UNIVERSITI
TEKNOLOGI
PETRONAS

DEVELOPMENT OF COMPOSITIONAL MODEL FOR PREDICTING VISCOSITY OF CRUDE OILS USING POLYNOMIAL NEURAL NETWORKS (PNN) INDUCED BY GROUP METHOD OF DATA HANDLING (GMDH)

By

Yong Wen Pin

Dissertation submitted in partial fulfillment of the requirements for the
Bachelor of Engineering (Hons) of Petroleum Engineering

August 2011

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan.

CERTIFICATION OF APPROVAL

Development of Compositional Model for
Predicting Viscosity of Crude Oils using Polynomial Neural Networks (PNN)
induced by Group Method of Data Handling (GMDH)

By
Yong Wen Pin

Dissertation submitted in partial fulfillment of the requirements for the
Bachelor of Engineering (Hons)
of Petroleum Engineering

August 2011

Approved by,

(Mr. Ali F. Mangi Alta'ee)

Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan.

CERTIFICATION OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this project, that the original work is my own except as specified in the references and acknowledgements, and that the original work contained herein have not been undertaken or done by unspecified sources or persons.

Yong Wen Pin

Petroleum Engineering & Geoscience Department
Universiti Teknologi PETRONAS
Bandar Seri Iskandar
31750 Tronoh
Perak Darul Ridzuan.

ABSTRACT

Viscosity or the internal resistance of the fluids to flow is the most important transport property that controls and influences the flow of oil through porous media and pipes. Accurate predictions of reservoir fluids are required in equation of state (EOS) based reservoir simulators. Due to time and money spent of experimental viscosity measurements, reliable viscosity models are developed for predicting crude oils viscosity. Throughout the years, although many of the common correlations were developed, laboratory measurements still cannot be replaced due to the complexities, varied composition and reservoir characteristics difference from different reservoirs. This study estimates crude oil viscosity by using a group method of data handling (GMDH) based on polynomial neural network (PNN). GMDH is an inductive algorithm for computer-based mathematical modeling using neural network with active neurons that optimizes model coefficients for predetermine mathematical equation and selects the optimal model complexity. The new model was built and tested using experimental measurements collected through literature search. The database consists of crude oils composition, viscosity, temperature and pressure from Middle East, North Sea and the others. Overall, the proposed model improved the prediction as compared to other viscosity model.

TABLE OF CONTENTS

LIST OF FIGURES

. LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background of Study

Viscosity is defined as the internal resistance of fluid to flow. It is expressed in deformation equation by

$$\tau = \mu\dot{\gamma} \qquad (1)$$

where τ is shear stress, $\dot{y}$ is the shear rate and μ is the viscosity. Refer to Figure 1, τ can be defined as F/A where F is the required force to keep the upper plate to move at constant velocity v in the x-direction and A is the area of plate in contact with the fluid. Therefore, fluid viscosity is the component when the force is transmitted through the fluid to the lower plate in a way that x-component of the fluid velocity depends linearly on the distance from the lower plate.

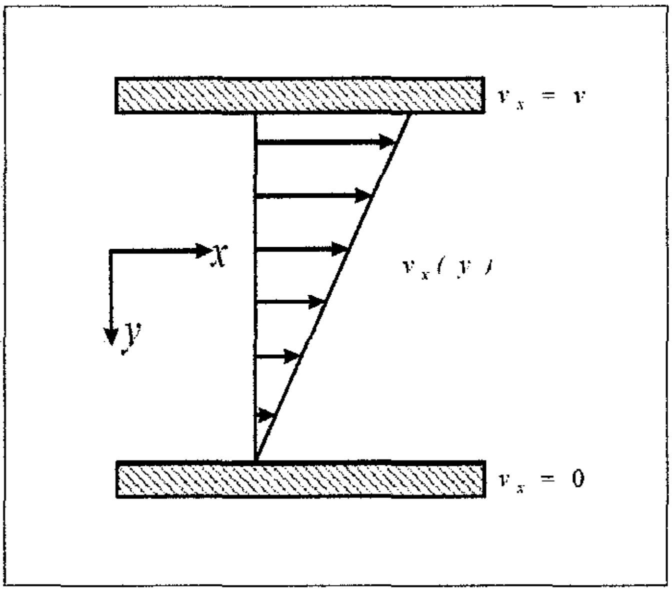


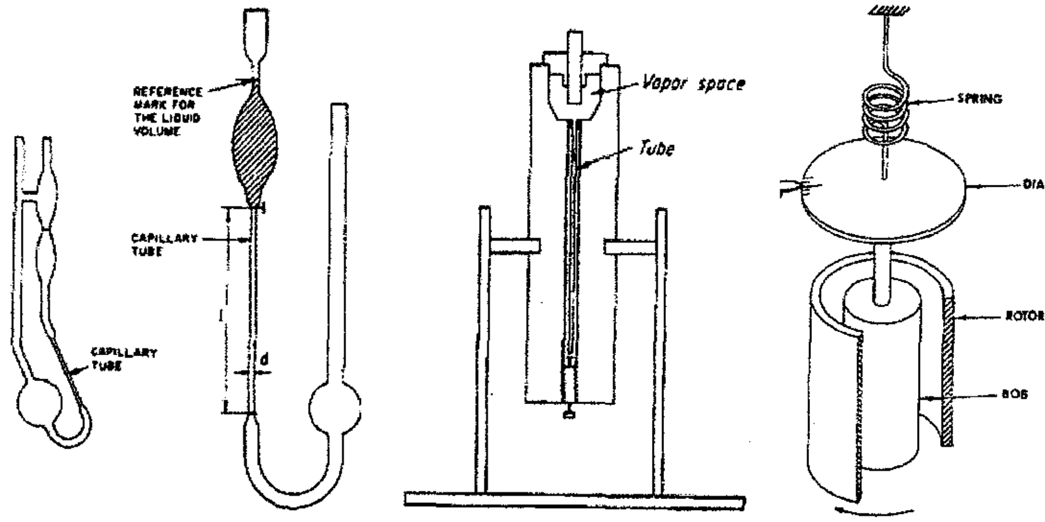
Figure 1: Steady-state velocity profile of fluid entrained between two flat surfaces [1]

In oil and gas industry, viscosity is generally expressed in terms of dynamic viscosity (μ) and kinematic viscosity (υ). As mentioned earlier, dynamic viscosity (the viscosity of liquid) is the ratio between the applied shear stress and the rate of shear. Meanwhile, kinematic viscosity is the ratio μ/ρ where ρ is fluid density. Table 1 below is the unit and dimensions of these terms.

	Symbol	cgs unit	SI unit	Dimension
Kinematic viscosity, υ	1 mm^2/s = 1 cSt 1 $m^2/s = 10^6$ cSt	cm^2/s	m^2/s	L^2/T
Dynamic viscosity, μ	1 $Dyn.s/cm^2$ = 100 cp 1 $Newton.s/m^2 = 10^3$ cp	$Dyn.s/cm^2$	$Newton.s/m^2$ (= $Pa.s$)	M/LT (FT/L^2)

Table 1: Unit and dimensions of viscosity [1]

Viscosity of liquids is measured by viscosimeter or viscometer. Three type of viscometer will be discussed here. The first viscometer is Ostwald viscometer (Figure 2a). In this viscometer, the viscosity is deduced from the comparison of the times required for a given capillary tube under specified initial head conditions with constant temperature. The second viscometer, falling (rolling) ball viscometer (Figure 2b) is based on Stoke's law for a sphere falling in a fluid under effect of gravity by dropping a polished steel ball into a glass tube of a somewhat larger diameter containing the liquid and recording the time required for the ball to fall at constant velocity through a specified distance between the reference marks. If the fluid flow in the tube remains in laminar range, the measurement results will be good. For non-Newtonian fluids, rotational type viscometer (Figure 2c) that consists of two concentric cylinders with annulus containing the liquid that viscosity is to be measured is often used. Both the outer cylinder or inner cylinder is rotated at a constant speed and the rotational deflection of the cylinder becomes a measure of the liquid's viscosity. [1]



(a) (b) (c)

Figure 2: Schematic diagram of viscometer [1]

Regardless whether the reservoir fluids are transmitting in subsurface or at surface, viscosity is important in designing pipelines, production equipment, well testing and reservoir simulation. In terms of crude oil viscosity, it ranges from 0.1cp for near critical to over 100cp for heavy oil and it is considered the most difficult oil property to calculate with a reasonable accuracy from correlations [2].

Reservoir oil viscosity is measured in laboratory using subsurface and surface samples. Whenever possible, oil viscosity should be determined by laboratory measurements at reservoir temperature and pressure although it is expensive to carry out PVT analysis at reservoir conditions. Sometimes, the measured surface oil viscosity from wellhead samples can be used as conversion to subsurface value but the accuracy of the results is challenged. If the laboratory data are not available, the engineers may refer to the published correlations.

The viscosity of crude oils can be classified into three groups in terms of their pressure, p:

- Dead oil viscosity, μ_{od}, the viscosity of crude oil at atmospheric pressure and system temperature.
- Saturated oil viscosity, μ_{ob}, the viscosity of crude oil at any pressure less than or equal to the bubble-point pressure.
- Undersaturated oil viscosity, μ_o, the viscosity of crude oil at a pressure above bubble point and reservoir temperature.

During production, the pressure of produced crude oils will change as they travel along the wellbore to the surface. At atmospheric pressure and reservoir temperature, there is no dissolved gas so the oil has the highest viscosity value of μ_{od}. As the pressure increases, gas solubility increases, resulting decrease in oil viscosity. When it reaches the bubble point pressure, the amount of gas in solution reaches its maximum R_{sb} while the oil viscosity reaches its minimum μ_{ob}. After the bubble point pressure, the increase of pressure will result the increase of viscosity of undersaturated crude oil due to the compression of the oil. Figure 2 showed the relationship between R_s and p with crude oil viscosity.

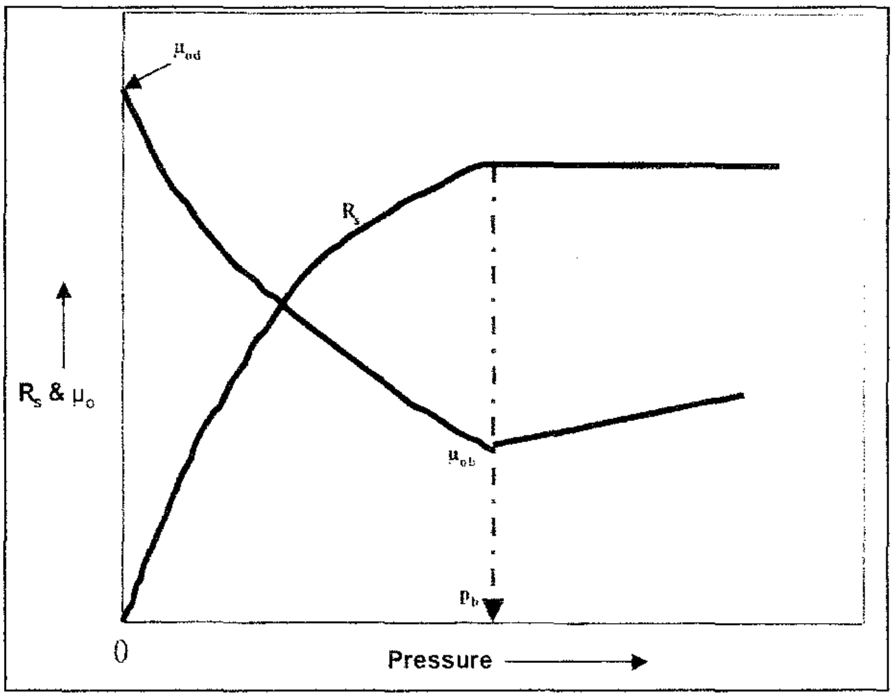


Figure 3: Crude oil viscosity as a function of R_s and p [2]

Depends on the available oil mixture data, correlations generally can be divided into correlations based on other measured PVT data like API and R_s and correlations based on oil composition. In general, there are two types of reservoir fluid model namely the black oil model and composition model. Black oil model consists of two components (solution gas and stock tank oil). The calculations are based on empirical correlations from experiment using B_o, R_s and other measured fluid properties. For compositional model, N components based on paraffin series are needed and the calculations are based on equation of state (EOS) using feed forward calculation of fluid properties. In this case, the crude oil viscosity calculation is based on the availability of the fluid properties or crude oil composition.

