The Determination of Petroleum Reservoir Fluid Properties: Application of Robust Modeling Approaches

By

Arash Kamari

(MSc in Petroleum Reservoirs Engineering, Azad University)

In fulfillment of the requirements of the Doctor of Philosophy in the field of Chemical Engineering, School of Engineering, College of Agriculture, Engineering and Science, University of KwaZulu-Natal, Durban, South Africa

March 2016

Supervisor: Prof Deresh Ramjugernath
Co-Supervisor: Prof Amir H. Mohammadi

PREFACE

The investigation or study presented in this thesis entitled "The Determination of Petroleum Reservoir Fluid Properties: Application of Robust Modeling Approaches" was undertaken at the University of KwaZulu-Natal in the School of Engineering, Howard College Campus, Durban, South Africa. The duration of the study was from November 2013 to February 2016. The study was supervised by Professor Deresh Ramjugernath, and Professor Amir H. Mohammadi.

This thesis has been submitted as the fully requirement for the award of the Doctor of Philosophy (PhD) degree in Chemical Engineering. The study presented in this thesis is my original work, unless otherwise stated.

The thesis has not previously been submitted for degree or examination at any other tertiary institute or university.

_____
Arash Kamari

As the candidate's supervisor, I agree to the submission of this thesis

_____
Professor D. Ramjugernath

DECLARATION 1 - PLAGIARISM

I, Arash Kamari, declare that:

1. The research reported in this thesis, except where otherwise indicated, is my original research.

2. This thesis has not been submitted for any degree or examination at any other university.

3. This thesis does not contain other persons' data, pictures, graphs or other information, unless specifically acknowledged as being sourced from other persons.

4. This thesis does not contain other persons' writing, unless specifically acknowledged as being sourced from other researchers. Where other written sources have been quoted, then:

a. Their words have been re-written but the general information attributed to them has been referenced

b. Where their exact words have been used, then their writing has been placed in italics and inside quotation marks, and referenced.

5. This thesis does not contain text, graphics or tables copied and pasted from the Internet, unless specifically acknowledged, and the source being detailed in the thesis and in the References sections.

Signed

……

DECLARATION 2 - PUBLICATIONS

DETAILS OF CONTRIBUTION TO PUBLICATIONS that form part and/or include research presented in this thesis (include publications in preparation, submitted, *in press* and published and give details of the contributions of each author to the experimental work and writing of each publication)

1) **Kamari, A.**, Sattari, M., Mohammadi, A.H., Ramjugernath, D. Reliable method for the determination of surfactant retention in porous media during chemical flooding oil recovery. ***Fuel*** 2015, 158, 122-128.
2) **Kamari, A.**, Sattari, M., Mohammadi, A.H., Ramjugernath, D. Rapid Model for the Estimation of Dew Point Pressures in Gas Condensate Systems. ***Journal of the Taiwan Institute of Chemical Engineers*** 2016, 60, 258-266.
3) **Kamari, A.**, Sattari, M., Mohammadi, A.H., Ramjugernath, D. Modeling of the Properties of Gasoline and Petroleum Fractions Using a Robust Scheme. ***Petroleum Science and Technology*** 2014, 33, 639-648.
4) **Kamari, A.**, Sattari, M., Mohammadi, A.H., Ramjugernath, D. Modeling of the Vaporization Enthalpies of Petroleum Fractions. ***Fluid Phase Equilibria*** 2016, 412, 25, 228–234.
5) **Kamari, A.**, Gharagheizi, F., Mohammadi, A.H., Ramjugernath, D. Prediction of Natural Gas Compressibility Factors Using a Corresponding States-Based Method. ***Journal of Molecular Liquids*** 2016, 216, 25–34.

ACKNOWLEDGEMENTS

All thanks to the God of heaven, sovereign Lord over all the nations, for granting me the grace and strength to complete this thesis.

First and foremost, I would like to thank both my supervisors Professors Deresh Ramjugernath and Amir H. Mohammadi who have been an essential part of my studies at University of KwaZulu-Natal. Without their invaluable guidance, inspiration and support, this research would never have come to fruition.

A special thanks to Dr Abdolhossein Hemmati-Sarapardeh from Amirkabir University of Technology, my old friend, who was always there to lend a helping hand, enriching the experience, and assisting when I needed help with the modeling process and the revisions requested by journals.

In addition, I am grateful to Dr. Mohammad Nikookar from the National Iranian Oil Company, my previous supervisor, for his support, and encouragement to excel. He was a strong motivating force in bringing me to this point in my research. I never forget him and his kindness.

Thanks are also due to all my fellow colleagues and staff members at the University of KwaZulu-Natal, and, in particular, at the Thermodynamics Research Unit in Chemical Engineering, for their assistance, support and collaboration.

This study would not have been possible without the sponsorship of the National Research Foundation (NRF) and the South African Research Chairs Initiative (SARChI) of the Department of Science and Technology).

Finally, I am sincerely thankful to my beloved parents and family for their tremendous support and encouragement throughout my life, and especially during these years of studying abroad. Without their love and faith in my abilities, I would never have got to where I am.

Table of Content

List of Tables

List of Figures

ABSTRACT

Petroleum reservoir fluid properties are regarded as vital parameters in a number of computations associated with petroleum engineering, including hydrocarbon reserve estimation and consequently economic efficiency evaluation, fluid flow in porous media issues, and improved and enhanced oil recovery. In addition to empirical methods, laboratory measurements, and equations of state are used to determine such properties. The two latter methods are generally expensive and time-consuming and require complex calculations. Therefore, it is timely to develop more rapid, accurate, and reliable models for the determination of petroleum fluid's physical properties.

This study sets out to show that the powerful modeling approaches developed here for the determination of reservoir fluid properties, provide an accurately and efficient alternative to existing methods. They are reliable and straightforward predictive techniques, which are simpler than existing approaches and with fewer computations. To this end, artificial intelligence methods are preferred as the means to develop reliable models for the estimation of petroleum reservoir fluid properties. Recently, artificial intelligence methods have gained popularity in solving complex nonlinear problems, and could be applied in reservoir modeling and characterization. Furthermore, models developed on the basis of artificial intelligence methods can perform prediction and generalization rapidly once trained. Therefore, in this study novel models for the prediction of petroleum reservoir fluid properties are applied in areas including reservoirs simulations, relevant reservoir engineering softwares, calculations associated with enhanced oil recovery processes, fluid flow in porous media, well-testing, and well-logging.

The present study aims to develop reliable models for petroleum reservoir properties using gene expression programming (GEP), artificial neural networks (ANNs), least square support vector machine (LSSVM), adaptive neuro-fuzzy inference system (ANFIS), and decision tree (DT) computational schemes. And, in order to tune the adjustable parameters associated with the algorithms mentioned above, different optimization techniques viz. couples simulated annealing (CSA), particle swarm optimization (PSO), and genetic algorithm (GA), are employed. In order to enable the development of these predictive models thousands of series of data on reservoirs from various geographical locations worldwide were assembled and collated.

The thermos-physical and petroleum reservoir fluid properties studied are surfactant retention, dew point pressure, natural gas compressibility factor, gasoline properties (specific gravity (SG), motor octane number (MON), research octane number (RON), Reid vapor pressure (RVP)), vaporization enthalpy of petroleum fractions, dead reservoir oil viscosity, saturated reservoir oil viscosity, under-saturated reservoir oil viscosity, solution gas-oil ratio, oil formation volume factor, bubble point pressure, wax disappearance temperature, asphaltene precipitation, and hydrocarbon–plus fractions properties viz. boiling point temperature, specific gravity, and molecular weight.

Many empirically derived correlations have been developed, tested and compared. A comparison of results show that the models are superior to those in literature in terms of accuracy and simplicity of use.

NOMENCLATURE

μ_o	Under-saturated reservoir oil viscosity, cP
μ_{ob}	Saturated oil viscosity, cP
μ_{od}	Dead oil viscosity, cP
API	Oil gravity
B_{ob}	Oil formation volume factor at bubble point pressure, bbl/STB
$C_{co\text{-}solvents}$	Co-solvent concentration, wt%
CX_w	Cumulative weight fraction.
C_7^+	Heptane-plus fraction
E_a	Average absolute percent relative error, %
E_i	Percent relative error, %
E_r	Average percent relative error, %
K_{abs}	Absolute permeability, mD
M	Solvent molecular weight, g/mol
MW_b	Bulk molecular weight, g/mol
MW_{C7+}	Molecular weight of heptane-plus fraction, g/mol
MW_{Sur}	Average molecular weight of the surfactant solution, g/mol
N	Number of data points
$n_,$	Number of moles
P	Pressure, psi, Pa, kPa
P_b	Bubble point pressure, psi
P_{ci}	Critical pressure of component i
P_d	Dew point pressure, psi
pH	Maximum effluent pH

P_{pc}	Pseudo-critical pressure
P_{pr}	Pseudo-reduced pressure, psi
R	Gas constant
R	Retention of surfactant, mg/g-Rock
R^2	Coefficient of determination
R_s	Solution gas-oil ratio, SCF/STB
R_{Si}	Initial solution gas oil ratio, SCF/STB
R_v	Solvent to oil dilution ratio, mL/g
r	Relevancy factor, %
S	Specific gravity
SG_b	Bulk specific gravity
SP_{C7+}	Specific gravity of pentane plus fraction
S_{PD}	Salinity of the polymer drive, ppm
T	Temperature, °C
T_b	Boiling point temperature, °R
T_{ci}	Critical temperature component i, °R
T_{pc}	Pseudo-critical temperature, °R
T_{pr}	Pseudo-reduced temperature
T_R	Reservoir temperature, °F
V	Volume
V_{ci}	Critical volume component i
W_t	Amount of asphaltene precipitated, wt%
y_i	Mole fraction of component i
z	Gas compressibility factor
γ	Relative weight of the summation of the regression errors

γ_g Gas specific gravity

σ^2 Squared bandwidth

ABBREVATIONS

AAPRE	Average absolute percent relative error
AARD	Average absolute relative deviation
ANN	Artificial neural network
ANFIS	Adaptive neuro-fuzzy inference system
APRE	Average percent relative error
ARD	Average relative deviation
AI	Artificial intelligence
CCE	Constant composition expansion
CVD	Constant volume depletion
DPP	Dew point pressure
EOR	Enhanced oil recovery
EoS	Equations of state
ET	Expression tree
ET	Expression tree
GA	Genetic algorithm
GC	Group contribution
GEP	Gene expression programming
GOR	Gas oil ratio
GFVF	Gas formation volume factor
GP	Gene programming
HFC	Hydrofluorocarbon
LINGO	Linear Interactive and General Optimizer

LSSVM	Least square support vector machine
MON	Motor octane number
MLP	Multilayer-perceptron
MR	Mobility ratio
MSE	Mean square error
OF	Objective function
OFVF	Oil formation volume factor
QSPR	Quantitative structure–property relationship
RMSE	Root mean square error
RON	Research octane number
RVP	Reid vapor pressure
SA	Simulated annealing
SCF	Standard cubic feet
SD	Standard deviation
SG	Specific gravity
STB	Stock tank barrel
SVM	Support vector machine
TAN	Total acid number of the oil, mg KOH/g-oil
WAT	Wax appearance temperature
WDT	Wax disappearance temperature

CHAPTER 1

1. Introduction

The computations for the estimation of reserves in an oil reservoir - predicting economic efficiency - needs a thorough and accurate knowledge of the reservoir fluid properties (De Ghetto and Villa, 1994).

