysis and Calorimetry” [Online]. Available: <http://www.ictac.org/>; Accessed: July 2013
- [76] H. L. Finke, J. F. Messerly, D. R. Douslin, S. H. Lee, and A. G. Osborn, “Comprehensive thermodynamic studies of seven aromatic hydrocarbons”, *The Journal of Chemical Thermodynamics*, vol 9, pp. 937–956, 1977.
- [77] R. Chirico, S. Knipmeyer, and W. Steele, “Heat capacities, enthalpy increments, and derived thermodynamic functions for naphthalene between the temperatures 5K and 440K,” *The Journal of Chemical Thermodynamics*, vol. 34, no. 11, pp. 1873–1884, 2002.
- [78] D. G. Archer, “Thermodynamic Properties of Synthetic Sapphire ( $\alpha$ -Al<sub>2</sub>O<sub>3</sub>), Standard Reference Material 720 and the Effect of Temperature-Scale Differences on Thermodynamic Properties” *J. Phys. Chem. Ref. Data* 22, p. 1441, (1993).
- [79] A. Diedrichs and J. Gmehling, “Measurement of heat capacities of ionic liquids by differential scanning calorimetry,” *Fluid Phase Equilibria*, vol. 244, no. 1, pp. 68–77, 2006.
- [80] J. Little, C. Moler, 1984, MATLAB R2012[computer software], Natick, Massachusetts, U.S.A.
- [81] F. E. Croxton, D. J. Cowden; *Applied General Statistics*. Pitman, p. page 625, 1968.

- [82] H.Y. Afeefy, J.F. Liebman, and S.E. Stein, "Neutral Thermochemical Data" in NIST Chemistry WebBook, NIST Standard Reference Database Number 69, Eds. P.J. Linstrom and W.G. Mallard, National Institute of Standards and Technology, Gaithersburg MD, 20899, <http://webbook.nist.gov>, (retrieved August 28, 2013).
- [83] R. H. Shumway, D. S. Stoffer, Time series analysis and its application, Springer, 2008.
- [84] K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares". *Quarterly of Applied Mathematics* **2**: 164–168, 1944.
- [85] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors" *Nature* **323**, p: 533 – 536, 1986.
- [86] P. F. Verhulst, "[Notice sur la loi que la population poursuit dans son accroissement](#)", *Correspondance mathématique et physique* **10**: p, 113–121, 1838.
- [87] N. Andrei, "Scaled conjugate gradient algorithms for unconstrained optimization", *Comput Optim Appl* , 38: p:401–416, 2007.
- [88] L. Grippo, S. Lucidi, "A globally convergent version of the Polak-Ribière conjugate gradient method", *Mathematical Programming*, Volume 78, Issue 3, pp 375-391, 1997.
- [89] "Purelin" [Online]. Available: <http://www.mathworks.com/help/nnet/ref/purelin.html>; Accessed: August 2013
- [90] M. Dorofki, A. H. Elshafie et al, "Comparison of Artificial Neural Network Transfer Functions Abilities to Simulate Extreme Runoff Data" *International Conference on Environment, Energy and Biotechnology*, vol 33, pp: 30-44, 2012.
- [91] M. D. Buhmann, "Radial basis functions: theory and implementations", University of Giessen, Cambridge University Press, 2003.
- [92] "Reibas" [Online]. Available: <http://www.mathworks.com/help/nnet/ref/tribas.html>; Accessed: August 2013
- [93] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *International Journal of Forecasting*, vol. 22, no. 4, pp. 679–688, Oct. 2006.
- [94] J. S. Armstrong and F. Collopy, "Error measures for generalizing about forecasting methods: Empirical comparisons," *International Journal of Forecasting*, vol. 8, no. 1, pp. 69–80, Jun. 1992.
- [95] Black, C. and Twu, C.H., AIChE Spring Meeting, Houston, March 1983
- [96] API procedure 4D3.1, API Technical Data Book - Petroleum Refining, Refining Department, 5th Edition, May 1992, American Petroleum Institute