1.2 Problem Statement

A direct viscosity measurement in reservoir condition the most ideal case but since this method is costly. In common practices, the fluid properties from the subsurface samples are measured in laboratory in surface condition under certain calibration and assumptions

On the other hand, sometimes due to the unavailability of reservoir data and the cost of laboratory viscosity measurement at different time, viscosity values are estimated based on the publish correlations. Many researchers are improving the prediction of gas and liquid viscosities but most of the methods from mathematically rigorous to completely empirical forms have the following limitations:

1. Applied for either gas phase or liquid phase
2. Applied within a limited range of composition, temperature and pressure
3. Reservoir fluids at different part of the world exhibit different behaviour
4. Separate density correlation is required when density involved in estimation
5. Inconsistent of prediction (different API might exhibit same viscosity value)

Since viscosity plays an important role in providing a better understanding on future development planning, a better viscosity model has to be proposed to eliminate as many limitations mentioned above to estimate the crude oil viscosity as close as the measured value.

1.3 Objectives & Scope of Study

The objective of this research is to propose a new compositional model for crude oil viscosity prediction using Polynomial Neural Networks (PNN) induced by Group Method of Data Handling (GMDH). This study is focused on the compositional crude oil viscosity model using a specific ranges of pressure, temperature, crude oil composition from methane until heptanes plus, specific gravity and molecular weight of heptanes plus.

In this project, crude oil PVT analysis composition data will be collected and filtered. Then, the data will be the input in GMDH software to develop the compositional model with a few trails using polynomial neural networks (PNN). Next, the results will be compared to other published correlations or models. The following are the scopes of study:

(1) Investigating the accuracy of other published viscosity models for predicting viscosity of Middle East, North Sea and the others crude oils from literature review

(2) Utilizing Polynomial Neural Networks (PNN) induced by Group Method of Data Handling (GMDH) in developing compositional model

(3) Comparing proposed compositional model with other models from literature

CHAPTER 2

LITERATURE REVIEW

In most cases, there is no simple rule or theory to predict the rate of change of crude oil viscosity with respect to temperature and pressure or composition. Due to the complexity of reservoir fluids, their composition will change during depletion of reservoir at different pressure and make it impossible to obtain a constant viscosity value that measured isothermally at the initial composition of reservoir fluids. Therefore, as mentioned by Adel M. Elsharkawy *et al.* [4], the viscosity experimental data are used to evaluate and calibrate the proposed compositional model and the compositional model is used to update the value of the viscosity as a function of the changing pressure, temperature and composition. Basically, viscosity models can be classified into: (i) empirical methods; (ii) corresponding state methods; (iii) equation of state (EOS) based viscosity models.

2.1 Empirical Correlations

Empirical models cover all kinds of crude oils namely dead, saturated and undersaturated crudes. The commonly used empirical models in the industry are the models developed by Beggs and Robinson (1975) [5], Labedi (1992) [6], Kartoatmodjo and Schmidt (1994) [7], and Elsharkawy and Alikhan (1999) [8]. In these models, crude oil viscosity is a strong function of temperature (T), pressure (P), separated gas gravity (γ_g), and tank oil gravity (API). The dead oil viscosity (μ_{od}) is a function of oil API gravity and reservoir temperature; saturated oil viscosity (μ_{ob}) was correlated to the dead oil viscosity and solution gas-oil ratio (R_s); above bubble point (no gas being dissolved into the oil), the pressure becomes the primary independent variable for predicting the viscosity of undersaturated oil (μ_o). A reminder that saturated oil viscosity is highly dependent on the gas-oil ratio, a function of the oil API gravity, gas gravity, reservoir pressure, and temperature.

2.2 Corresponding State Models

According to principle of corresponding states, the dimensionless properties of all fluids have the same numerical values at the same reduced conditions and it provides the basis for the development of correlations and estimation methods using knowledge of the properties of, at least, one reference fluid [4]. Initially, Hanley *et al.*[9] used the viscosity of methane as the reference fluid to correlate viscosity as a function of the temperature and density. Later, Pedersen and Fredenslund [10] proposed an improvement of the model by extending the method of Pedersen *et al.* to $T\mathrm{r} < 0.4$ using additional experimental viscosity data. All of these methods use the acentric factor as an interpolation parameter of the reference fluid properties. In 1991, Petersen *et al.*[11] proposed a method with the molecular weight as an interpolation parameter instead of the acentric factor using methane and *n*-decane as reference components but this model suitable for viscosity predictions of light petroleum fractions only. In 1995, Moharam and Fahim [12] used decane and eicosane as the reference fluids in predicting the viscosity of heavy crude oils and petroleum fractions of average molecular weight higher than 142 and the model relaxes the limitation imposed on reduced temperature $T\mathrm{r} > 0.476$ to $T\mathrm{r} > 0.4$.

2.3 EOS Based Viscosity Models

EOS based viscosity models applicable to both reservoir gases and liquids viscosity calculation at the desired temperature and pressure based on their composition from methane through heptane-plus, hydrogen sulfide, nitrogen, and carbon dioxide together with the molecular weight and specific gravity of the heptane-plus fractions. Lohrenz *et al.*[13] developed the first crude oil compositional viscosity prediction method based on the residual viscosity concept and the law of corresponding states. Later, Little and Kennedy [14] developed EOS based viscosity model using the van der Waals EOS but six coefficients involved in the model were not generalized. Both models above are very sensitive to the density, which is normally determined by a separate correlation and may be very inaccurate for high-viscosity fluids. Lawal [15] also proposed a viscosity model based on the four-parameter Lawal-Lake-Silberberg EOS but it is not a predictive model because specific constants are required for each substance and poor results were obtained when applied to the C_{7+} fraction containing reservoir oils.

2.4 Group Method of Data Handling (GMDH)

GMDH was originated in 1968 by Prof. A.G. Ivakhnenko [16, 17]. This modeling tool was successfully used in areas such as weather prediction, medical informatics, spectrum analysis and energy forecasting but still has not been much utilized in the oil industry and it has been proved successfully for predicting PVT properties of crude oils by E.A. Osman and R. E. Abdel-Aal. (2002) [18]. GMDH is well coordinated with the state of art data mining techniques. The central role in optimal complexity detection is given to data split into training and testing samples.

Basically, GMDH applies gradual model complication and self-organized model selection to select the optimally-complex mathematical model. Using Core algorithms, different model structures and a number of models are selected to predict testing data in best way. The training data mentioned before is used for fitting of model coefficients while testing data is used for measurement of model accuracy. The multilayered threshold algorithm of GMDH model is based on polynomial function generated from a given set of variables [19]. For example, a data set of two input (independent variables) x_1 and x_2 and output (dependent variable) y can form a quadratic polynomial function $y = a_0 + a_1*x_1 + a_2*x_2 + a_3*x_1*x_2 + a_4*x_1^2 + a_5*x_2^2$.

The working algorithm is presented in Figure 1. At the top, it is the input for each predictor variable. Each neuron in the second layers draws two of the input variables as its input and third layer neurons draw their inputs from two of the previous layer. The process continues until the final layer (bottom) draws two inputs from the previous layer and produces a single value as the output of the network. Each neuron in neural network induced by GMDH is called partial description (PD). During the process, non linear regression determines the coefficients of each polynomial on training data set while external selection criteria and test data set are used to select optimally complex models.

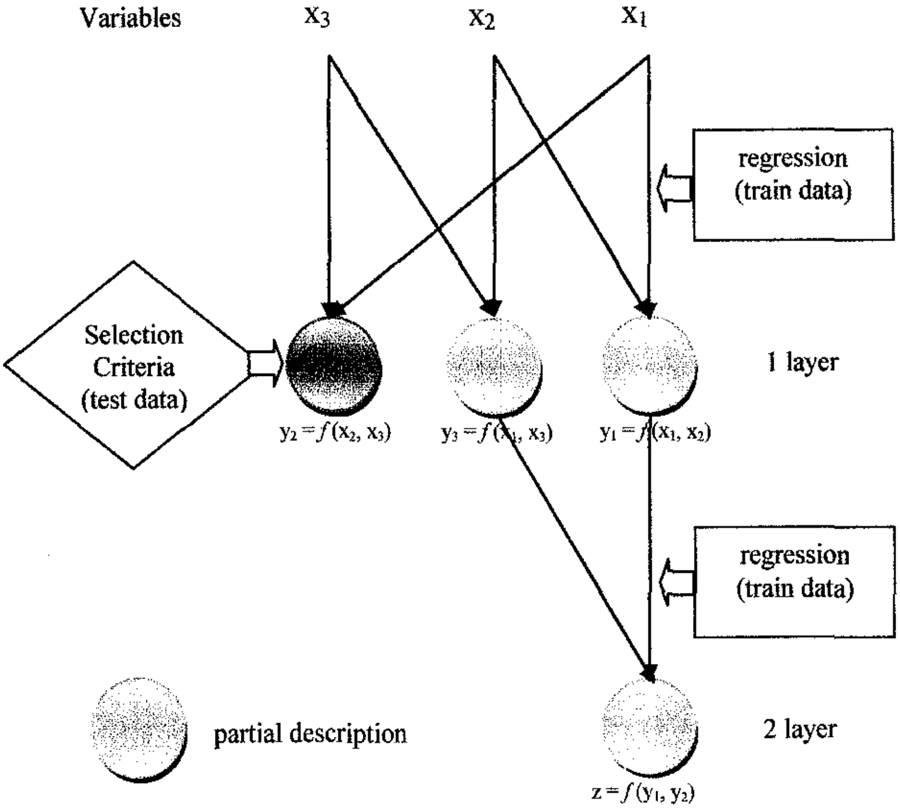


Figure 4: GMDH polynomial neural network structure

a. Literature review

In order to understand the discussed study, papers had been reviewed and studies were conducted along the project development stages. During the initial stage, empirical correlations were reviewed. EOS based viscosity model was intended to be developed by referring to PVT analysis and phase behavior textbooks. Later, GMDH method was reviewed and a compositional model was developed based on this method.

b. Data collection

Based on the literature quoted as reference 4, PVT data from Middle East, North Sea and the other places were collected. The needed data are pressure, temperature, laboratory measured viscosity, mole fraction of crude oil compositions (C_1 - C_{7+}), mole fraction of non-hydrocarbon compositions (N_2, CO_2, H_2S), specific gravity of C_{7+} and molecular weight of C_{7+}.

c. Model Development

During the model development process, Mathematica, Microsoft Excel and VariReg software have been attempted. Nonlinear regression analysis in Mathematica was the first attempted but unfortunately the version available did not support multiple response data. The coding will be attached in the appendix. After that, Microsoft Excel Analysis ToolPax had been attempted with a few trials but the result was not better than the published correlations. At last, GMDH method was attempted using VariReg software.

d. Results and Discussions

After the compositional model had been developed, statistical and graphical analysis will be used to portrait the results and compare with other published correlations.

e. Conclusions

After the results had been discussed, it is time to conclude the project by summarizing the findings along the project. Last but not least, the objectives were reviewed in order to determine whether they have been achieved.