The properties associated with pressure-volume-temperature (PVT) parameters are accounted as fundamental to many petroleum engineering calculations. Of particular relevance for this study, PVT parameters are used in computations related to enhanced and improved oil recovery, properties estimation of hydrocarbon flowing, analyses related to well testing and well logging, computations associated with material balance (in order to undertake reserves estimation and prediction of petroleum reservoir performance), reservoir oil production and the regulation of surface and reservoir oil and gas volumes.

In other words, it can be agreed that reaching reliable solutions is impossible for many petroleum engineering problems without precise predicts of PVT properties of petroleum reservoir fluids (Khoukhi, 2012).

1.1. Petroleum Reservoir Fluid Properties

The properties of reservoir fluid (Elsharkawy and Alikhan, 1997) are normally determined from bottom-hole and/or surface recombined samples. The fluid properties are required for a large number of reservoir engineering calculations, which include, selection of the enhanced oil recovery (EOR) method for a reservoir candidate, estimation of hydrocarbon reserves, performance prediction, calculations related to the production operation, production optimization, well-testing studies, and fluid flow through porous media (Al-Marhoun, 1988; Al-Marhoun, 2004; Alizadeh et al., 2013; Elsharkawy and Alikhan, 1997; Frashad et al., 1996; Hemmati-Sarapardeh et al., 2013b; Kamari et al., 2015c; Kamari and Mohammadi, 2014; Kamari et al., 2014f; Kamari et al., 2014h). As a result, reservoir fluid properties are significant parameters in petroleum engineering calculations. These are obtained through laboratory measurements, theoretical methods, and/or empirically derived correlations. PVT fluid properties of primary importance are the solution gas oil ratio (GOR), bubble point pressure (BPP), gas formation volume factor (GFVF) and oil formation volume factor (OFVF) (Sutton and Farshad, 1990).

PVT properties are normally acquired experimentally by conducting laboratory tests (El-Sebakhy, 2009a). However, the estimation of PVT properties by laboratory experiments is difficult and complex because cores or rock samples containing petroleum reservoir fluid are not easily acquired and are usually only available from isolated and limited well locations (Khoukhi, 2012). Furthermore, the production horizon does not always warrant the cost of an in-depth petroleum reservoir fluid investigation because it may still be essential to have precise predictions of the fluid's physical properties and therefore PVT properties must still be obtained by the utilization of empirically derived correlations (Saleh et al., 1987). As a result, empirical correlations are frequently employed in the place of time-consuming and costly experimental PVT tests. Another method to determine PVT properties of crude oils is equations of state, which are proposed based on knowledge of detailed reservoir fluid composition. However, accurate determination of such compositions requires much time and cost, and moreover it includes numerous complex numerical calculations (Alimadadi et al., 2011).

The accurate prediction of crude oil viscosity is critical for the petroleum engineer. As a simple definition, the reservoir oil viscosity is the resistance of fluid flow through porous media. It is fundamental to such issues as the design of enhanced oil recovery

processes, the evaluation of fluid flow rate through porous media, the estimation of hydrocarbon reserves, design of operation and production systems equipment and pipelines, production reduction and future performance issues, and the development of reservoir and production simulation software (Al-Marhoun, 2004; Alizadeh et al., 2013; Elsharkwy and Gharbi, 2000; Hemmati-Sarapardeh et al.; Kamari et al., 2014h; Labedi, 1992; Naseri et al., 2005). In addition, this fundamental parameter plays a key role in studying the deposition of wax during crude oil production in the transportation pipelines (Obanijesu and Omidiora, 2008).

Importantly, the efficiency of thermal EOR processes including steam injection, hot water injection, and steam assisted gravity drainage, is strongly associated with reducing the viscosity of heavy oil. Furthermore, the oil production capacity of hydrocarbon reservoirs is directly related to the oil viscosity, so that reservoirs with low viscosities have a higher production capacity (Xu et al., 2010).

It is critically important to develop reliable methods for establishing reservoir fluid properties in order to enable the design of thermal EOR techniques, to address production capacity and petroleum transportation issues.

Gas condensate reservoirs have been growing in importance as a hydrocarbon resource for energy supply. In these reservoirs, well deliverability often reduces when the bottom-hole pressure drops below the dew point pressure (DPP). As a definition, dew point pressure is the pressure at which a considerably larger amount of the gas phase is in equilibrium with a significantly smaller amount of liquid phase (Shokir, 2008). As a result, the calculation of dew point pressure plays a significant role in hydrocarbon reservoir engineering. Gas condensate reservoirs differ in their thermodynamic and flow behavior compared to common gas reservoirs.

Due to decreasing oil production and a general increase in the demand for oil and oil products, as well as concerns about the future of hydrocarbon reserves, near saturation of techniques for optimization of production facilities, and oil price volatility, considerable work is being undertaken on enhanced oil recovery techniques (Kamari et al., 2014g). There is a recent trend to continue production from mature crude oil reservoirs and to assess options for increasing their ultimate oil recovery (Kamari and Mohammadi, 2014). In this vein, EOR techniques have gained attention in the petroleum

industry for their potential to recover larger amounts of oil from depleted reservoirs compared with conventional production methods (Al Adasani and Bai, 2011).

An EOR technique to optimize oil recovery requires a decrease in the saturation of residual oil by means of a reduction in the interfacial tension existing between phases. Among chemical flooding methods, the use of surfactants decreases interfacial tension between oil and water, enabling recovery of much of the saturated residual oil in hydrocarbon reservoirs. After water flooding, the injection of surfactants as an oil recovery method has been implemented for more than 35 years in depleting oil reservoirs, especially in the United States of America (Ferrell et al., 1988; Garrett, 1972; Green and Willhite, 1998; Maerker and Gale, 1992). Typically, surfactant injection is an expensive oil recovery method in comparison with other recovery processes like gas injection and thermal recovery techniques. However, in the early 21st Century surfactant flooding has experienced an increase in interest because of higher oil prices (Iglauer et al., 2010). In other words, surfactant-implemented oil recovery is influenced by economics. Questions relate to the cost of surfactants and the development of a practical EOR process aimed at minimizing the use of surfactant (Novosad et al., 1981). Therefore, a better understanding of the retention of these emulsions by crude oils is of paramount importance for progress in the use of this technique for EOR.

The determination of physical properties of substances is, in general, a difficult task. In the case of determining the physical properties of petroleum products the difficulty is even greater as petroleum is a complex substance, the full composition of which is not entirely known (Litani-Barzilai et al., 1997).

Gasoline which is a petroleum product fractionated from crude oil is a complex mixture of combustible and volatile compounds (Mendes et al., 2012). It is one of the most recognized petroleum products as it is used as a fuel for transportation globally (Murty and Rao, 2004). Moreover, approximately 70% of the crude oil produced is normally processed into gasoline in integrated petroleum refineries (Murty and Rao, 2004).

Gasoline fractions are normally separated into several hydrocarbon groups, viz. aromatics, olefins and naphthenics, iso-paraffins and n-paraffins (de Oliveira et al., 2004; Teixeira et al., 2007). Additionally, the specific gravity (SG), motor octane number (MON), research octane number (RON), and Reid vapor pressure (RVP) are very important

gasoline properties because they characterize the performance and therefore the price of fuels (Albahri, 2014).

The RVP provides an indication of gasoline's tendency to evaporate. The higher the RVP the easier it is for a fuel to evaporate (Teixeira et al., 2009). SG plays a key role in assessing the performance and operational-ability of the engine (Aleme et al., 2009). Ratings or octane numbers are associated with the power and efficiency of a petrol or diesel engine operated using gasoline. It is used to evaluate the quality of the gasoline and also to classify the gasoline by price and grade (viz. premium or regular) (Assis et al., 2013; Doble et al., 2003). There are different kinds of engines, and consequently different test conditions, which result in the two most common octane rating scales, viz. MON and RON. The MON and RON are determined by assessing the fuel in a test engine under rigorous and controlled conditions (Assis et al., 2013).

Heavy paraffins in hydrocarbon reservoir's fluid including crude oil and gas condensate reserves possess high potential of forming wax (Ji et al., 2004). Wax deposit ingredients are located between pure paraffins to asphalthenes. In this wide variation, wax deposits include a lot of constituents such as microcrystalline paraffins, asphalthenes, resins. Wax deposition is mainly affected by thermodynamic conditions and the reservoir fluid's composition (García and Carbognani, 2001). One of the main characteristics of wax is wax appearance temperature (WAT) which causes wax to precipitate in pipe lines and transferring tools when the temperature is reduced below this value (Daridon et al., 2002; Moradi et al., 2013b). In wax appearance temperature, which is also known as cloud point, first crystals of paraffins are formed as a result of dropping temperature. This process is reversible and consequently, the temperature in which the last precipitated paraffin is re-solved in fluid is known as wax disappearance temperature (WDT). The difference between the WDT and WAT can be noticeable (Parsa et al., 2014; Wang et al., 2003). WDT always relates to equilibrium conditions.

The main problems in transferring equipment and reservoirs associated with wax deposition, which give rise to research and development, include reduction in production rate, imposing damage in well bore, increased power needs, fuel filter and process implement damage (Ji et al., 2004). One of the most serious consequence of wax deposition is reduction in diameter of pipelines and equipment which leads to cessation of procedure. To avoid wax deposition in pipes and equipment, the transferring route of

fluid should be maintained in conditions in which wax cannot be precipitated. This could be achieved either by mixing the oil or increasing the temperature (Daridon et al., 2002; Moradi et al., 2013b).