## Appendix A: Matlab Code for the PLS based Correlation and trained ANN

This Appendix presents the MATLAB code for all the programs developed in the partial least square based correlation and trained artificial neural net

### PLS based Correlation

```
Datap = xlsread('sep data.xlsx','Sheet1','C2:F155');

alpha=Datap(:,4)';
Density=Datap(:,2)';
MW=Datap(:,1)';
Tb=Datap(:,3)';

Xc=[Density'];
yc=alpha';

comp=1;

[XL,y1,XS,YS,beta,PCTVAR,MSE,stats] = plsregress(Xc,yc,comp);
yfit = [ones(size(Xc,1),1) Xc]*beta;

figure(2)
plot(yc,yfit,'o')

TSS = sum((yc-mean(yc)).^2);
RSS = sum((yc-yfit).^2);
Rsquared = 1 - RSS/TSS;

figure(5)
plot(yc)
hold on
plot(yfit,'r')
legend('Real Output','Predicted Output')

n=length(alpha);
RMESP=sqrt(sumsqr(yfit-yc)/n)
coe=corr(yfit,yc)
```

## Trained ANN

```
Datap = xlsread('API Elemental Analysis for Mixtures.xlsx','Training data set','AM5:AP158');
Datat= xlsread('API Elemental Analysis for Mixtures.xlsx','Sin fitet curve','D38:F41');
Datat1= xlsread('API Elemental Analysis for Mixtures.xlsx','Sin fitet curve','D10:E35');

Densityt=Datat(:,2)';
MWt=Datat(:,1)';
Tbt=Datat(:,3)';
Densityt1=Datat1(:,2)';
MWt1=Datat1(:,1)';
alpha=Datap(:,2)';
Density=Datap(:,3)';
MW=Datap(:,1)';
Tb=Datap(:,4)';

Test1=[MWt;Densityt];
Test2=[MWt1;Densityt1];
inputs=[MW;Density];
targets=[alpha];

hiddenLayerSize = 5;
net = fitnet(hiddenLayerSize);

net.inputs{1}.processFcns = {'removeconstantrows','mapminmax'};
net.outputs{2}.processFcns = {'removeconstantrows','mapminmax'};

net.divideFcn = 'dividerand'; % Divide data randomly
net.divideMode = 'sample'; % Divide up every sample
net.divideParam.trainRatio = 70/100;
net.divideParam.valRatio = 15/100;
net.divideParam.testRatio = 15/100;

net.trainFcn = 'trainlm';
net.layers{1}.transferFcn = 'tribas';

net.performFcn = 'mse'; % Mean squared error

net.plotFcns = {'plotperform','plottrainstate','ploterrhist', ...
'plotregression', 'tribas'};

[net,tr]=train(net,inputs,targets);

outputs = net(inputs);
errors = gsubtract(targets,outputs);
performance = perform(net,targets,outputs);

trainTargets = targets .* tr.trainMask{1};
valTargets = targets .* tr.valMask{1};
testTargets = targets .* tr.testMask{1};
```



```
trainPerformance = perform(net,trainTargets,outputs)
valPerformance = perform(net,valTargets,outputs)
testPerformance = perform(net,testTargets,outputs);
```

```
view(net)
```

```
i=[1:length(targets)]
j=[1:length(MWt)]
```

```
output=sim(net,inputs(:,i))
OPT=sim(net,Test1(:,j))
targets(:,i)
```

```
MAE=sum(abs(targets-output))/length(targets)
MSE=sum((targets-output).^2)/length(targets)
```

```
TSS = sum((targets-mean(targets)).^2);
RSS = sum((targets-output).^2);
Rsquared = 1 - RSS/TSS;
```

```
n=length(alpha);
RMESP=sqrt(sumsqr(output-targets)/n)
```