3.2 Model Development Methodology using VariReg software

The following are the steps to develop the proposed composition model:

1. Load training data
2. Load test data from file
3. Select PNN tab
4. Key in degree of polynomials for individual neurons
5. Select various radio buttons (A-E):
    a. Selection of full polynomials (no subset selection) or one of the subset selection algorithms for generating neurons.
    b. Selection of criterion for subset selection in each neuron as well as for deciding when to stop the building of the network.
    c. Selection of whether the inputs to the neurons are taken only from the immediately preceding layer or also prom the original input variables.
6. Select maximum number of inputs for each neuron.
7. Select maximum number of neurons in each layer.
8. (Optional) Draw surface of the model right after building it.
9. (Optional) Seeking optimum of the output variable using the built model
10. (Optional) Saving the predictions of the built model to a file right after building it.
11. Information on the just built model:
    a. "Total number of generated layers" – the total number of the generated layers of the network (the last layer is discarded)
    b. "Number of layers" – the number of layers in the final network
    c. "Used input variables" – the list of input variables used in the final network
    d. "The number of used input variables" – the number of input variables used in the final network
    e. "Crit value" – used criterion's value for the final network

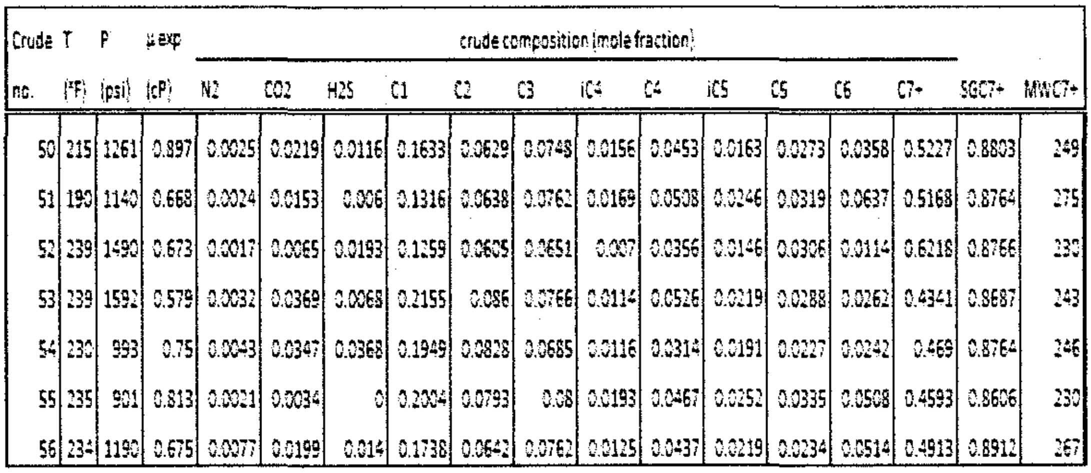
(The software user interface is available in the Appendix)

CHAPTER 4

RESULTS and DISCUSSIONS

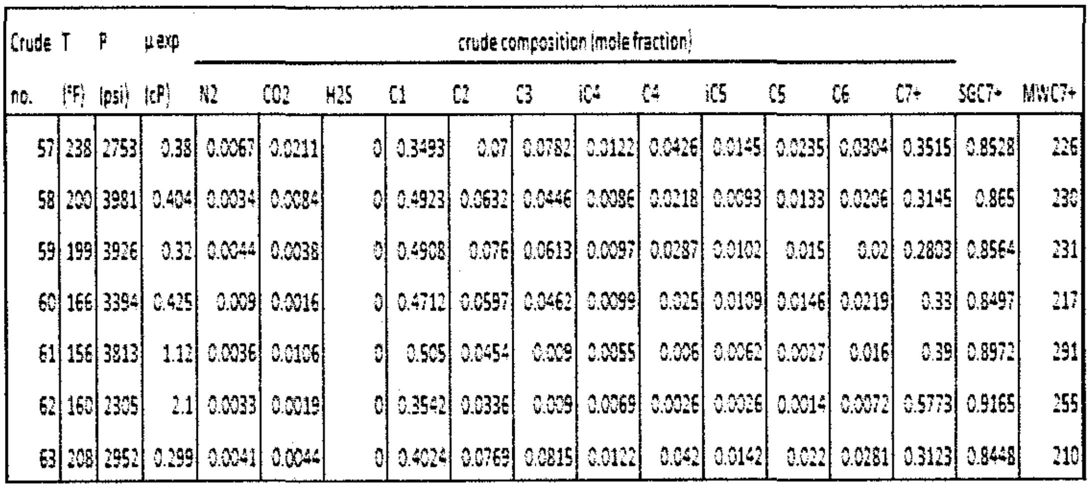
4.1 PVT Data Collections

The new compositional model are developed from a database consists of crude oils composition, viscosity, temperature and pressure from Middle East, North Sea and the other places. 75 crude oil samples are collected through literature search where 49 crude oil samples from Kuwait [4], 7 crude oil samples from Abu Dhabi [12], 7 crude oil samples from North Sea [10] and 12 crude oil samples from Lawal [15] paper. The composition and properties of the crude oil samples used in the study are shown in Tables 2-5.



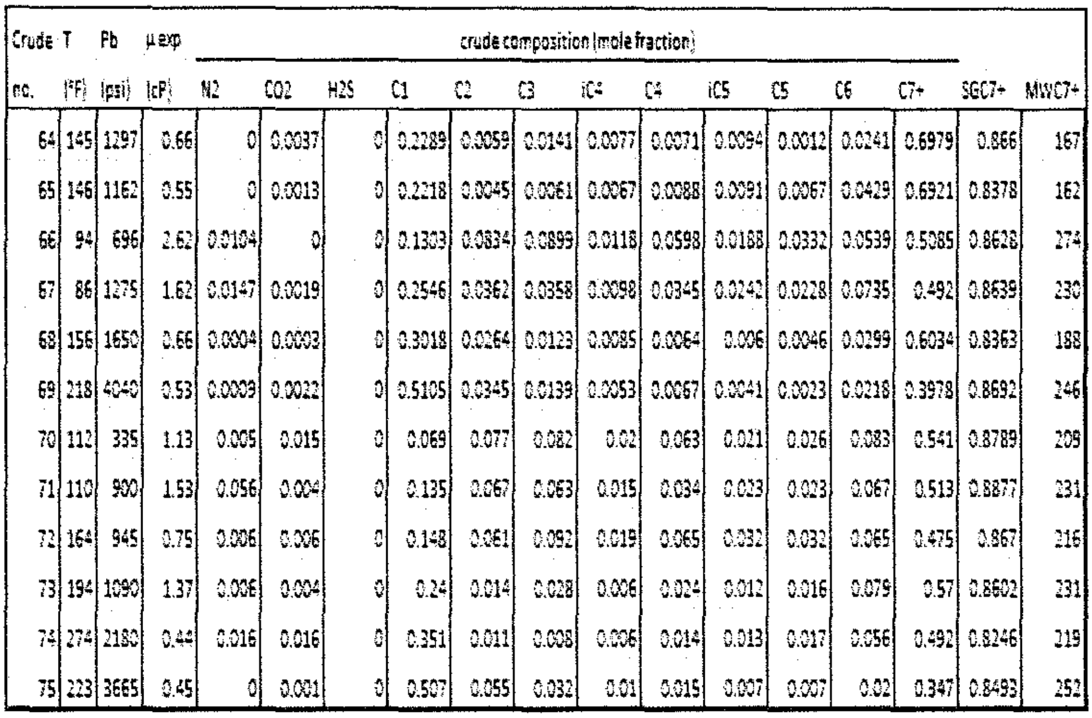
Crude	T	P	μ exp	crude composition (mole fraction)													
no.	(°F)	(psi)	(cP)	N2	CO2	H2S	C1	C2	C3	iC4	C4	iC5	C5	C6	C7+	SGC7+	MWC7+
50	215	1261	0.897	0.0025	0.0219	0.0116	0.1633	0.0629	0.0748	0.0156	0.0453	0.0163	0.0273	0.0358	0.5227	0.8803	249
51	190	1140	0.668	0.0024	0.0153	0.006	0.1316	0.0638	0.0762	0.0169	0.0508	0.0246	0.0319	0.0637	0.5168	0.8764	275
52	239	1490	0.673	0.0017	0.0065	0.0193	0.1259	0.0605	0.0651	0.007	0.0356	0.0146	0.0306	0.0114	0.6218	0.8766	230
53	239	1592	0.579	0.0032	0.0369	0.0068	0.2155	0.086	0.0766	0.0114	0.0526	0.0219	0.0288	0.0262	0.4341	0.8687	243
54	230	993	0.75	0.0043	0.0347	0.0368	0.1949	0.0828	0.0685	0.0116	0.0314	0.0191	0.0227	0.0242	0.469	0.8764	246
55	235	901	0.813	0.0021	0.0034	0	0.2004	0.0793	0.08	0.0193	0.0467	0.0252	0.0335	0.0508	0.4593	0.8606	230
56	234	1190	0.675	0.0077	0.0199	0.014	0.1738	0.0642	0.0762	0.0125	0.0437	0.0219	0.0234	0.0514	0.4913	0.8912	267

Table 2: Composition and Viscosity of Crude Oils from UAE



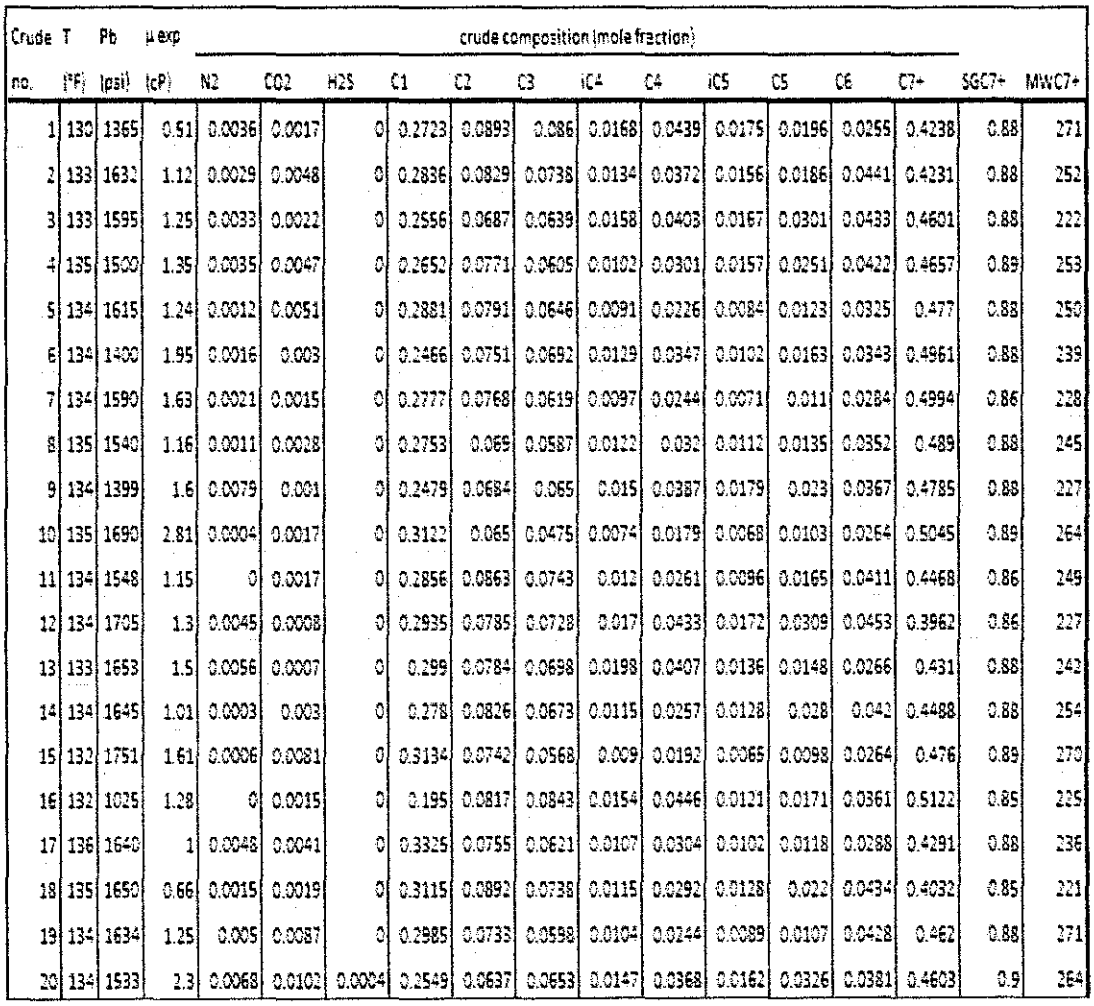
Crude	T	P	μ exp	crude composition (mole fraction)													
no.	(°F)	(psi)	(cP)	N2	CO2	H2S	C1	C2	C3	iC4	C4	iC5	C5	C6	C7+	SGC7+	MWC7+
57	238	2753	0.38	0.0067	0.0211	0	0.3493	0.07	0.0782	0.0122	0.0426	0.0145	0.0235	0.0304	0.3515	0.8528	226
58	200	3981	0.404	0.0034	0.0084	0	0.4923	0.0632	0.0446	0.0086	0.0218	0.0093	0.0133	0.0206	0.3145	0.865	230
59	199	3926	0.32	0.0044	0.0038	0	0.4908	0.076	0.0613	0.0097	0.0287	0.0102	0.015	0.02	0.2803	0.8564	231
60	166	3394	0.425	0.009	0.0016	0	0.4712	0.0597	0.0462	0.0099	0.025	0.0109	0.0146	0.0219	0.33	0.8497	217
61	156	3813	1.12	0.0036	0.0106	0	0.505	0.0454	0.009	0.0055	0.006	0.0062	0.0027	0.016	0.39	0.8972	291
62	160	2305	2.1	0.0033	0.0019	0	0.3542	0.0336	0.009	0.0069	0.0026	0.0026	0.0014	0.0072	0.5773	0.9165	255
63	208	2952	0.299	0.0041	0.0044	0	0.4024	0.0769	0.0815	0.0122	0.042	0.0142	0.022	0.0281	0.3123	0.8448	210