Asphaltene molecules are complex structures and are the heaviest part of crude oils and cannot be dissolved in light hydrocarbon solvents such as n-pentane, n-heptane, and n-decane (Zanganeh et al., 2015). The precipitation of asphaltenes is directly associated with the stability condition of crude oil, which includes changes in the reservoir pressure and temperature as well as the chemical composition of petroleum fractions (Junior et al., 2006; Srivastava et al., 1999). As a result, precipitation of asphaltene is known as a problematic phenomenon in the petroleum industry, and in particular during oil production from hydrocarbon reservoirs to the pipelines.

The precipitation of asphaltene can lead to serious engineering problems including reduced oil recovery and relative permeabilities, blockage of rock pores, and reduction of flow rate. Importantly, asphaltene precipitation causes the reduction of relative permeabilities with changes of reservoir rock wettability from water-wet rock to oil-wet rock which can decreased the oil recovery factor (Amin et al., 2010; Soorghali et al., 2014). Additionally, wellbore damage caused by the blocking of the rock pores, and the reduction of processing facility capability with the plugging of surface pipelines during petroleum production, are two other serious problems associated with asphaltene precipitation (Buckley and Wang, 2002).

Asphaltene precipitation is also an important issue for enhanced oil recovery processes, and in particular, carbon dioxide flooding. During injection of carbon dioxide into wells, contact between the oil and injected CO_2 can change the reservoir fluid properties and also its phase equilibrium conditions, and consequently cause precipitation of heavy and complex hydrocarbon mixtures or asphaltenes (Monger and Trujillo, 1991). Here it is worth mentioning that the type and amount of asphaltene precipitated from the crude oils may differ from one geographical location to another because of non-uniformity of petroleum reservoirs in terms of oil gravity and density (Kazemzadeh et al., 2015).

The boiling point temperature, molecular weight, and specific gravity are accounted as the fundamental properties of heptane–plus (C_{7+}) components of petroleum fractions

or the heavier and more complex mixture of crude oil, so that accurate determination of such important properties is required to characterize and estimate the thermos-physical properties of crude oils. Additionally, application of equations of state (EoS) in the prediction of phase behaviour of reservoir fluid and PVT calculations, which are required for designing and operating the refinery distillation columns reservoir simulation, need accurate and reliable characterization of crude oil, in particular, the heavier components (Riazi, 1989; Riazi, 1997). Furthermore, the volumetric and phase behaviours of volatile oils and gas-condensates are quite sensitive to characterization of heavier petroleum fractions such as heptane–plus components (Riazi, 1997).

A simple definition of vaporization enthalpy is that it is the difference between the enthalpies of the vapor and liquid phases at the same equilibrium pressure and temperature. Vaporization enthalpy (ΔH^{vap}) is the energy needed to transform a quantity of liquid substance into a gas phase at its boiling point temperature (Parhizgar et al., 2013). As a result, the vaporization enthalpy for petroleum fractions and hydrocarbon components is a property that is used in many chemical disciplines, as well as is in the oil and gas industries. From a thermodynamic perspective, vaporization enthalpies can be applied in processing and transportation facilities for the optimization and design of oil and gas production and for heat flux calculations, as well as for the estimation of some physical phenomena like the solubility parameters of hydrocarbons (Mohammadi and Richon, 2007). The vaporization enthalpy of pure components, in particular hydrocarbons and petroleum fractions, is a key and fundamental thermodynamic property which is related to the specific gravity (S), boiling point temperature (T_b), and molecular weight (M) through various thermodynamic relationships.

Natural gas, as a multi-component mixture, is composed of broadly different components, methane as the key component along with more useful components such as carbon dioxide (CO_2), nitrogen (N_2), ethane (C_2H_6), propane (C_3H_8), and weightier hydrocarbon components (Sanjari and Lay, 2012a). Natural gas is a clean and cheap source of energy, compared to other hydrocarbon-based materials like oil and coal. It also has a longer predicted future availability than crude oil and coal (BP, 2006). Furthermore, there has been growing understanding of the importance of natural gas as a means of meeting the world energy requirements due to its adaptability, profusion and clean burning (Wang and Economides, 2009). It is therefore important to develop reliable

predictive methods for the physical properties related to natural gas, such as the gas compressibility factor (*z*-factor), for optimal exploitation and usage.

The gas compressibility factor is a key thermodynamic parameter in chemical and petroleum engineering disciplines. Factors such as the phase equilibria of various hydrocarbon and non-hydrocarbon mixtures, PVT behavior, upstream and downstream calculations, material balances, assessment of underground gas reserves, gas reservoir simulations, well-testing analysis and calculations associated with the processing of gasses (Chamkalani et al., 2013c; Heidaryan et al., 2010b). Moreover, the importance and role of *z*-factor cannot be overemphasized in process engineering calculations and in lower complexity simulations within a thermodynamics context.

1.2. Artificial Intelligence

We have seen that on account of economic and technical requirements, petroleum engineers seek a rapid means to obtain accurate values for the properties of petroleum reservoir fluid and that there are some shortcomings in the existing methods. This study sets out to show that reservoir fluid properties can be usefully determined through empirical methods. Reliable and simple predictive techniques, which are simpler than existing approaches and with fewer computations, are possible. To this end, artificial intelligence methods are the chosen means to develop reliable models for the estimation of petroleum reservoir fluid properties.

In recent years, intelligent/smart methods have been progressively employed in petroleum and chemical calculations (Esfahani et al., 2015; Fathinasab et al., 2015; Ghiasi et al., 2013; Hosseinzadeh and Hemmati-Sarapardeh, 2014; Kamari et al., 2015a; Kamari et al., 2014a; Kamari et al., 2015b; Kamari et al., 2014d; Kamari et al., 2014e; Mohaghegh et al.; Nejatian et al., 2014; Talebi et al., 2014; Zendehboudi et al., 2014). Artificial neural networks (ANN), support vector machines (SVM), decision tree (DT), fuzzy logic (FL), and genetic algorithms (GA) are the most commonly-used artificial intelligence (AI) methods which are applied in oil and natural gas reservoirs simulation, production, enhanced oil recovery (EOR), optimization, drilling automation, process control, and data mining (Al-Bulushi et al., 2009; Aminian and Ameri, 2005; Aminzadeh, 2005; Asadisaghandi and Tahmasebi, 2011; Darabi et al., 2010; El-Sebakhy, 2009b; Huang et al., 2003; Irani and Nasimi, 2011; Jafari Kenari and Mashohor, 2013; Jahanandish et al., 2011; Kamari et al.,

2014h; Kamyab et al., 2010b). As a consequence, artificial intelligence techniques have gained popularity in solving complex nonlinear problems (Al-Fattah and Al-Naim, 2009).

The ANN methodology is a fast and accurate method for the prediction of reservoir properties, which can be applied in reservoir modeling and characterization. Artificial neural networks are able to solve complex nonlinear and classification problems, and they can perform prediction and generalization rapidly (Gharbi, 1997). In the presence of a small size of dataset, the ANN technique may lead to an overfitting problem during training/learning phase, which potentially diminishes performance for capability, applicability and generalization (Al-Anazi and Gates, 2012). Although ANN has demonstrated some successful applications for the estimation of reservoir rock and fluid properties (Bhatt and Helle, 2002; Saemi et al., 2007; Tahmasebi and Hezarkhani, 2012), the basic training/learning mathematical algorithm has been designed for application with large sample sizes. Hence, for a given small size of dataset, extensive experiments with various training/learning methods are required to perform an accurate regression (Kaviani et al., 2008).

In recent decades, an important smart technique, namely support vector machine (SVM), has rapidly gained popularity due to its excellent performance in solving complex classification and regression problems (Al-Anazi and Gates, 2012). The SVM technique has found many applications in various fields of science and engineering including petroleum engineering and geology, pattern recognition in medical science, and speech and text detection. (Choisy and Belaid, 2001; Gao et al., 2001; Li et al., 2000; Ma et al., 2001; Van Gestel et al., 2001).

The least square support vector machine (LSSVM) (Suykens and Vandewalle, 1999) is a variation of conventional SVM. The main differences between SVM and LSSVM are that the LSSVM utilizes square errors instead of nonnegative errors in the cost function and equality constraints instead of inequality constraints, as opposed to the conventional SVM (Li et al., 2012). In other words, the least squares support vector machine (LSSVM) is an improved version of the classical SVM approach, which considers equality constraints in place of inequalities for classical SVM (Suykens and Vandewalle, 1999). In the LSSVM models, the problem of local minima does not occur. As a result, the main drawbacks of LSSVM are fewer tuning parameters (typically two parameters) and lack of sparseness for solutions, respectively (Yang et al., 2010).

In recent years, the genetic algorithm (Holland, 1975) mathematical approach has been applied increasingly as a reliable optimization method for various targets in petroleum engineering. More recently, the original form of GA was improved and a new form known as genetic programming (GP) was introduced. In the GP mathematical approach, the solutions are treated as nonlinear structures of parse trees (treated as functions) instead of fixed length binary solutions (Cramer, 1985; Koza, 1992). An additional development was introduced by Ferreira (Ferreira, 2001) who presented the gene expression programming (GEP) scheme as a new modified and variant version of the classical GP approach.

Another important AI technique is the decision tree algorithm which constructs models for the purpose of regression or classification with the structural shape of a tree. It divides a dataset into smaller subsections while a related decision tree is incrementally and simultaneously developed. The outcome is a tree with decision nodes and leaf nodes. A decision node has two or more divisions, each signifying values for the feature tested. A leaf node represents a decision on the target. The top decision node in a tree relates to the best predictor called a root node. Decision trees can process both categorical and numerical data (Erdogan et al., 2001; Heinze et al., 1995; Sethi and Chatterjee, 1977).

1.3. Scope of Study

The current study claims to put forward a more efficient, accurate and reliable predictive model for the determination of petroleum reservoir fluid properties. To this end, the largest possible databank was assembled from two main sources: actual field data collected and previously published data available in the literature. This databank therefore contains a wide range of reservoir rock and fluid properties coming from the various geographical regions of world.

In order to obtain the information for the databank, robust artificial intelligence strategies were utilized to determine petroleum reservoir fluid properties. Additionally, the Leverage methodology and parameter-dependency approach were employed to detect suspended and/or outlier data points existing in the dataset, and to investigate the effects of different input variables on the output parameters (petroleum reservoir fluid properties).

The results obtained from the use of the newly developed models are compared against the actual field and literature data, equations of state, and also previously published empirically-derived correlations available in literature.

Finally, in order to evaluate the accuracy and capability performance of the developed empirical model, and to provide a comparative study, graphical and statistical error analyses are performed.