Table 3: Composition and Viscosity of Crude Oils from the North Sea



Crude	T	Pb	μ exp	crude composition (mole fraction)													
no.	(°F)	(psi)	(cP)	N2	CO2	H2S	C1	C2	C3	iC4	C4	iC5	C5	C6	C7+	SGC7+	MWC7+
64	145	1297	0.66	0	0.0037	0	0.2289	0.0059	0.0141	0.0077	0.0071	0.0094	0.0012	0.0241	0.6979	0.866	167
65	146	1162	0.55	0	0.0013	0	0.2218	0.0045	0.0061	0.0067	0.0088	0.0091	0.0067	0.0429	0.6921	0.8378	162
66	94	696	2.62	0.0104	0	0	0.1303	0.0834	0.0899	0.0118	0.0598	0.0188	0.0332	0.0539	0.5085	0.8628	274
67	86	1275	1.62	0.0147	0.0019	0	0.2546	0.0362	0.0358	0.0098	0.0345	0.0242	0.0228	0.0735	0.492	0.8639	230
68	156	1650	0.66	0.0004	0.0003	0	0.3018	0.0264	0.0123	0.0085	0.0064	0.006	0.0046	0.0299	0.6034	0.8363	188
69	218	4040	0.53	0.0009	0.0022	0	0.5105	0.0345	0.0139	0.0053	0.0067	0.0041	0.0023	0.0218	0.3978	0.8692	246
70	112	335	1.13	0.005	0.015	0	0.069	0.077	0.082	0.02	0.063	0.021	0.026	0.083	0.541	0.8789	209
71	110	900	1.53	0.056	0.004	0	0.135	0.067	0.063	0.015	0.034	0.023	0.023	0.067	0.513	0.8877	231
72	164	945	0.75	0.006	0.006	0	0.148	0.061	0.092	0.019	0.065	0.032	0.032	0.065	0.475	0.867	216
73	194	1090	1.37	0.006	0.004	0	0.24	0.014	0.028	0.006	0.024	0.012	0.016	0.079	0.57	0.8602	231
74	274	2180	0.44	0.016	0.016	0	0.351	0.011	0.008	0.006	0.014	0.013	0.017	0.056	0.492	0.8246	219
75	223	3665	0.45	0	0.001	0	0.507	0.055	0.032	0.01	0.015	0.007	0.007	0.02	0.347	0.8493	252

Table 4: Composition and Viscosity of Crude Oils from Lawal paper



Crude	T	Pb	μ exp	crude composition (mole fraction)													
no.	(°F)	(psi)	(cP)	N2	CO2	H2S	C1	C2	C3	iC4	C4	iC5	C5	C6	C7+	SGC7+	MWC7+
1	130	1365	0.51	0.0036	0.0017	0	0.2723	0.0893	0.086	0.0168	0.0439	0.0175	0.0196	0.0255	0.4238	0.88	271
2	133	1632	1.12	0.0029	0.0048	0	0.2836	0.0829	0.0738	0.0134	0.0372	0.0156	0.0186	0.0441	0.4231	0.88	252
3	133	1595	1.25	0.0033	0.0022	0	0.2556	0.0687	0.0639	0.0158	0.0403	0.0167	0.0301	0.0433	0.4601	0.88	222
4	135	1500	1.35	0.0035	0.0047	0	0.2652	0.0771	0.0605	0.0102	0.0301	0.0157	0.0251	0.0422	0.4657	0.89	253
5	134	1615	1.24	0.0012	0.0051	0	0.2881	0.0791	0.0646	0.0091	0.0226	0.0084	0.0123	0.0325	0.477	0.88	250
6	134	1400	1.95	0.0016	0.003	0	0.2466	0.0751	0.0692	0.0129	0.0347	0.0102	0.0163	0.0343	0.4961	0.88	239
7	134	1590	1.63	0.0021	0.0015	0	0.2777	0.0768	0.0619	0.0097	0.0244	0.0071	0.011	0.0284	0.4994	0.86	228
8	135	1540	1.16	0.0011	0.0028	0	0.2753	0.069	0.0587	0.0122	0.032	0.0112	0.0135	0.0352	0.489	0.88	245
9	134	1399	1.6	0.0079	0.001	0	0.2479	0.0684	0.065	0.015	0.0387	0.0179	0.023	0.0367	0.4785	0.88	227
10	135	1690	2.81	0.0004	0.0017	0	0.3122	0.065	0.0475	0.0074	0.0179	0.0068	0.0103	0.0264	0.5045	0.89	264
11	134	1548	1.15	0	0.0017	0	0.2856	0.0863	0.0743	0.012	0.0261	0.0096	0.0165	0.0411	0.4468	0.86	249
12	134	1705	1.3	0.0045	0.0008	0	0.2935	0.0785	0.0728	0.017	0.0433	0.0172	0.0309	0.0453	0.3962	0.86	227
13	133	1653	1.5	0.0056	0.0007	0	0.299	0.0784	0.0698	0.0198	0.0407	0.0136	0.0148	0.0266	0.431	0.88	242
14	134	1645	1.01	0.0003	0.003	0	0.278	0.0826	0.0673	0.0115	0.0257	0.0128	0.028	0.042	0.4488	0.88	254
15	132	1751	1.61	0.0006	0.0081	0	0.3134	0.0742	0.0568	0.009	0.0192	0.0065	0.0098	0.0264	0.476	0.89	270
16	132	1025	1.28	0	0.0015	0	0.195	0.0817	0.0843	0.0154	0.0446	0.0121	0.0171	0.0361	0.5122	0.85	225
17	136	1640	1	0.0048	0.0041	0	0.3325	0.0755	0.0621	0.0107	0.0304	0.0102	0.0118	0.0288	0.4291	0.88	236
18	135	1650	0.66	0.0015	0.0019	0	0.3115	0.0892	0.0738	0.0115	0.0292	0.0128	0.022	0.0434	0.4032	0.85	221
19	134	1634	1.25	0.005	0.0087	0	0.2985	0.0733	0.0598	0.0104	0.0244	0.0089	0.0107	0.0428	0.462	0.88	271
20	134	1533	2.3	0.0068	0.0102	0.0004	0.2549	0.0637	0.0653	0.0147	0.0368	0.0162	0.0326	0.0381	0.4603	0.9	264

Table 5: Composition and Viscosity of Kuwaiti Crude Oils

21	135	1548	1.74	0.0022	0.0137	0	0.2792	0.0651	0.0516	0.0086	0.0166	0.0074	0.0084	0.0259	0.5213	0.89	255
22	134	1825	3.52	0.0019	0.0251	0	0.3189	0.0751	0.0659	0.0124	0.0353	0.0135	0.0216	0.0329	0.3974	0.92	324
23	134	1825	3.39	0.0018	0.0246	0	0.3026	0.073	0.0636	0.0119	0.0342	0.014	0.0214	0.0322	0.4208	0.91	299
24	133	1565	1.56	0.0044	0.0083	0	0.2775	0.0754	0.069	0.0142	0.0319	0.0148	0.0162	0.0311	0.4572	0.9	272
25	134	1580	1.91	0.0014	0.0054	0	0.2944	0.0796	0.0619	0.0107	0.0289	0.0125	0.02	0.0384	0.4468	0.9	290
26	132	1380	2.01	0.0015	0.0039	0	0.2755	0.0786	0.063	0.0102	0.0257	0.0099	0.0159	0.0368	0.479	0.9	152
27	134	1680	1.9	0.0026	0.0126	0	0.2827	0.06	0.0485	0.0087	0.0273	0.0153	0.0247	0.0426	0.475	0.91	274
28	132	1738	2.02	0.002	0.008	0	0.3142	0.076	0.0723	0.0142	0.0346	0.0089	0.0108	0.0099	0.4491	0.89	251
29	132	1920	1.77	0	0.013	0	0.3453	0.0724	0.0531	0.0087	0.0192	0.0073	0.0118	0.0348	0.4344	0.89	281
30	133	1655	1.86	0.0054	0.005	0	0.2779	0.0689	0.0634	0.0125	0.032	0.0134	0.0195	0.035	0.467	0.89	214
31	135	1880	1	0.0041	0.0065	0	0.3021	0.0689	0.066	0.0163	0.0425	0.0177	0.0285	0.0427	0.4047	0.88	247
32	133	1788	1.54	0.002	0.0119	0	0.3226	0.0757	0.0597	0.0111	0.0274	0.0122	0.0195	0.0363	0.4216	0.88	134
33	133	1805	1.52	0.001	0.0116	0	0.3293	0.0778	0.0606	0.0108	0.0265	0.0114	0.0178	0.037	0.4162	0.88	279
34	135	1500	1.74	0.002	0.0049	0	0.2946	0.0851	0.0663	0.0107	0.029	0.0123	0.0214	0.0379	0.4359	0.9	146
35	135	1500	1.91	0.0015	0.0048	0	0.2851	0.0817	0.0636	0.0107	0.0308	0.0143	0.023	0.0387	0.4459	0.9	142
36	133	1715	1.05	0.0049	0.0047	0	0.3155	0.0866	0.0792	0.0147	0.0401	0.0155	0.0192	0.0445	0.3751	0.9	289
37	130	1670	1.55	0.0015	0.0065	0	0.3048	0.0727	0.0515	0.0069	0.0135	0.0078	0.0118	0.0276	0.4954	0.87	239
38	135	1735	1.48	0.0045	0.0051	0	0.3056	0.0725	0.0728	0.0144	0.038	0.0155	0.0221	0.0385	0.411	0.89	268
39	134	1530	2.1	0.0023	0.0106	0.0001	0.2475	0.0678	0.068	0.0175	0.0426	0.016	0.0275	0.0389	0.4612	0.91	256
40	135	1766	1.99	0.0019	0.0012	0	0.2862	0.0866	0.0652	0.0114	0.0364	0.0114	0.0157	0.0362	0.4478	0.93	274
41	140	1730	3.01	0.0021	0.0072	0	0.3312	0.0744	0.0607	0.0094	0.0228	0.0067	0.0112	0.0305	0.4438	0.9	271
42	168	2505	1.3	0.0008	0.0137	0.0069	0.3597	0.0867	0.0594	0.0085	0.0212	0.0058	0.0092	0.031	0.3971	0.9	274
43	208	1877	0.79	0	0.0112	0.0021	0.2695	0.1076	0.0923	0.0096	0.0401	0.009	0.0172	0.0333	0.4081	0.88	249
44	230	2110	0.49	0	0.0125	0	0.3335	0.0924	0.0607	0.0063	0.0177	0.0053	0.0109	0.0278	0.4329	0.85	252
45	230	3730	0.25	0	0.0068	0	0.4879	0.1057	0.0621	0.0073	0.0273	0.0284	0.0172	0.028	0.248	0.85	218
46	241	3335	0.44	0.0005	0.0085	0	0.4105	0.1107	0.0707	0.0096	0.0386	0.0104	0.0185	0.0319	0.2901	0.85	198
47	240	3630	0.43	0.0006	0.0094	0	0.4444	0.1076	0.0618	0.008	0.0324	0.0091	0.0172	0.0257	0.2838	0.85	195
48	241	3120	0.45	0.0004	0.0078	0	0.4091	0.1072	0.0664	0.0087	0.0357	0.0095	0.0167	0.0251	0.3134	0.85	202
49	243	3180	0.46	0.0006	0.0085	0	0.407	0.1054	0.0655	0.0087	0.0354	0.0094	0.0178	0.0281	0.3136	0.85	195