In summary, these are the steps that were undertaken to conduct the modelling processes:

I. Determine the most important properties associated with petroleum reservoir fluid properties, as well as the parameters related to the oil and gas production operation.

II. Collect the most comprehensive database from different sources, taking into account screening of the duplicated and erroneous data.

III. Select the most appropriate artificial intelligence method for model development with respect to the nature of each petroleum reservoir fluid property.

IV. Evaluate the accuracy and performance capability of the models developed through various graphical and statistical error analyses.

V. Conduct a comparative study of different artificial intelligence methods, and compare the results to previously published empirical correlations, and equations of state.

This study is organized as follows:

A detailed background on petroleum reservoir fluid properties is presented in Section 2, based on a literature review of previously published works.

Section 3 presents the databanks collected in this study, and investigates effective variables for accurate prediction of petroleum reservoir fluid properties.

In Section 4, the computational methodologies pursued in this study for developing reliable methods to determine the petroleum reservoir fluid properties are covered.

The results obtained by means of the newly developed methods are presented and discussed in Section 5. In this section, the results obtained are compared with literature-reported data, as well as previously published empirical correlations, and equations of states. In order to assess the validity of the models developed in this study, graphical and statistical error parameters are also analysed and discussed.

Finally, future prospects and conclusions of the current study are briefly expressed in Section 6.

CHAPTER 2

2. General Background

Growing global demand for oil and its products, reduction of the natural production of oil resources, concerns about the future of hydrocarbon reserves, production optimization topics, and finally, oil prices have, in recent years, led to considerable research effort into reservoir fluid, which roughly comprises oil, gas, and water. As a result, several studies have been published on the prediction of petroleum reservoir fluid properties and production parameters. Furthermore, the large expansion in industrial projects, and population growth, have led to global attention on petroleum engineering, and in particular, reservoirs and production engineering.

Much experimental work, simulation studies, analytical research, and intelligent predictive models have been developed for accurate determination of petroleum reservoir fluid properties. However, because of the complex nature of reservoir fluid, previously published approaches may not be widely applicable in a variety of reservoir conditions.

There is therefore much scope for the development of simple, robust and reliable models for the accurate estimation of petroleum reservoir fluid properties.

2.1. Surfactant Retention

In the recovery of residual oil, the loss of surfactant can reduce the technical and economic feasibility of recovery (Glover et al., 1979; Yassin et al., 2013). The retention of surfactant molecules is therefore a fundamental factor in chemical surfactant flooding. The adsorption of surfactants by porous rocks is dependent on the rock characteristics (mineralogical and morphological), type of surfactant used, and types of electrolytes existing in the solution (Austad, 1993).

Several studies have been undertaken in order to investigate the retention of surfactants. Standnes and Austad (Standnes and Austad, 2000) used 14 different surfactants for spontaneous counter-current imbibition into oil-wet chalk cores by altering wettability. They indicated that cationic surfactants can recover oil. Liu (Liu, 2008) conducted laboratory experiments taking into consideration the type of the rock and surfactant. Results from the study by Liu (Liu, 2008) indicated that anionic surfactants have much lower adsorption into a sandstone surface than non-ionic surfactants. Yassin et al. (Yassin et al., 2013) developed an artificial intelligent model for prediction of surfactant retention by using the least squares support vector machine algorithm. They used experimental dynamic surfactant retention data over a wide range of conditions. Their results indicated that the values predicted by their model were in agreement with experimental surfactant retention data.

For the determination of the retention of surfactant in porous media, there is an empirical correlation proposed by Solairaj et al. (Solairaj, 2011; Solairaj et al., 2012). The correlation is a multi-variable regression used to investigate the impact of different parameters on surfactant retention during chemical flooding under different conditions. Solairaj et al. (Solairaj, 2011; Solairaj et al., 2012) predicted surfactant retention based on the following input parameters, which included average molecular weight of the surfactant solution, maximum effluent pH, reservoir temperature, total acid number of the oil, mobility ratio, salinity of the polymer, drive co-solvent concentration, as follows:

$$R = a_1 TAN + a_2 T + a_3 C_{co-solvent} + a_4 S_{PD} + a_5 pH + a_6 MR + a_7 MW_{Sur} + C \qquad (2.1)$$

where R denotes the retention of surfactant, TAN is total acid number of the oil, T denotes reservoir temperature, $C_{co\text{-}solvents}$ is co-solvent concentration, S_{PD} stands for salinity of the

polymer drive, pH is maximum effluent pH, MR is value of mobility ratio, and MW_{Sur} is average molecular weight of the surfactant solution, as well as some constants as follows:

a_1=-0.0538725; a_2=-0.0001459; a_3=-0.4855773; a_4=0.0000002;

a_5=-0.0275395; a_6=0.0383129; a_7=0.0000072; C=0.4846366

2.2. Dew Point Pressure of Gas Condensate Reservoirs

The accurate prediction of dew point pressure in gas condensate reservoirs is important for the evaluation of their performance, because there is a reduction in the rate of gas condensate production with an increase of liquid (Seyed Mohammad Javad Majidi 2013). A number of researchers have studied the effect of dew point pressure on well productivity, e.g. Fevang and Whitson (Ø. Fevang 1996), Fan et al. (D. Afidick, 1994), Afidick et al. (Fan and Whitson, 1998), Barnum et al. (Barnum, 1995), and Smits et al. (Eilerts, 1942). The studies conclude that there is a considerable reduction in well generation in gas condensate wells under certain conditions, *e.g.* near wellbore condensate aggregation.

The determination of the dew point pressure in gas condensate reservoirs, both experimentally and theoretically, has been investigated by several researchers. For the experimental determination of the DPP, the constant composition expansion (CCE) and constant volume depletion (CVD) are the two most commonly-used laboratory measurement methods (Shokir, 2008). While laboratory measurement of DPP is reliable, it is expensive and time-consuming. Hence, there is a preference to determine DPP using an empirical method called equations of state (EoS) (Shokir, 2008). EoSs are usually not, however, able to accurately simulate the phase behavior of light oil and gas condensate reservoirs, particularly in the retrograde region (Hadi Rostami-Hosseinkhani 2014).

Over the years, many research studies have attempted to provide a global model for the prediction of DPP in gas condensate systems, on the basis of temperature, hydrocarbon composition, and C_{7+}. In 1942, Kurata and Katz (F. Kurata, 1942) developed a correlation to predict the critical properties of volatile hydrocarbon mixtures. Unfortunately, they neglected the effect of composition due to the limited DPP data used

for the model development. Eilerts and Smith (Eilerts, 1942) proposed a relationship between temperature, pressure, composition, boiling point of the fluid, and gas oil ratio based on research in the Palam field. In 1945, Olds et al.(R. Olds, 1945) developed a correlation to predict the dew point pressure (in graphical and tabular forms) by using the characteristics of oil and gas samples obtained from the primary separator of a well in the Paloma field. They also studied the impact of the elimination of intermediate molecular weight on DPP. They showed that the intermediate molecular weight components have a significant influence on DPP.

Olds et al. (R. Olds, 1949) experimentally studied the volumetric behavior for various mixtures of gas condensate samples which were collected from the San Joaquin Valley field. The correlation developed by them provided a relationship between the retrograde DPP and gas-oil ratio, temperature, and stock tank API oil gravity. The results obtained showed that the effect of temperature was minimal in comparison to the influence of modifying the compositions. Modification of the composition was undertaken by eliminating the intermediate components (Elsharkawy, 2002b). In 1950, Reamer and Sage (H. Reamer, 1950) investigated existing correlations, with respect to higher gas-oil ratio samples by combining five different pairs of fluid from a typical field in Louisiana. In their study, the effect of temperature and gas-oil ratio on DPP was investigated. They concluded that the complexity of the effect of composition on DPP is the main reason for a lack of a global model for predicting DPP.

In 1952, Organic and Golding (E. Organick, 1952) studied the dew point pressure in condensate gas and volatile-oil mixtures. They introduced a simple correlation in the form of working charts which had an error of approximately 8% (Elsharkawy, 2002b). The correlation developed was not able to describe some materials like pure components and non-complex mixtures. Nemth and Kennedy (L. Nemeth, 1967) developed an extended relationship between dew point pressure, temperature, composition, and characteristics of the C_{7+} fraction of the hydrocarbon fluid. They used multiple-variable regression analysis in their correlation development. In 1996, Crogh (Crogh, 1996) improved the Nemeth and Kennedy (L. Nemeth, 1967) correlation. Their correlation enabled better prediction of DPP, neglecting reservoir temperature. Humoud and Al-Marhoun (Humoud, 2001) developed an empirical model using different gas condensate samples extracted from the Middle East. The correlation developed is a relationship

between the DPP of a gas condensate fluid with its reservoir temperature, the primary separator pressure and temperature, primary separator gas-oil ratio, pseudo reduced pressure and temperature, relative densities of separator, and heptane-plus fraction.

In 1996, Carlson and Cawston (M.R. Carlson, 1996) studied the effect of H_2S on DPP. They concluded that an increase in H_2S content decreases the volume of liquid drop out. Marruffo et al. (I. Marruffo, 2001) proposed a correlation which used 146 PVT analyses data from Western Venezuela (Anaco) fields to determine the DPP in gas condensate reservoirs. In 2002, Elsharkawy (Elsharkawy, 2002b) developed a relationship between DPP of the gas condensate reservoirs and some properties, including temperature, molecular weight, composition of hydrocarbon fluid and specific gravity of the C_{7+} components. They used 340 experimentally measured data within a pressure range of 1560-11830 psi and temperature ranging from 40 to 340 °F. Gonzales et al. (A. González, 2003) used neural network modeling to predict DPP. The reservoir temperature, hydrocarbon and non-hydrocarbon composition, molecular weight and specific gravity of the C_{7+} were applied as the inputs of the network developed. In 2008, Shokir (Shokir, 2008) proposed a model using mathematical genetic programming, neglecting the effect of specific gravity of C_{7+}. The study showed that the model developed is more acceptable (precise) compared to previous methods. Sarkar et al. (R. Sarkar, 1991) proposed equations of state to model the behavior of the reservoir fluid phase, but they could not accurately simulate the complex hydrocarbon behavior, such as the retrograde gas condensate reservoir phase.

It appears that the most commonly used methods for the calculation of dew point pressure are Elsharkawy (Elsharkawy, 2002a), Shokir (Shokir, 2008), and Nemeth and Kennedy (Nemeth and Kennedy, 1967) as follows:

Elsharkawy (Elsharkawy, 2002a)

$$P_d = A_0 + A_1 T + A_2 z_{H_2S} + A_3 z_{CO_2} + A_4 z_{N_2} + A_5 z_{C_1} + A_6 z_{C_2} + A_7 z_{C_3} + A_8 z_{C_4} + A_9 z_{C_5} + A_{10} z_{C_6} + A_{11} z_{C_{7+}}$$
$$+ A_{12} MW_{C_{7+}} + A_{13} SG_{C_{7+}} + A_{14} (z_{C_{7+}} MW_{C_{7+}}) + A_{15} \left(\frac{MW_{C_{7+}}}{SG_{C_{7+}}}\right) + A_{16} \left(\frac{z_{C_{7+}} MW_{C7+}}{SG_{C_{7+}}}\right) + A_{17} \left(\frac{z_{C_{7+}}}{z_{C_1} + z_{C_2}}\right)$$
$$+ A_{18} \left(\frac{z_{C_{7+}}}{z_{C_3} + z_{C_4} + z_{C_5} + z_{C_6}}\right)$$

(2.2)

Shokir (Shokir, 2008)

$$P_d = B_1 + B_2 + B_3 + B_4 \tag{2.3}$$

where

$$B_1 = 201.875481(z_{C_{7+}}(((T_r(((z_{C_3} - (z_{H_2S} - Z_{CO_2})) - (z_{C_6} - (Z_{CO_2} - z_{C_4}))) - z_{C_2})) - ((z_{C_4}(((Z_{CO_2} - z_{C_4}) - (MW_{C_{7+}} - z_{N_2})) - (MW_{C_{7+}}{}^2 z_{C_5}))) - z_{C_{7+}})) - (z_{H_2S} - ((z_{N_2}(T(z_{C_1}{}^2 - z_{C_{7+}}))) - (MW_{C_{7+}} - (z_{C_2} - z_{H_2S})))))) + 38456.87953z_{C_5}$$

(2.4)

$$B_2 = 0.000007((T((((z_{CO_2} - MW_{C_{7+}}) - z_{C_{7+}})((T - MW_{C_{7+}}) - (z_{CO_2} - T))) - ((z_{H_2S} - T)((MW_{C_{7+}} - z_{C_3})MW_{C_{7+}})))z_{N_2}) + 225500.9399z_{C_5}$$

(2.5)

$$B_3 = 120586.9719(z_{C_1}((((z_{H_2S}z_{C_3}) - (z_{C_5} - z_{C_{7+}}))z_{H_2S}) - (((z_{C_{7+}} - z_{C_1})(z_{C_{7+}} - z_{C6})) - (z_{H_2S}z_{N_2}{}^2)))) + 72.6908MW_{C_{7+}}$$

(2.6)

$$B_4 = -1962.40851(z_{C_5}(MW_{C_{7+}} - z_{C_1}{}^2)) - 253385.67764((z_{C_{7+}}((z_{CO_2}z_{C_3}) - (z_{C_4} - z_{C_{7+}})))(z_{CO_2}(z_{C_3}(z_{C_3} - MW_{C_{7+}})))) - 13358.59271z_{C_4} + 4676.933602z_{C_2} - 6567.9 \tag{2.7}$$

Nemeth and Kennedy (Nemeth and Kennedy, 1967)

$$\ln(P_d) = A_1(z_{C_2} + z_{CO_2} + z_{H_2S} + z_{C_6} + 2(z_{C_3} + z_{C_4}) + z_{C_5} + 0.4z_{C_1} + 0.2z_{N_2}) + A_2 SG_{C_{7+}} + A_3\left(\frac{z_{C_1}}{z_{C_{7+}} + 0.002}\right)$$
$$+ A_4 T + A_5(z_{C_{7+}} MW_{C_{7+}}) + A_6(z_{C_{7+}} MW_{C_{7+}})^2 + A_7(z_{C_{7+}} MW_{C_{7+}})^3 + A_8\left[\frac{MW_{C_{7+}}}{SG_{C_{7+}} + 0.0001}\right] + A_9\left[\frac{MW_{C_{7+}}}{SG_{C_{7+}} + 0.0001}\right]^2$$
$$+ A_{10}\left[\frac{MW_{C_{7+}}}{SG_{C_{7+}} + 0.0001}\right]^3 + A_{11}$$

(2.8)

where in the equations P_d denotes the dew point pressure in gas condensate reservoirs, MW_{C7+} is the molecular weight of the heptane-plus fraction, SG_{C7+} is the specific gravity of the heptane plus fraction, T_R is the reservoir temperature and $z_{C_1}, z_{C_2}, z_{C_3}, z_{C_4}, z_{C_5}, z_{C_6} z_{C_{7+}}, z_{CO_2}, z_{N_2}, z_{H_2S}$ are compositions of methane, ethane, propane, butanes, pentanes, hexanes, heptane-plus, carbon dioxide, nitrogen, and hydrogen sulfide, respectively. Furthermore, the adjustable parameters of Elsharkawy (Elsharkawy, 2002a), and Nemeth and Kennedy (Nemeth and Kennedy, 1967) methods are presented in **Table 2. 1**.

Table 2. 1 List of the adjustable parameters applied in the Elsharkawy (Elsharkawy, 2002a), and Nemeth and Kennedy (Nemeth and Kennedy, 1967) methods.

Parameter	Elsharkawy's model	Nemeth and Kennedy's model
A_0	4268.85	
A_1	0.094056	-2.0623054
A_2	-7157.87	6.6259728
A_3	-4540.58	$-4.4670559*10^{-3}$
A_4	-4663.55	$1.0448346*10^{-4}$
A_5	-1357.56	$3.2673714*10^{-2}$
A_6	-7776.1	$-3.6453277*10^{-3}$
A_7	-9967.99	$7.4299951*10^{-5}$
A_8	-4257.1	$-1.1381195*10^{-1}$
A_9	-1417.1	$6.2476497*10^{-4}$
A_{10}	691.5298	$-1.0716866*10^{-6}$
A_{11}	40660.36	$1.746622*10^{+1}$
A_{12}	205.26	
A_{13}	-7260.32	
A_{14}	-352.413	
A_{15}	-114.519	
A_{16}	8.133	
A_{17}	94.916	
A_{18}	238.252	

2.3. Bubble Point Pressure and Oil Formation Volume Factor

Several PVT empirical correlations to determine the fluid's physical properties viz. bubble point pressure and oil formation volume factor, have been developed between 1947 and the present. Standing (Standing, 1947a) and Katz (Katz, 1942) first provided PVT empirical correlations graphically for the estimation of PVT properties such as OFVF and bubble point pressure in 1947 and 1950, respectively. The Standing (Standing, 1947a) correlation is based on experimental tests conducted on 105 cores from 22 crude oils in California State and also the Katz (Katz, 1942) method employs bubble point

pressure, oil API gravity, gas specific gravity, reservoir temperature and solution gas oil ratio in order to calculate the oil formation volume factor. Vazquez and Beggs (Vazquez and Beggs, 1980) presented a new PVT empirical correlation to estimate solution gas–oil ratio, viscosity associated with under-saturated oil reservoirs and oil formation volume factor based on the laboratory measurement of 600 cores collected from various regions of world. Subsequently, the results indicated that their proposed empirical correlation for the estimation of OFVF has an average error of 4.7%.

Glaso (Glaso, 1980) performed regression and graphical analyses for the oil PVT properties by using data related to 45 core samples mostly collected from the North Sea region. Their results, related to oil formation volume factor and bubble point pressure, were of an average errors equal to 1.28 and 20.43 %, respectively. Al-Marhoun (Al-Marhoun, 1992) proposed a PVT empirical correlation for oil formation volume factor based on experimental PVT data mostly extracted from North America and the Middle East, Dokla and Osman (Dokla and Osman, 1992) used experimental PVT data from UAE for development of bubble point pressure and oil formation volume factor correlations, and Petrosky and Farshad (Petrosky Jr and Farshad, 1998) developed under-saturated isothermal oil compressibility, oil formation volume factor and solution gas–oil ratio correlations using PVT data from the Gulf of Mexico.

Arabloo et al. (Arabloo et al., 2014) implemented two constrained multivariable search techniques including a generalized reduced gradient algorithm and successive linear programming for the development of two empirical correlations to determine the oil formation volume factor and bubble point pressure. To pursue their objective, they utilized experimental data associated with various geographical domains worldwide. The results indicated that their OFVF and bubble point pressure correlations had average relative errors equal to 2.24 and 18.9 %, respectively. Recently, smart techniques have been increasingly employed to predict PVT properties (Boukadi et al., 1999; Elsharkawy, 1998; Gharbi and Elsharkawy, 1997a; Gharbi and Elsharkawy, 1997b). However, these intelligent methods have some drawbacks. For example, a symbolic equation, necessary in a large dataset for prediction and over-fitting problems, was not given.

2.4. Saturated, Under-Saturated, and Dead Oil Viscosities

The viscosity of reservoir oil depends on many thermo-physical factors including reservoir pressure, reservoir temperature, solution gas–oil ratio properties, bubble point pressure properties, the gravity of gas and oil, and the composition of oil mixture (Riazi and Al-Sahhaf, 1996; Torabi et al., 2011). Normally, the measurement of reservoir oil viscosity is undertaken by conducting experimental methods that simulate reservoir conditions under particular reservoir temperatures. Nevertheless, experimentally measuring the reservoir oil viscosity at different temperatures is sometimes impossible and/or unprofitable due to the high price of the sampling equipment and the related tests (Torabi et al., 2011). Therefore, empirically derived methods can be useful in estimating the reservoir oil viscosity where experimental data is not available.

In regard to the pressure issue, there are several empirical and/or semi-empirical correlations available for calculation of dead oil viscosities, saturated, and under-saturated, which will be explained in more detail in the next section. Moreover, there are some models for the prediction of viscosity on the basis of the corresponding states method (Johnson, 1991; Johnson et al., 1987; Teja and Rice, 1981). The corresponding states-based method has not found acceptance because it requires complex calculations/computations as well as information on the composition of fluid for its viscosity estimation (Elsharkawy and Alikhan, 1999; Naseri et al., 2005).

Additional methods include smart techniques like the neural networks and support vector machines approaches, which are applied to determine reservoir oil viscosities. But these methods sometimes have over-fitting problems and also require large databases and finally, do not give a symbolic equation for future applications (Al-Anazi and Gates, 2010; Al-Anazi and Gates, 2012; Parhizgar et al., 2013). Obanijesu and Omidiora (Obanijesu and Omidiora, 2009) developed an artificial neural network model for the determination of oil viscosities of Nigerian crudes. Hemmati-Sarapardeh et al. (Hemmati-Sarapardeh et al., 2014b) proposed a smart predictive model based on the least squares version of the support vector machines mathematical scheme for the prediction of Iranian oil field viscosities. The results illustrated agreements between the experimental data and the results.