Table 5 (continue): Composition and Viscosity of Kuwaiti Crude Oils

4.2 Viscosity Model

In order to obtain a practical model with reasonable accuracy of result, several attempts have been made. VariReg, a software tool for general purpose multidimensional regression modelling was used for the models development. This freeware provides means for creating "full" polynomial regression models, sparse polynomial models and employing subset selection algorithms [20, 21]. In the software, some input GMDH parameters were varied in terms of:

- degree of polynomials for individual neurons,
- algorithm for individual neurons,
- maximum number of inputs for each neuron,
- maximum number of neurons for each layer,
- inputs from preceding layer or preceding layer and from original input variables

For the criterion for model selection, the model evaluation and hyperparameter selection is done using small sample corrected Akaike's Information Criterion (AICC) using the following equation:

$$AICC = n\ln(MSE) + 2k + (2k(k+1))/(n-k-1) \qquad (2)$$

From 75 world crude oils database from Kuwait, UAE, North Sea and Lawal paper, the proposed empirical models were developed. The viscosity of crude oil can be expressed as a function of temperature, pressure and several oil composition groups. In the model, the following variables were used:

- x_0, T = temperature (degrees Fahrenheit)
- x_1, P = pressure (pound per square inch absolute, psi)
- x_2, G_L = mole fractions of the light components (C_1 and C_2),
- x_3, G_m = mole fractions of the intermediate components (C_3 through C_6),
- x_4, G_h = mole fractions of components heavier than heptanes, C_{7+} fraction,
- x_5, G_{non} = mole fractions of non-hydrocarbon components (N_2, CO_2 and H_2S)

After GMDH was applied, the best statistical model varying different GMDH internal parameters was selected from varied data distribution between train and test data sets and model complexity. Table 6 presented the compassion results from the GMDH models.

Model Number	Test data set, %	Polynomial degree	Function	MSE	RRMSE	AARE
1	20	2	$y = f(x_0, x_2, x_5)$	0.2655	0.704275739	0.415010488
2	20	3	$y = f(x_0, x_4, x_5)$	0.2908	0.737067	0.359043
3	20	4	$y = f(x_2, x_3, x_4)$	40.8589	8.737445	2.06185
4	10	2	$y = f(x_0, x_2, x_5)$	0.2614	0.698926666	0.40000463
5	10	3	$y = f(x_0, x_4, x_5)$	0.1568	0.541276	0.260662
6	10	4	$y = f(x_0, x_2, x_5)$	1.6029	1.730591899	0.45225865

Table 6: GMDH Models

In VariReg the main results are displayed in RRMSE form where 0 is perfect fit; value approaching 1 means the model is no better than a mean value of *y*; value largely exceeding 1 the model probably means overfits the training data. From Table 6, model number 2 and model number 5 showed the best fitted models with the least AARE. These models used T (x_0), G_h (x_4), G_{non} (x_5) as their variables to calculate the viscosity values. Since the test data set was selected by hold up the training data set, the number of training data set will be less with increasing percentage of test data set. Therefore, model number 5 showed better fitting than model number 2 because more data from model number 2 is used as test data set.

After learning that test data set will affect the model fitting, various trials have been attempted. In order to minimise the training data set input lost due to usage in test data, the similar training data set and test data set were used in developing the compositional model. This method results in a compositional model that is capable of predicting the crude oil viscosity with MSE of 0.041781893, RRMSE of 0.279405788 and AARE of 0.086901677. The proposed compositional model has the following form:

$$\begin{aligned}\mu_o \quad &= a_1 + a_2*T + a_3*T^2 - a_4*T^3 + a_5*T^4 - a_6*T^5 + a_7*G_L - a_8*T*G_L \\ &+ a_9*T^2*G_L - a_{10}*T^3*G_L + a_{11}*T^4*G_L - a_{12}*G_L^2 + a_{13}*T*G_L^2 - a_{14}*T^2*G_L^2 \\ &+ a_{15}*T^3*G_L^2 + a_{16}*G_L^3 - a_{17}*T*G_L^3 - a_{18}*T^2*G_L^3 - a_{19}*G_L^4 + a_{20}*T*G_L^4 \\ &- a_{21}*G_L^5 + a_{22}*G_{non} - a_{23}*T*G_{non} + a_{24}*T^2*G_{non} - a_{25}*T^3*G_{non} + a_{26}*T^4*G_{non} \\ &- a_{27}*G_L*G_{non} - a_{28}*T*G_L*G_{non} + a_{29}*T^2*G_L*G_{non} - a_{30}*T^3*G_L*G_{non} \\ &+ a_{31}*G_L^2*G_{non} + a_{32}*T*G_L^2*G_{non} + a_{33}*T^2*G_L^2*G_{non} - a_{34}*G_L^3*G_{non} \\ &- a_{35}*T*G_L^3*G_{non} + a_{36}*G_L^4*G_{non} - a_{37}*G_{non}^2 + a_{38}*T*G_{non}^2 - a_{39}*T^2*G_{non}^2 \\ &+ a_{40}*T^3*G_{non}^2 - a_{41}*G_L*G_{non}^2 - a_{42}*T*G_L*G_{non}^2 + a_{43}*T^2*G_L*G_{non}^2 \\ &+ a_{44}*G_L^2*G_{non}^2 - a_{45}*T*G_L^2*G_{non}^2 + a_{46}*G_L^3*G_{non}^2 - a_{47}*G_{non}^3 + a_{48}*T*G_{non}^3 \\ &+ a_{49}*T^2*G_{non}^3 + a_{50}*G_L*G_{non}^3 - a_{51}*T*G_L*G_{non}^3 - a_{52}*G_L^2*G_{non}^3 + a_{53}*G_{non}^4 \\ &- a_{54}*T*G_{non}^4 - a_{55}*G_L*G_{non}^4 + a_{56}*G_{non}^5 \qquad (3)\end{aligned}$$

where

$a_1 = -8478.58349606828$

$a_2 = 55.7362649737557$

$a_3 = 0.0317567844848732$

$a_4 = 0.00244544177916937$

$a_5 = 1.31281534907682E-5$

$a_6 = 2.43041542877451E-08$

$a_7 = 77823.8744820531$

$a_8 = 574.825569957195$

$a_9 = 3.32289855754005$

$a_{10} = 0.0123543066137851$

$a_{11} = 1.97741181759314E-5$

$a_{12} = 251925.279192096$

$a_{13} = 849.0509876425$

$a_{14} = 1.85805197448581$

$a_{15} = 0.00479457122014624$

$a_{16} = 428350.408327296$

$a_{17} = 659.387877657145$

$a_{18} = 2.85065013356515$

$a_{19} = 359455.090075369$

$a_{20} = 1584.23213582731$

$a_{21} = 80637.6949098182$

$a_{22} = 484740.067189686$

$a_{23} = 3926.90745511693$

$a_{24} = 17.7320700483948$

$a_{25} = 0.0334115078920825$

$a_{26} = 0.000204674937622932$

$a_{27} = 2078397.26136335$

$a_{28} = 9700.52977447604$

$a_{29} = 71.8440983550474$

$a_{30} = 0.454556152809839$

$a_{31} = 10375786.2236233$

$a_{32} = 624.041916585568$

$a_{33} = 210.867090335141$

$a_{34} = 18541552.8677577$

$a_{35} = 71130.5137912637$

$a_{36} = 20628488.0562239$

$a_{37} = 15539999.8022648$

$a_{38} = 332380.871618826$

$a_{39} = 1859.52284112302$

$a_{40} = 1.70447348965117$

$a_{41} = 12312465.5509495$

$a_{42} = 81582.9393949258$

$a_{43} = 1671.76488670865$

$a_{44} = 27332488.2517499$

$a_{45} = 418685.338687488$

$a_{46} = 33613498.2901565$

$a_{47} = 342924264.74869$

$a_{48} = 1804340.33355108$

$a_{49} = 3921.88539760773$

$a_{50} = 861970372.030063$

$a_{51} = 5149971.11606209$

$a_{52} = 209011409.93615$

$a_{53} = 2309360358.65222$

$a_{54} = 18715455.5408219$

$a_{55} = 415750768.871321$

$a_{56} = 8355290234.78248$

4.3 Statistical Error Analysis

Error analysis is utilized to check the accuracy of the models. The error measures used in the preset work are mean squared error (MSE), standard deviation (STD) relative root mean squared error (RRMSE), and average absolute relative error (AARE). Equations for those parameters are given below. Summary of statistical comparisons between models and correlations are presented in Table 6 and Table 7.

1. Mean squared error (MSE):

It is the sum of square error divide by n, the number of data, defined by:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - F(x_{(i)})\right)^2 \qquad (4)$$

2. Standard deviation (STD):

It captures how irregular the problem is, defined by:

$$STD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - \bar{y}\right)^2} \qquad (5)$$

3. Relative root mean squared error (RRMSE):

It is the root mean squared error divide by standard deviation, defined by:

$$RRMSE = \frac{RMSE}{STD} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - F(x_{(i)})\right)^2}}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - \bar{y}\right)^2}} \qquad (6)$$

4. Average absolute relative error (AARE):

It measures the relative absolute deviation from the experimental values, defined by:

$$\mathrm{AARE} = \frac{1}{n}\sum_{i=1}^{n}\left(y_{(i)} - F(x_{(i)})\right) \qquad (7)$$

4.4 Graphical Error Analysis

Graphical tools assist in visualizing the performance and accuracy of a correlation or a model. Crossplots were employed as the graphical analysis technique. In this technique, all the estimated values are plotted against the measured values. Then a 45° straight line that represents a perfect correlation line between the estimated versus actual data paoints is drawn on the crossplot. The tighter the cluster about the unity slope line, the better the agreement between the experimental and the predicted results.

Figure 6, 7 & 8 present crossplots of predicted viscosity versus the actual laboratory measured viscosity for proposed model, Adel's compositional model 3a and Adel's compositional model 3b. From the crossplots, it is clear that the proposed compositional model outperforms the other two compositional models by having the tighter cluster about the unity slope line while others indicate higher scattering range.

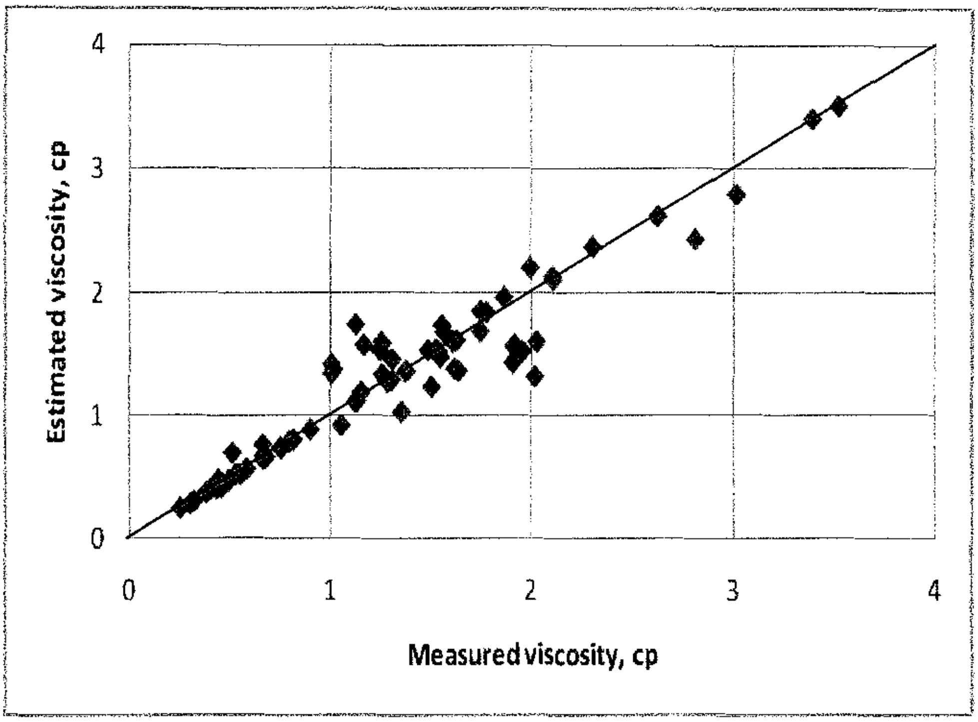


Figure 6: Crossplot of measured versus estimated viscosity
(Proposed compositional model)

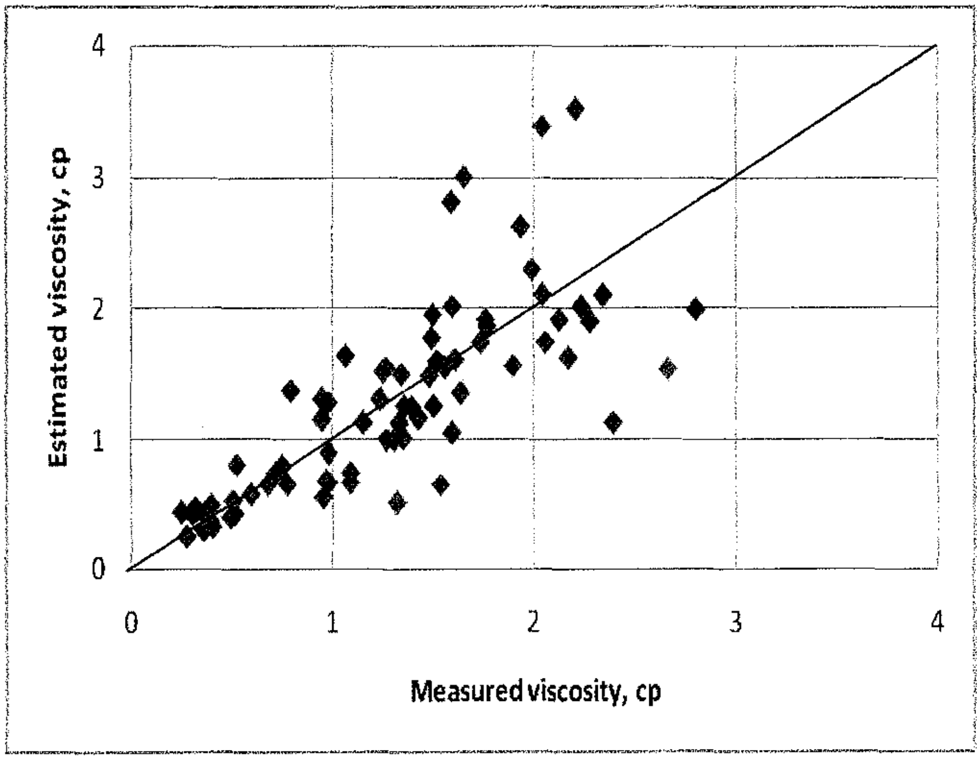


Figure 7: Crossplot of measured versus estimated viscosity
(Adel's compositional model 3a)

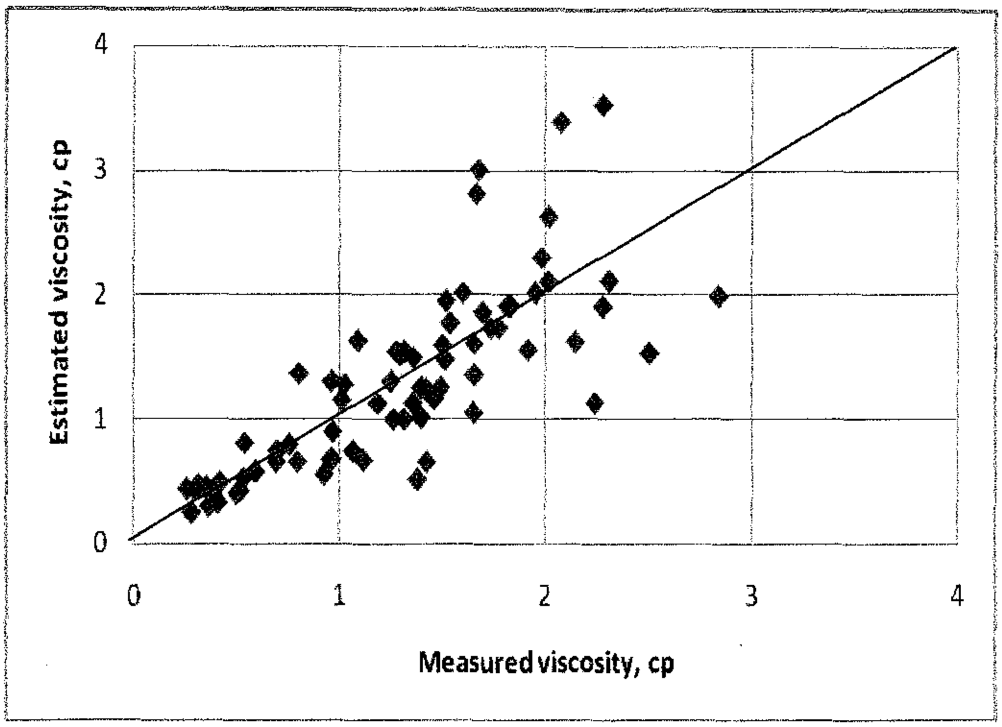


Figure 8: Crossplot of measured versus estimated viscosity
(Adel's compositional model 3b)

4.5 Discussions

Based on Adel M. Elsharkawy *et al.* [4], the 49 Kuwaiti crudes were first evaluated using Beggs and Robinson (1975) [5], Labedi (1992) [6], Kartoatmodjo and Schmidt (1994) [7], and Elsharkawy and Alikhan (1999) [8] correlations. In that literature review, Elsharkawy and Alikhan [8] model has the smallest AAD value, 20.86%, followed by Beggs and Robinson (1975) [5] model (ADD = 26.4%), Kartoatmodjo and Schmidt [7] model (ADD = 32.11%) and Labedi [6] model (ADD = 33.45%). Elsharkawy and Alikhan [8] model showed the best result among the other three correlations because it was developed using Middle East crudes that have some physical and chemical similarities to the Kuwaiti crudes. These empirical models mentioned above are not suitable for accurate viscosity prediction as the viscosity prediction might give the same API gravity although the compositions had changed.

Due to the limitation of empirical models mentioned above, the proposed model considered the compositional of reservoir fluids. Figure 9 shows the comparison between the measured and predicted viscosities by Adel M. Elsharkawy *et al.* [4] compositional models (3a and 3b) and proposed compositional model using

75 crude oil samples. This figure indicates that the proposed model closely matches the experimentally measured viscosity.

Based on the calculated viscosity data presented in Lawal paper, EOS-based models such as were compared with the proposed model using 12 crude oil samples from Lawal paper. The Average Absolute Relative Error (AARE) is the lowest in the proposed model with only 0.8%, followed by Lawal [15] model (26%), Little and Kennedy [14] model (41.69%) and Lohrenz *et al.* [13] model (89.28%). It is believed that the viscosity calculation by Lawal methods is questionable because of highly inaccuracy of AARE.

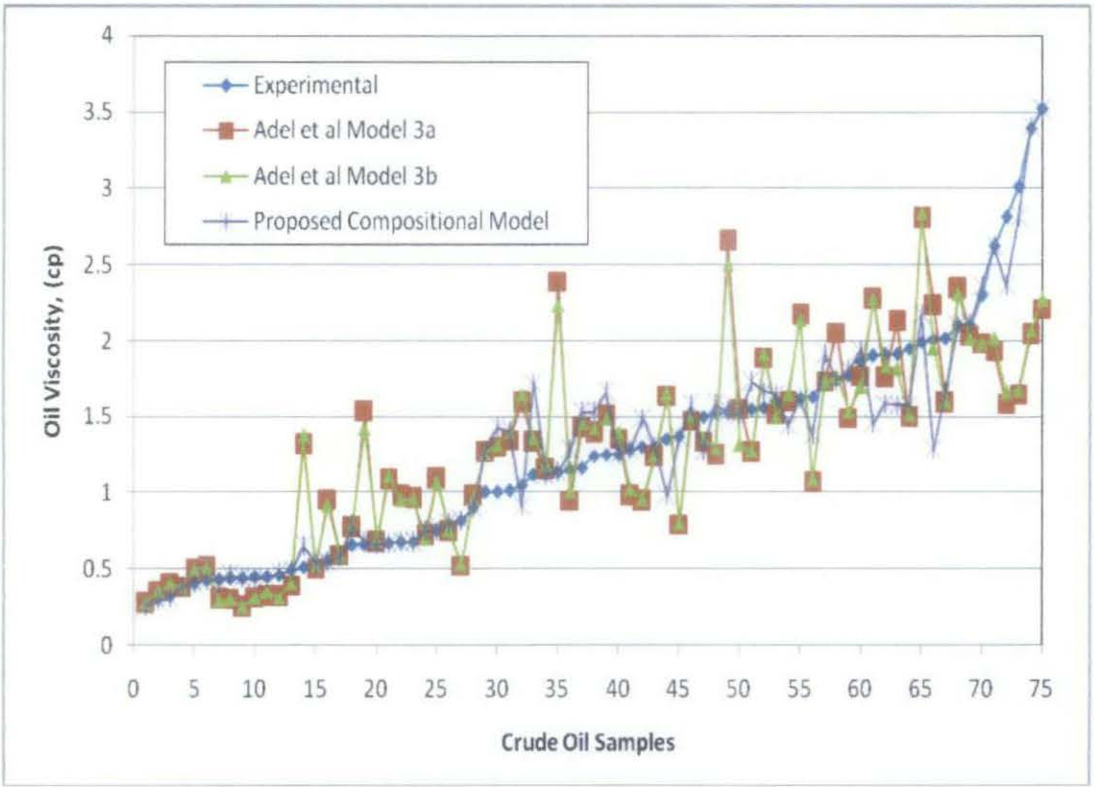


Figure 9: Viscosities comparisons for 75 crude oil samples

Table 7 shows summarize the viscosity calculation results of 75 crude oils from different places of the world along with the data ranges that have been applied in the development of compositional models. These results indicates that the proposed model is more accurate than the other compositional models proposed by Adel M. Elsharkawy *et al.* [20] by displaying only 8.7% of Average Absolute Relative Error (%).