The list of the previously published dead, saturated, and under-saturated oil viscosity models investigated in this study as well as their ranges and formulas have been summarized in **Tables 2. 2-4**.

Table 2. 2 List of the previously published dead oil viscosity models investigated in this study as well as their ranges and formulas.

Method	Origin of Data	T (K)	API	μ_{od} (cP)	Formula
Beal (Beal, 1946)	US	310-394	10.1-52	0.865-1550	$\mu_{od} = \left(0.32 + 1.8 \times \frac{10^7}{API^{4.53}}\right)\left(\frac{360}{T+200}\right)^{x}; x = e^{[2.302585\left(0.43+\frac{8.33}{API}\right)]}$
Beggs and Robinson (Beggs and Robinson, 1975)	-	294-419	16-58	-	$\mu_{od} = 10^{x} - 1; x = e^{[3.0324-0.02023API]}\ T^{-1.163}$
Glaso (Glaso, 1980)	North Sea	283-422	20-48	0.6-39	$\mu_{od} = \left(\frac{3.141 \times 10^{10}}{T^{3.44}}\right)\log(API)^{[0.313\log(T)-36.447]}$
Kaye (Kaye, 1985)	Offshore California	334-412	7-41	-	$\mu_{od} = 10^{[T^{-0.65}10^{(2.203-0.0254API)}]} - 1, API \leq 12;\ \mu_{od} = 10^{[T^{-0.65}10^{(2.305-0.03354API)}]} - 1, API > 12;$
Al-Khafaji (Al-Khafaji et al., 1987)	-	289-422	15-51	-	$\mu_{od} = \frac{10^{(4.9563-0.00488T)}}{(API + \frac{T}{30} - 14.29)^{2.709}}$
Petrosky (Petrosky, 1990)	Gulf of Mexico	319-415	25-46	0.72-10.25	$\mu_{od} = \frac{2.3511 \times 10^7}{T^{2.10255}}\log(API)^{[4.59388\log(T)-22.82792}$
Egbogah and Ng (Egbogah and Ng, 1990)	-	288-353	5-58	-	$\mu_{od} = 10^{x} - 1; x = 10^{[1.8653-2.5086\times10^{-2}\ API-0.56441\log(T)]}$
Labedi (Labedi, 1992)	Libya	311-425	32-48	0.66-4.79	$\mu_{od} = 10^{9.224}\ API^{-4.7013}\ T^{-0.6739}$
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994b)	Worldwide	300-433	14-59	0.5-586	$\mu_{od} = \frac{1.6 \times 10^9}{T^{2.8177}}\log(API)^{[5.7526\log(T)-26.9718]}$

Bennison (Bennison, 1998)	North Sea	277-422	11-20	6.4-8396	$\mu_{od} = 10^{[-0.8021API+23.8765]}\ T^{[0.31458API-9.21592]}$
Elsharkawy and Alikhan (Elsharkawy and Alikhan, 1999)	Middle East	311-422	20-48	0.6-33.7	$\mu_{od} = 10^{x} - 1; x = e^{[2.16924-0.02525API-0.68875\ \log(T)]}$
Hossain *et al*. (Hossain et al., 2005)	Worldwide	273-375	7-22	12-451	$\mu_{od} = 10^{[-0.71523API+22.13766]}\ T^{[0.269024API-8.268047]}$
Naseri *et al*. (Naseri et al., 2005)	Iran	314-421	17-44	0.75-54	$\mu_{od} = 10^{[11.2699-4.2699\ \log(API)-2.052\ \log(T)]}$
Alomair *et al*. (Alomair et al., 2011)	Kuwait	293-433	10-20	1.78-11360	

Table 2. 3 List of the previously published saturated oil viscosity models investigated in this study as well as their ranges and formulas.

Method	Source of Data	Solution GOR (SCF/STB)	Saturation Pressure (MPa)	μ_{od}, cP	Formula
Chew and Connally I (Chew and Connally Jr, 1959)	US	51-3544	0.91-38.92	0.37-50	$\mu_{ob} = A\mu_{od}^{B}; A = 0.2 + \frac{0.8}{10^{0.00081R_s}}$; $B = 0.43 + \frac{0.57}{10^{0.00072R_s}}$
Chew and Connally II (Chew and Connally Jr, 1959)	US	51-3544	0.91-38.92	0.37-50	$\mu_{ob} = A\mu_{od}^{B}; A = 10^{[R_s(2.2\times10^{-7}R_s - 7.4\times10^{-4})]}$; $B = 0.68\times10^{[-8.62\times10^{-5}R_s]} + 0.25\times10^{[-1.1\times10^{-3}R_s]} + 0.062\times10^{[-3.74\times10^{-3}R_s]}$
Chew and Connally III (Chew and Connally Jr, 1959)	US	51-3544	0.91-38.92	0.37-50	$\mu_{ob} = A\mu_{od}^{B}$ $A = 0.987583 - 0.1746773\times10^{-2}R_s + 0.2067531\times10^{-5}{R_s}^2 - 0.1310529\times10^{-8}{R_s}^3 + 0.3229416\times10^{-12}{R_s}^4$ $B = 0.9900216 - 0.112183\times10^{-2}R_s + 0.1427879\times10^{-5}{R_s}^2 - 0.9440539\times10^{-9}{R_s}^3 + 0.2312365\times10^{-12}{R_s}^4$
Beggs and Robinson (Beggs and Robinson, 1975)	-	20-2070	0.91-36.30	-	$\mu_{ob} = A\mu_{od}^{B}; A = \frac{10.715}{(R_s+100)^{0.515}}$; $B = \frac{5.44}{(R_s+150)^{0.338}}$
Al-Khafaji (Al-Khafaji et al., 1987)	-	0-2100	-	-	$\mu_{ob} = A\mu_{od}^{B}; X_1 = \log(R_s)$ $A = 0.247 + 0.2824X_1 + 0.5657X_1^2 - 0.4065X_1^3 + 0.0631X_1^4$ $B = 0.894 + 0.0546X_1 + 0.07667X_1^2 - 0.0736X_1^3 + 0.01008X_1^4$
Khan *et al*. (Khan et al., 1987)	Saudi Arabia	24-1901	0.74-29.75	0.13-77.4	$\mu_{ob} = \frac{0.09\gamma_g^{0.5}}{\frac{1}{R_{sb}^3}[\frac{T+459.67}{459.67}]^{4.5}(1-\gamma_o)^3}$; if $P < P_b \rightarrow \mu_o = \frac{\mu_{ob}\, e^{[-2.5\times10^{-4}(P-P_b)}}{(\frac{P}{P_b})^{0.14}}$
Petrosky (Petrosky, 1990)	Gulf of Mexico	21-1855	10.85-65.85	0.21-7.4	$\mu_{ob} = A\mu_{od}^{B}; A = 0.1651 + \frac{0.6165}{10^{(6.0866\times10^{-4}R_s)}}$; $B = 0.5131 + \frac{0.5109}{10^{(1.1831\times10^{-3}R_s)}}$
Labedi (Labedi, 1992)	Libya	13-3533	0.41-43.83	0.115-3.72	$\mu_{ob} = \frac{10^{(2.344-0.03542API)}\,\mu_{od}^{0.6447}}{P_b^{0.426}}$; at $P < P_b \rightarrow \mu_o = \frac{\mu_{ob}}{1-(10^{-3.876}\,P_b^{05423}API^{1.1302}\,(1-\frac{P}{P_b})}$

Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994b)	Worldwide	2.3-572	0.10-41.74	0.1-6.3	$\mu_{ob} = -0.06821 + 0.9824X_2 + 4.034 \times 10^{-4}X_2^2$; $X_1 = 0.43 + 0.5156 \times 10^{[-8.1\times10^{-4}R_s]}$; $X_2 = [0.2001 + 0.8428 \times 10^{[-8.45\times10^{-4}R_s]}]\mu_{od}^{X_1}$
Elsharkawy and Alikhan (Elsharkawy and Alikhan, 1999)	Middle East	10-3600	0.69-25.51	0.05-21	$\mu_{ob} = A\mu_{od}^{B}; A = \frac{1241.935}{(R_s + 641.026)^{1.12410}}$; $B = \frac{1768.841}{(R_s + 1180.335)^{1.06622}}$
Hossain *et al*. (Hossain et al., 2005)	Worldwide	19-493	0.83-43.24	3.6-360	$\mu_{ob} = A\mu_{od}^{B}$ $A = 1 - 0.001718831R_s + 1.58081 \times 10^{-6}{R_s}^2$; $A = 1 - 0.002052461R_s + 3.47559 \times 10^{-6}{R_s}^2$
Naseri *et al*. (Naseri et al., 2005)	Iran	255-4116	2.89-40.68	0.11-18.15	$\mu_{ob} = 10^{1.1145}P_b^{-0.4956}\mu_{od}^{0.9961}$
Bergman and Sutton (Bergman and Sutton, 2007)	Worldwide	6-6525	0.45-71.02	0.21-4277	$\mu_{ob} = A\mu_{od}^{B}$; $A = \frac{1}{1 + (\frac{R_s}{344.198})^{0.855344}}$; $B = \frac{0.617677}{1 + (\frac{R_s}{567.953})^{0.819326}}$

Table 2. 4 List of the previously published under-saturated oil viscosity models investigated in this study as well as their ranges and formulas.