<table>
<tr><th>Data Ranges</th><th colspan="2">Maximum Value</th><th>Minimum Value</th></tr>
<tr><td>T range, °F</td><td colspan="2">274</td><td>86</td></tr>
<tr><td>P range, psi</td><td colspan="2">4040</td><td>335</td></tr>
<tr><td>G_L</td><td colspan="2">0.5936</td><td>0.146</td></tr>
<tr><td>G_m</td><td colspan="2">0.305</td><td>0.0297</td></tr>
<tr><td>G_h</td><td colspan="2">0.6979</td><td>0.248</td></tr>
<tr><td>G_{non}</td><td colspan="2">0.0758</td><td>0.0007</td></tr>
<tr><td>Compositional Model</td><td>3-a (Adel *et al.*)</td><td>3-b (Adel *et al.*)</td><td>Proposed Model</td></tr>
<tr><td>MSE</td><td>0.2212</td><td>0.2009</td><td>0.041781893</td></tr>
<tr><td>RRMSE</td><td>0.642902262</td><td>0.612661</td><td>0.279405788</td></tr>
<tr><td>AARE</td><td>0.271669236</td><td>0.263783</td><td>0.086901677</td></tr>
</table>

Table 7: Comparison of the proposed model to Adel *et al.* compositional models

CHAPTER 5

CONCLUSION & RECOMMENDATIONS

The viscosity compositional model developed from Polynomial Neural Networks (PNN) induced by Group Method of Data Handling (GMDH) is applicable for crude oil only. The new model successfully predicts 75 crude oil samples from different parts of the world with an AARE (%) of 8.7%. The proposed model uses only three input variables, temperature, T, mole fractions of the light components (methane and ethane), G_L and mole fractions of non-hydrocarbon components (N_2, CO_2 and H_2S), G_{non} for the viscosity estimation.

It eliminates the tedious procedure for characterization of heptanes plus, splitting of the heavy fraction, complex mixing rule, which are needed for the EOS-based viscosity model. At the same time, it improves the accuracy of results as compared with the widely used empirical correlations that use oil API gravity as input parameter.

Overall, GMDH is a very useful technique and should be applied more as valuable alternative modeling tool. The proposed compositional model is only applicable for certain range of data at bubble point pressure and it showed a better crude oil viscosity prediction

For further improvement, more data are recommended to be collected for a more accurate and generalized composition model more with a wider range of application.

NOMENCLATURE

PNN =Polynomial Neural Networks
GMDH = Group Method of Data Handling
EOS =Equation of State
PD =Partial description
AICC =Akaike's Information Criterion
T =Shear stress
$\dot{y}$ =Shear rate
F =Force
v =Velocity
A =Area
μ =Dynamic viscosity
υ =Kinematic viscosity
ρ =Fluid density
μ_{od} =Dead oil viscosity
μ_{ob} =Saturated oil viscosity
μ_o =Undersaturated oil viscosity
R_{sb} =Solution gas oil ratio at bubble point
R_s =Solution gas oil ratio
API =API Gravity
T = Temperature
P = Pressure
G_L = Mole fractions of the light components (C_1 and C_2)
G_m = Mole fractions of the intermediate components (C_3 through C_6)
G_h = Mole fractions of components heavier than heptanes, C_{7+} fraction
G_{non} = Mole fractions of non-hydrocarbon components (N_2, CO_2 and H_2S)
AAD = Average absolute deviation
MSE =Mean squared error
STD =Standard deviation
RRMSE =Relative root mean squared error
AARE =Average absolute relative error

REFERENCES

[1] O. Torsaeter, M. Abtahi: "Experimental Reservoir Engineering Laboratory Work Book," Department of Petroleum Engineering and Applied Geoscience, Norwegian University of Science and Technology (**2003**)

[2] Tarek Ahmed: "Equation of State and PVT Analysis, *Applications for Improved Reservoir Modeling*," Gulf Publishing Company, Houston, Texas (**2007**). ISBN 1-933762-03-9

[3] D. H. Xu and A. K. Khurana: "A Simple and Efficient Approach for Improving the Prediction of Reservoir Fluid Viscosity", **1996**, SPE Paper No. 105983

[4] Adel M. Elsharkawy; Suad A.Hassan; Yousef S. Kh. Hashim; Mohamed A. Fahim: "New compositional models for calculating the viscosity of crude oils", *Idn. Eng. Chem. Res.* **2003**, *42*, 4132-4142

[5] Beggs, H. D.; Robinson, J. R. Estimating the viscosity of crude oil system. *J. Pet. Technol.* **1975**, *27*, 1140-1149.

[6] Labedi, R. Improved correlations for predicting the viscosity of light crudes. *J. Pet. Sci. Eng.* **1992**, *8*, 221-234.

[7] Kartoatmodjo, T.; Schmidt, Z. Large data bank improves crude physical property correlations. *Oil Gas J.* **1994**, *4*, July, 51-55.

[8] Elsharkawy, A. M.; Alikhan, A. A. Models for predicting the viscosity of Middle East crude oils. *Fuel* **1999**, *78*, 891-903.

[9] Hanley, H. J. M.; McCarty, R. D.; Haynes, W. M. "Equations for the viscosity and thermal conductivity coefficients of methane". *Cryogenics* **1975**, July, 413-415.

[10] Pedersen, K. S.; Fredenslund, Aa. "An improved corresponding states model for the prediction of oil and gas viscosities and thermal conductivities". *Chem. Eng. Sci.* **1987**, *42*, 182-186.

[11] Petersen, K. A.; Knudsen, K.; Fredenslund, A. "Prediction of Viscosities of Hydrocarbon Mixtures." *Fluid Phase Equilib.* **1991**, 70, 293-308.

[12] Moharam, H. M.; Fahim, M. A. Prediction of the viscosity of heavy petroleum fractions and crude oils using a corresponding states method. *Ind. Eng. Chem. Res.* **1995**, *34*, 4140-4144.

[13] Lohrenz, J., Bray. B. C. and Clark, C. R.: "Calculating Viscosities of Reservoir Fluids from their Compositions", *JPT*, **October 1964**, 1171-1176.

[14] Little, J. E.; Kennedy, H. T. "A correlation of the viscosity of hydrocarbon systems with pressure, temperature and composition". *Soc. Pet. Eng. J.* **1968**, 157-162.

[15] Lawal, A. S. "Prediction of vapor and liquid viscosities from the Lawal-Lake-Silberberg equation of state". Presented at the 5th Symposium on Enhanced Oil Recovery, Tulsa, OK, **April 1986**; SPE/DOE Paper No. 14926.

[16] A.G. Ivakhnenko. *Heuristic Self-Organization in Problems of Engineering Cybernetics*. Automatica 6: pp.207–219, 1970.

[17] A.G. Ivakhnenko. *Polynomial Theory of Complex System*. IEEE Trans. on Systems, Man and Cybernetics, Vol. SMC-1, No. 4, Oct. 1971, pp. 364-378.

[18] E. A. Osman and R. E. Abdel-Aal: "Abductive Networks: A New Modeling Tool for the Oil and Gas Industry." Presented at the SPE Asia Pacific Oil and Gas Conference and Exhibition, Melbourne, Australia, **8-10 October 2002**; SPE Paper No. 77882.

[19] Learning algorithms. In *GMDH Shell Documentation*. Retrieved **July 10, 2011**, from http://www.gmdhshell.com/docs/learning_algorithms

[20] Jekabsons G., Lavendels J. A Comparison of Heuristic Methods for Polynomial Regression Model Induction. Mathematical modelling and analysis, The Baltic Journal on Mathematical Applications, Numerical Analysis and Differential Equations, Buikis A. (ed.), Vol. 13, No. 1, ISSN print: 1392-6292, ISSN online: 1648-3510, Technika, Vilnius, Lithuania, 2008, pp. 17-27.

[21] Jekabsons G. VariReg ver. 0.10.1. A software tool for performing regression modeling using various modeling methods. User's manual - 2009

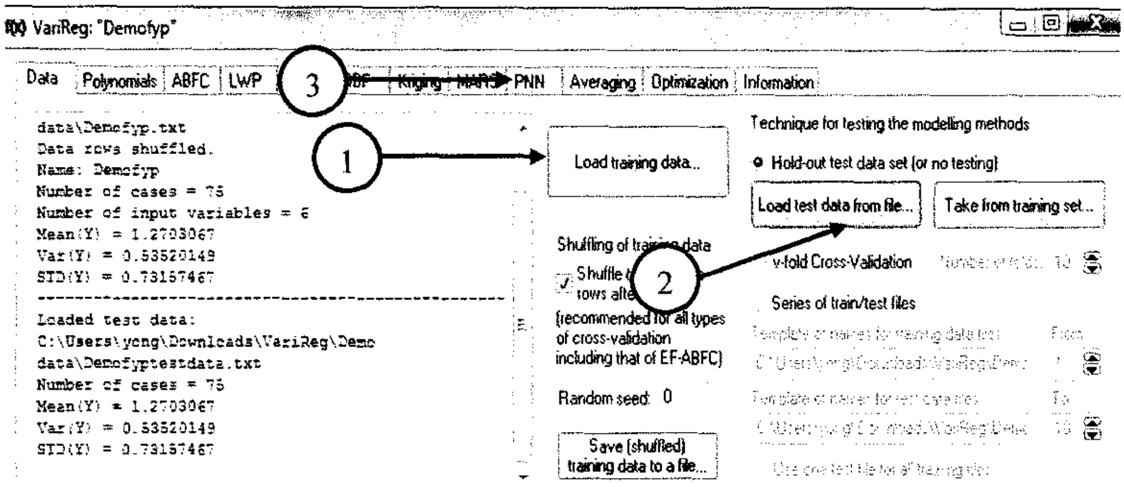
APPENDICES

Mathematica Coding:

```
SetDirectory[$UserDocumentsDirectory]

"C:\\Users\\yong\\Documents"

FileNames["labdata.dat"]

{"labdata.dat"}

Needs["NonlinearRegression`"]

data = Import["labdata.dat", "Table"]

{{"{\\rtf1\\ansi\\ansicpg1252\\deff0\\deflang17417{\\fonttbl{\\f0\\\
fswiss\\fcharset0", "Arial;}}"}, {"{\\*\\generator", "Msftedit",
 "5.41.21.2507;}\\viewkind4\\uc1\\pard\\f0\\fs20", 133, 1595, 0.88,
 0.3243, 0.2101, 0.4601, "1.25\\par"}, {135, 1500, 0.89, 0.3423,
 0.1838, 0.4657, "1.35\\par"}, {134, 1615, 0.88, 0.3672, 0.1495,
 0.477, "1.24\\par"}, {134, 1400, 0.88, 0.3217, 0.1776, 0.4961,
 "1.95\\par"}, {134, 1590, 0.86, 0.3545, 0.1425, 0.4994,
 "1.63\\par"}, {135, 1540, 0.88, 0.3443, 0.1628, 0.489,
 "1.16\\par"}, {134, 1399, 0.88, 0.3163, 0.1963, 0.4785,
 "1.6\\par"}, {135, 1690, 0.89, 0.3772, 0.1163, 0.5045,
 "2.81\\par"}, {134, 1548, 0.86, 0.3719, 0.1796, 0.4468,
 "1.15\\par"}, {"}"}, {}}

NonlinearRegress[data,
 a1*T^a2*P^a3 *Subscript[\[Gamma], c7plus]^a4*Subscript[G, l]^a5*
 Subscript[G, m]^a6*Subscript[G, c7plus]^a7, {a1, a2, a3, a4, a5, a6,
 a7}, {T, P, Subscript[\[Gamma], c7plus], Subscript[G, l],
 Subscript[G, m], Subscript[G, c7plus]}]

NonlinearRegress::bddata: The data argument of NonlinearRegress must \
be a matrix. Weighted regression is specified using the Weights \
option. The data format {{x11, x12, ..., {y11, ..., y1m}}, {x21, x22, \
..., {y21, ..., y2m}}, ..., {{xn1, xn2, ..., {yn1, ..., ynm}}} is \
reserved for multiple response data, which will be supported in the \
future. >>

NonlinearRegress[{{"{\\rtf1\\ansi\\ansicpg1252\\deff0\\deflang17417{\\\
fonttbl{\\f0\\fswiss\\fcharset0", "Arial;}}"}, {"{\\*\\generator",
  "Msftedit", "5.41.21.2507;}\\viewkind4\\uc1\\pard\\f0\\fs20", 133,
  1595, 0.88, 0.3243, 0.2101, 0.4601, "1.25\\par"}, {135, 1500, 0.89,
  0.3423, 0.1838, 0.4657, "1.35\\par"}, {134, 1615, 0.88, 0.3672,
  0.1495, 0.477, "1.24\\par"}, {134, 1400, 0.88, 0.3217, 0.1776,
  0.4961, "1.95\\par"}, {134, 1590, 0.86, 0.3545, 0.1425, 0.4994,
  "1.63\\par"}, {135, 1540, 0.88, 0.3443, 0.1628, 0.489,
  "1.16\\par"}, {134, 1399, 0.88, 0.3163, 0.1963, 0.4785,
  "1.6\\par"}, {135, 1690, 0.89, 0.3772, 0.1163, 0.5045,
  "2.81\\par"}, {134, 1548, 0.86, 0.3719, 0.1796, 0.4468,
  "1.15\\par"}, {"}"}, {}}, a1 P^a3 T^a2 \!
\*SubsuperscriptBox[\(G\), \(c7plus\), \(a7\)] \!
\*SubsuperscriptBox[\(G\), \(l\), \(a5\)] \!
\*SubsuperscriptBox[\(G\), \(m\), \(a6\)] \!
\*SubsuperscriptBox[\(\[Gamma]\), \(c7plus\), \(a4\)], {a1, a2, a3,
 a4, a5, a6, a7}, {T, P, Subscript[\[Gamma], c7plus], Subscript[G,
 l], Subscript[G, m], Subscript[G, c7plus]}]
```

APPENDICES

VariReg software user interface:



```
data\Demofyp.txt
Data rows shuffled.
Name: Demofyp
Number of cases = 75
Number of input variables = 6
Mean(Y) = 1.2703067
Var(Y) = 0.53520149
STD(Y) = 0.73157467
--------------------------------------------------------
Loaded test data:
C:\Users\yeng\Downloads\VariReg\Demo
data\Demofyptestdata.txt
Number of cases = 75
Mean(Y) = 1.2703067
Var(Y) = 0.53520149
STD(Y) = 0.73157467
```

Technique for testing the modelling methods

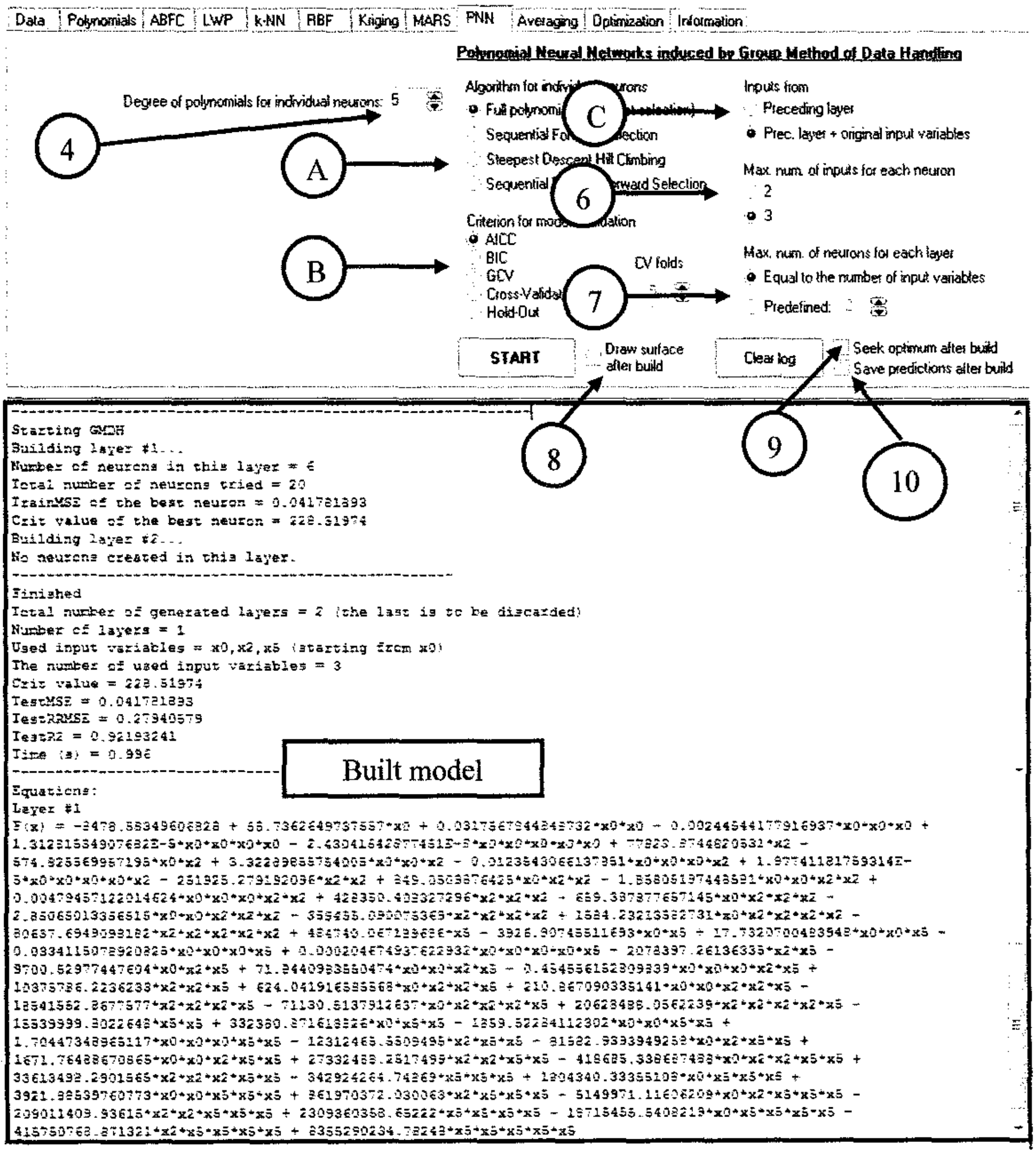
● Hold-out test data set (or no testing)

Shuffling of training data

☑ Shuffle the rows after loading (recommended for all types of cross-validation including that of EF-ABFC)

Random seed: 0

Tab "PNN" user interface:



Polynomial Neural Networks induced by Group Method of Data Handling

Degree of polynomials for individual neurons: 5

Algorithm for individual neurons
● Full polynomial (no selection)
○ Sequential Forward Selection
○ Steepest Descent Hill Climbing
○ Sequential Floating Forward Selection

Criterion for model validation
● AICC
○ BIC
○ GCV
○ Cross-Validation
○ Hold-Out

CV folds

Inputs from
○ Preceding layer
● Prec. layer + original input variables

Max. num. of inputs for each neuron
○ 2
● 3

Max. num. of neurons for each layer
● Equal to the number of input variables
○ Predefined:

☐ Draw surface after build

☐ Seek optimum after build

☐ Save predictions after build

```
--------------------------------------------------------------
Starting GMDH
Building layer #1...
Number of neurons in this layer = 6
Total number of neurons tried = 20
TrainMSE of the best neuron = 0.041721893
Crit value of the best neuron = 228.51974
Building layer #2...
No neurons created in this layer.
--------------------------------------------------------
Finished
Total number of generated layers = 2 (the last is to be discarded)
Number of layers = 1
Used input variables = x0,x2,x5 (starting from x0)
The number of used input variables = 3
Crit value = 228.51974
TestMSE = 0.041721893
TestRRMSE = 0.27940579
TestR2 = 0.92193241
Time (s) = 0.996
--------------------------------------------------------
Equations:
Layer #1
F(x) = -2478.55349606828 + 55.7362649737557*x0 + 0.0317567944243732*x0*x0 - 0.00244544177916937*x0*x0*x0 +
1.31231534907682E-5*x0*x0*x0*x0 - 2.43041642877451E-8*x0*x0*x0*x0*x0 + 77823.8744820531*x2 -
574.825569957195*x0*x2 + 5.32229855754005*x0*x0*x2 - 0.0123543066137351*x0*x0*x0*x2 + 1.97741181759314E-
5*x0*x0*x0*x0*x2 - 251925.279152056*x2*x2 + 349.0503876425*x0*x2*x2 - 1.55805197448591*x0*x0*x2*x2 +
0.00479457122014624*x0*x0*x0*x2*x2 + 428350.408327296*x2*x2*x2 - 659.387877657145*x0*x2*x2*x2 -
2.85065013356515*x0*x0*x2*x2*x2 - 355455.090075369*x2*x2*x2*x2 + 1524.23213382731*x0*x2*x2*x2*x2 -
80657.6949098182*x2*x2*x2*x2*x2 + 484740.067123626*x5 - 3926.90745511693*x0*x5 + 17.7320700493548*x0*x0*x5 -
0.0334115072920825*x0*x0*x0*x5 + 0.000204674937622932*x0*x0*x0*x0*x5 - 2078397.26136335*x2*x5 -
9700.52977447604*x0*x2*x5 + 71.2440933550474*x0*x0*x2*x5 - 0.454556152309339*x0*x0*x0*x2*x5 +
10375786.2236233*x2*x2*x5 + 624.041916585568*x0*x2*x2*x5 + 210.867090335141*x0*x0*x2*x2*x5 -
18541562.8677577*x2*x2*x2*x5 - 71130.8137912637*x0*x2*x2*x2*x5 + 20628486.0562239*x2*x2*x2*x2*x5 -
15539999.3022648*x5*x5 + 332360.871618826*x0*x5*x5 - 1859.52284112302*x0*x0*x5*x5 +
1.70447348965117*x0*x0*x0*x5*x5 - 12312465.5509495*x2*x5*x5 - 81522.9393949252*x0*x2*x5*x5 +
1671.76488670865*x0*x0*x2*x5*x5 + 27332488.2517499*x2*x2*x5*x5 - 418685.338687488*x0*x2*x2*x5*x5 +
33613492.2901565*x2*x2*x2*x5*x5 - 342924264.74269*x5*x5*x5 + 1204340.33355109*x0*x5*x5*x5 +
3921.98539760773*x0*x0*x5*x5*x5 + 361970372.030063*x2*x5*x5*x5 - 5149971.11606209*x0*x2*x5*x5*x5 -
209011409.93615*x2*x2*x5*x5*x5 + 2309360358.65222*x5*x5*x5*x5 - 18715456.5408219*x0*x5*x5*x5*x5 -
415750768.871321*x2*x5*x5*x5*x5 + 8355290234.78248*x5*x5*x5*x5*x5
```