Method	Origin of Data	P (MPa)	P_b (MPa)	μ_{ob} (cP)	μ_o (cP)	Formula
Beal (Beal, 1946)	USA	-	-	0.142-127	0.16-315	$\mu_o = \mu_{ob} + [0.001(P - P_b)](0.024\mu_{ob}^{1.6} + 0.038\mu_{ob}^{0.56})$
Vazquez and Beggs (Vazquez and Beggs, 1980)	Worldwide	0.87-65.50	-	-	0.117-148	$\mu_o = \mu_{ob}\, e^{[(5.50318\times10^{-5}+3.77163\times10^{-5}\mu_{ob}^{0.278})(P-P_b)]}$
Khan *et al.* (Khan et al., 1987)	Saudi Arabia	-	0.74-33.05	0.13-77.4	0.13-71	$\mu_o = \mu_{ob}\, e^{[9.6\times10^{-5}(P-P_b)]}$
Petrosky (Petrosky, 1990)	Gulf of Mexico	11.03-70.67	10.85-65.86	0.211-3.546	0.22-4.09	$\mu_o = \mu_{ob} + 1.3449\times10^{-5}(P - P_b)10^{X_2}; X_1 = \log(\mu_{ob}); X_2 = -1.0146+1.3322X_1 - 0.4876X_1^2 - 1.15036X_1^3$
Labedi (Labedi, 1992)	Libya	-	0.41-43.84	0.115-3.72	-	$\mu_o = \mu_{ob} + \frac{\mu_{od}^{0.9036} P_b^{0.6151}}{10^{(2.488+0.01976\ API)}}(\frac{P}{p_b} - 1)$
Orbey and Sandler (Orbey and Sandler, 1993)	-	5.10-100.00	-	0.217-3.1	0.225-7.3	$\mu_o = \mu_{ob}\, e^{[\alpha(P-P_b)]}; \alpha = 6.89\times10^{-5}$
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994b)	Worldwide	0.17-41.47	0.17-32.92	0.168-184.86	0.168-517.03	$\mu_o = 1.0081\mu_{ob} + 1.127\times10^{-3}(P - P_b)(-6.517\times10^{-3}\mu_{ob}^{1.8148} + 0.038\mu_{ob}^{1.59})$
Elsharkawy and Alikhan (Elsharkawy and Alikhan, 1999)	Middle East	8.87-68.94	-	-	0.2-5.7	$\mu_o = \mu_{ob} + \frac{10^{-2.0771}(P - P_b)\mu_{od}^{1.19279}}{\mu_{ob}^{0.40712} P_b^{0.7941}}$
Hossain *et al.* (Hossain et al., 2005)	Worldwide	2.07-23.44	0.83-43.24	3.6-360	3-517	$\mu_o = \mu_{ob} + [0.004481(P - P_b)](0.555955\mu_{ob}^{1.068099} - 0.527737\mu_{ob}^{1.063547})$

2.5 Solution Gas–Oil Ratio

Over the years, various empirical methods have been reported for the determination of reservoir fluid properties related to oil samples from different geographical locations. To this end, in one of the first attempts, Elam (Elam, 1957) in 1957 proposed a correlation for the estimation of saturation pressure as a function of temperature, gas specific gravity, oil gravity and solution GOR using as a basis of 231 data points for Texas crude oil. One year later, Lasater (Lasater, 1958) presented a bubble point-pressure correlation for black oil data taken from Canada, western and mid-continental United States and South America. His model was developed using 158 samples of 137 various crude oils. He reported an average error of 3.8% for his model. He also observed that the existence of CO_2 in crude oil samples results in an increment in the saturation pressure. Vasquez and Beggs (Vazquez and Beggs, 1980) proposed some empirically derived methods for the estimation of reservoir fluid properties using a universal databank collected from various regions of the world. Moreover, they separated the experimentally obtained data into two classes. The first group contained oils with gravities less than 30 °API. The second group contained oils with gravities of more than 30 °API. In contrast with Lasater's results (Lasater, 1958), they found that CO_2 content decreases the saturation pressure.

In 1983, Ostermann (Ostermann et al., 1983) developed two correlations for the estimation of the saturation pressure of crude oil samples taken from different regions in Alaska based on a limited number of data points. Al-Marhoun (Al-Marhoun, 1988) developed an empirical correlation applying data gathered from the Middle East region. In 1990, Rollins et al. (Rollins et al., 1990) proposed an empirically derived method to calculate the stock-tank gas–oil ratio as a function of oil API gravity, separator pressure and temperature, and gas gravity. In the same year, Sutton and Farshad (Sutton and Farshad, 1990) reviewed several PVT correlations and compared the accuracy of several PVT parameters in each model for application in the Gulf of Mexico. In their study, Glaso's correlations (Glaso, 1980) provided acceptable results for calculation of saturation pressure, solution GOR, and OFVF. They reported that Vazquez and Begg's correlations (Vazquez and Beggs, 1980) had higher accuracy for solution GOR for more than 1400 SCF/STB and saturation pressures of more than 7000 psi.

In 1992, Dokla and Osman (Dokla and Osman, 1992) studied 51 crude oil samples from UAE and developed a new empirical methods for OFVF, saturation pressure and solution GOR. They reported that PVT correlations should be derived using local data sets because universal correlations are not always accurate enough. In addition, Omar and Todd (Omar and Todd, 1993) developed models for OFVF and saturation pressure on the basis of Standing's correlations (Standing, 1947b) using 93 PVT datasets from Malaysian oil reservoirs. Their models showed better accuracy for Malaysian oil samples. Furthermore, Petrosky and Farshad (Petrosky Jr and Farshad, 1993) proposed some empirically derived methods for the determination of reservoir fluid properties using data collected from the Gulf of Mexico. They showed that the empirical methods proposed outperformed other methods developed for the Gulf of Mexico, including those of Standing (Standing, 1947b), Vasquez and Beggs (Vazquez and Beggs, 1980), Glaso (Glaso, 1980), and Al-Marhoun (Al-Marhoun, 1988). Elsharkawy et al. (Elsharkawy et al., 1995) also compared different correlations to characterize Kuwaiti crude oils using a limited number of oil samples.

Ghetto et al. (Ghetto et al., 1994) proposed empirical methods for the calculation of saturation pressure, solution GOR, OFVF, oil compressibility, and oil viscosity for heavy and extra-heavy oils. The data used in developing the correlations came from reservoir fluid samples extracted from the Mediterranean Basin, Africa, and the Persian Gulf. In 1998, Khairy et al. (Khairy et al., 1998) developed empirical methods for the estimation of saturation pressure and bubble point OFVF. They compared their model with nine published correlations. In 1999, Velarde and McCain (Velarde et al., 1999) developed a set of empirical methods for calculating solution GOR and OFVF, and modified OFVF using 195 laboratory tests. In 2007, Mazandarani and Asghari (Mazandarani and Asghari, 2007) tuned Al-Marhoun 's attempted to find a modified correlation (Al-Marhoun, 1988) for Iranian field data using about fifty fluid samples collected from different Iranian oil fields. In 2008, Taghaz et al. (Taghaz et al., 2008) tested the accuracy of PVT correlations to determine the solution GOR of Libyan oils using about 1600 data points from different oil fields in the Sirte basin. They concluded that no correlation is suitable for Libyan oils. In 2012, Shafiie et al. (Shafiie et al., 2012) optimized Standing (Standing, 1947b) and McCain correlations for solution GOR and OFVF, based on Iranian crude oil samples, and developed a new model using Genetic Algorithms. In 2014 Arabloo et al. (Arabloo et al.,

2014) developed simple and accurate empirical methods for the prediction of saturation pressure and OFVF using a large databank compiled from various geographical locations.

Here, it is worth mentioning that, in addition to empirical methods, smart techniques have been implemented for the estimation of reservoir fluid properties and petroleum engineering problems (Esfahani et al., 2015; Hosseinzadeh and Hemmati-Sarapardeh, 2014; Kamari et al., 2015a; Kamari et al., 2014a; Kamari et al., 2015b; Kamari et al., 2014c; Kamari et al., 2014d; Kamari et al., 2014e; Zendehboudi et al., 2013b). **Tables 2. 5** and **2. 6** summarize the most important methods available in the literature in order to determine the petroleum reservoir fluid properties.

Table 2. 5 Author defined ranges for bubble point pressure, solution GOR, OFVF, and compressibility correlations

Petroleum reservoir fluid properties	Standing	Lasater	Glaso	Kartoatmodjo	Vasques-Beggs	Al-Marhoun	Rollins-McCain Creeger	Petrosky-Farshad	Lebadi
Tank-oil gravity (^{0}API)	16.5 to 63.8	17.9 to 51.1	22.3 to 48.1	14.4 to 58.95	15.3 to 59.5	19.4 to 44.6	18 to 53.5	16.3 to 45	32.2 to 48
Bubblepoiot pressure (psia)	130 to 7000	48 to 5780	165 to 7142	0 to 6040	15 to 6055	130 to 3573	–	1574 to 6523	520 to 6358
Reservoir temperature (^{0}F)	100 to 258	82 to 272	80 to 280	75 to 320	170 (mean)	74 to 240	–	114 to 288	128 to 306
OFVF at bubblepoint (bbl/STB)	1.024 to 2.15	-	1.025 to 2.588	1.022 to 2.747	1.028 to 2.226	1.032 to 1.997	–	1.1178 to 1.6229	1.088 to 2.92
Solution GOR (scf/STB)	20 to 1425	3 to 2905	90 to 2637	0 to 2890	0 to 2199	26 to 1602	–	217 to 1406	–
Separator gas gravity (air=I)	–	–	–	0.4824 to 1.668	0.511 to 1.351	–	0.579 to 1.124	-	–
Total surface gas gravity (air-1)	0.59 to 0.95	0.574 to 1.223	0.65 to 1.276	–	–	0.752 to 1.367	–	0.5781 to 0.8519	–
Separator pressure (psia)	265 to 465	15 to 605	415 (mean)	100	60 to 565	–	29.7 to 314.7	–	34.7 to 789.7
Separator temperature (^{0}F)	100 (mean)	34to 106	125 (mean)	38 to 294	76 to 150	–	60 to 150	–	60 to 220
Reservoir pressure (psia)	–	–	–	10 to 6000	141 to 9515	20 to 3573	-	1700 to 10692	–
Stock-tank GOR (scf/STB)	–	–	–	–	–	–	4 to 220	–	–
Separator GOR (scf/STB)	–	–	–	–	–	–	12 to 1742	–	–

Table 2. 6 Average errors reported by authors in their original papers.

	Bubblepoiot pressure	Solution GOR	OFVF	Isothermal compressibility	Dead-oil viscosity	Gas-saturated oil viscosity	Undersaturated oil viscosity
Author	Standing	Vasquez-Beggs	Standing	Labedi	Beggs-Robinson	Beggs-Robinson	Vasquez-Beggs
Average error (%)	4.8	-0.7	1.17	3	-0.64	-1.83	-7.541
Author	Lasater	Kartoatmodjo	Vasquez-Beggs	Kartoatmodjo	Egbogah-Jack	Kartoatmodjo	Kartoatmodjo
Average error (%)	3.8	23.2	4.7	23.7	-5.13	16.1	6.9
Author	Glaso	Rollins-McCain Creeger	Glaso	Petrosky-Farshad	Kartoatmodjo	Labedi	Majeed-Kattan-Salman
Average error (%)	1.28	3	-0.43	6.66	39.6	-2.38	1.188
Author	Al-Marhoun		Kartoatmodjo		Labedi		Labedi
Average error (%)	3.66		2		-2.61		-3.1
Author	Kartoatmodjo						
Average error (%)	20.2						

2.6. Asphaltene Precipitation

A significant number of attempts have been made by researchers to address asphaltene precipitation during petroleum production. Loureiro et al. (Loureiro et al., 2015) studied the effect of carbon dioxide and n-heptane on the behavior of asphaltene precipitated. To this end, they employed ultraviolet–visible (UV–vis) spectrometry to monitor phase behaviour of asphaltene precipitation. Hemmati-Sarapardeh et al. (Hemmati-Sarapardeh et al., 2013a) developed a SARA fraction based model using a least square support vector machine (LSSVM) algorithm for the estimation of asphaltene precipitation of Iranian crudes. They indicated that the results obtained by their model were in satisfactory agreement with the corresponding experimental data.

Zendehboudi et al. (Zendehboudi et al., 2014) performed laboratory tests on asphaltene precipitation to observe the influence of mixture composition temperature, pressure, pressure drop, and dilution ratio. They also compared the results with those of an artificial neural network. They found that temperature and pressure drop have significant impacts on the precipitation of asphaltene. Ju et al. (Ju et al., 2013) developed a 3D multiphase indicating the carbon dioxide transport into a reservoir, and precipitation of asphaltene. In the study, they observed the influence of asphaltene precipitation on the petroleum production trend during CO_2 injection. The results obtained in the study indicated that the permeability and production rate reduce with the precipitation of asphaltene.

Huang et al. (Lei et al., 2010) conducted an experimental investigation as well as a modeling approach to study asphaltene precipitation inducted with carbon dioxide flooding. Shahebrahimi and Zonnouri (Shahebrahimi and Zonnouri, 2013) developed a thermodynamics model for the determination of asphaltene precipitation. The model is composed of Flory–Huggins as well as None Random Two Liquid (NRTL) models. The results demonstrated a satisfactory accuracy between the model values and experimental data.

Rassamdana et al. (Rassamdana et al., 1996) proposed the two-variables scaling equation as a function of the solvent to oil dilution ratio (R), and solvent molecular weight (M) for predicting the phase behaviour of asphaltene precipitation. Next, Rassamdana and Sahimi (Rassamdana and Sahimi, 1996) modified the scaling equation suggested by

Rassamdana et al. (Rassamdana et al., 1996) by observing the influence of temperature (T). The asphaltene scaling equation proposed by Rassamdana et al. (Rassamdana et al., 1996) is formulated as follows:

$$X = \frac{R_v}{M^z} \tag{2.9}$$

$$Y = \frac{W}{{R_v}^{z'}} \tag{2.10}$$

where R_v stands for the solvent to oil dilution ratio, M expresses the solvent molecular weight, and z and z' denote the tuning parameters of equations above. Regardless of oil and precipitant applied in the experiments, z and z' should be considered 2 and −2 (Rassamdana et al., 1996). The scaling equations mentioned above can be expressed with a new form through a polynomial function as below:

$$Y = A_1 + A_2X + A_3X^2 + A_4X^3 \quad (X \geq X_c) \tag{2.11}$$

where X_c denotes the value of X at the onset point of asphaltene precipitation, and A_1, A_2, A_3, and A_4 are considered as the scaling coefficients. As mentioned earlier, Rassamdana and Sahimi (Rassamdana and Sahimi, 1996) modified the scaling equation by considering the influence of temperature as follows:

$$x = \frac{X}{T^{C_1}} \tag{2.12}$$

$$y = \frac{Y}{X^{C_2}} \tag{2.13}$$

$$y = b_1 + b_2x + b_3x^2 + b_4x^3 \quad (x \geq x_c) \tag{2.14}$$

In Eqs. (2.8–2.14), X and Y are the same variables which are defined in Eqs. 2.9 and 2.10, C_1 and C_2 denote the adjustable parameters of the asphaltene precipitation scaling equation (a good fit of the experimental data of asphaltene precipitation with the predicted values by Eqs. (2.8–2.14) is obtained at C_1=0.25 and C_2=1.6), and b_1, b_2, b_3, and b_4 are considered as the coefficients of third-order polynomial scaling equation regarding effect of temperature.

2.7. Wax Disappearance Temperature

Several experimental and mathematical studies have been conducted regarding the measurement and modelling of wax deposition (Kamari et al., 2013c; Kamari et al., 2014d; Mohammadi et al., 2011; Mohammadi et al., 2012b; Mohammadi et al., 2012d; Mohammadi et al., 2006). Moradi et al. predicted wax disappearance temperature by means of artificial neural networks (Moradi et al., 2013b). Ji et al. used a set of data which included a number of binary and multi-component systems to develop a new thermodynamic model for prediction of WDT (Ji et al., 2004).

A correlation for wax disappearance temperature was proposed by Moradi et al. for various pressure conditions (Moradi et al., 2013a). Ghanaei et al. developed a novel thermodynamic model to predict formation of wax at high pressure conditions and, a simplified thermodynamic model which works based on sensitivity analysis together with a new improved predictive UNIQUAC to forecast wax formation from paraffinic compounds (Ghanaei et al., 2012; Ghanaei et al., 2014; Ghanaei et al., 2007). In other research by this group, 25 groupings of five models (regular solution theory, UNIFAC, predictive UNIQUAC, predictive Wilson and ideal solution model) for the explanation of solid and liquid phases have been analyzed to get the highest agreement between experimental and predicted data (Ghanaei et al., 2006).

To illustrate an improved representation of wax deposition of hydrocarbon fluid, Zuo et al. established a solid-solution model (Zuo et al., 2001). Daridon et al. designed a high pressure device which uses a polarizing microscope to control visually the liquid–solid phase conversions in intricate waxy systems (Daridon et al., 2002). Banki et al. developed a new mathematical model to predict wax deposition in transferring equipment for laminar flow (Banki et al., 2008). Kelechukwu et al. proposed an empirical model by advanced feed forward neural network to forecast deposition of wax in production systems (Modesty Kelechukwu et al., 2013). Li et al. analyzed the pressure impact on WDT/WAT of crude oil with different water cuts (Li and Jing, 2010).

Coutinho et al. undertook a review to compare different methods of cloud point measurements (Coutinho and Daridon, 2005). Won modeled new thermodynamic correlations for wax appearance temperatures and the compositions of the deposited waxes in a broad range of temperature (Won, 1989). Chen et al. developed a

thermodynamic model for wax deposition and novel correlations for properties of solid-solid and solid-liquid conversions. A thermodynamic model to forecast phase equilibriums of crude oils was also proposed (CHEN and ZHAO, 2006; Chen et al., 2007).

2.8. Hydrocarbon–plus (C_{7+}) Properties of Crude Oils and Gas-Condensates

An adequate and reliable determination and characterization of properties of heptane–plus components of petroleum fractions increases the accuracy of calculations related to the PVT calculations (Whitson, 1983). To this end, Whitson (Whitson, 1983) developed a distribution method for the estimation of molecular weight of heptane–plus components of crude oils as follows (Riazi, 1989):

$$F(M) = \frac{(M-\eta)^{\alpha-1} \exp\left[-\frac{M-\eta}{\beta}\right]}{\beta^{\alpha}\,\Gamma(\alpha)} \tag{2.15}$$

where *F(M)* stands for the possibility density function, *M* denotes the molecular weight of heptane–plus components of crude oil, and α, β, η express the distribution parameters of the model. An improvement on the model presented above has been performed by Whitson et al. (1986) (Whitson et al., 1989) to characterize the molecular weight and the boiling point temperature of crude oil and gas-condensate samples.

Another successful distribution model has been proposed by Riazi (1989) (Riazi, 1989) for characterization and estimation of the boiling point temperature, molecular weight, and specific gravity of heptane–plus (C_{7+}) components of crude oils and gas-condensate samples as follows (Riazi, 1989):

$$P^{*} = [\frac{A}{B}\ \ln(\frac{1}{1-x})]^{\frac{1}{B}} \tag{2.16}$$

where $P^{*}=(P-P_0)/P$ and *P* are considered for the molecular weight , specific gravity , and or boiling point temperature, *x* stands for cumulative weight, mole, and or volume fractions, and finally *A*, *B* and P_0 are the model parameters which can be obtained by linear regression of the data used (Riazi, 1989).

Moreover, there are studies for characterization of heavier components of petroleum fractions such as that reported by Moradi et al. (Moradi et al., 2011). An artificial neural network model was developed for the estimation of the molecular weight, specific gravity, and boiling point temperature. The results obtained by Moradi et al. (Moradi et al., 2011) indicated a good fit between the values predicted by the ANN modelling and the actual data used.

2.9. Vaporization Enthalpy of Petroleum Fractions

Vaporization enthalpy is important from both the experimental and theoretical perspective because of its use in engineering optimization and design, and thus experimental techniques, correlations, and estimation models have been developed to provide greater thermodynamic insight (Fang et al., 2003). For hydrocarbon components, Vetere (Vetere, 1979; Vetere, 1995) developed two empirical correlations for the calculation of vaporization enthalpy using two variables, viz. molecular weight and normal boiling temperature. Riazi and Daubert (Riazi and Daubert, 1980) proposed an empirical correlation for predicting the vaporization enthalpy as a function of T_b and S. Both the Vetere and Riazi and Daubert correlations showed an estimation error of approximately 7%.

Mohammadi and Richon (Mohammadi and Richon, 2007) developed a simple correlation for vaporization enthalpy as a function of the T_b and S. The correlation is capable of calculating the vaporization enthalpies of pure hydrocarbon components and petroleum fractions. They also proposed an artificial neural network (ANN) tool for comparison of the results obtained from their empirical correlation. The results showed agreement between their empirical correlation, the ANN model, and experimental values. Parhizgar et al. (Parhizgar et al., 2013) proposed an empirical method for determination of vaporization enthalpies of pure hydrocarbon components and petroleum fractions using genetic programming which is a function of the T_b and S. Their results indicated that their correlation can calculate the vaporization enthalpy of both pure hydrocarbon components and petroleum fractions with an average absolute relative deviation (AARD) of approximately 1.35%.

2.10. Gasoline Properties

There have been many attempts to determine the properties of gasoline, especially with regard to characterization of its quality. Mendes et al. (Mendes et al., 2012) utilized distillation curves (ASTM D86) to estimate the RON and MON efficiently. Balabin et al. (Balabin et al., 2007) compared the most popular methods for the determination of properties, and consequently characteristics, of gasoline, with the use of near infrared spectroscopy data. Murty and Rao (Murty and Rao, 2004) developed a neural network-based model for estimating the RON. They compared the results obtained with a multiple linear regression technique accessible in the open literature. In their opinion, the advantage of the neural network-based model over the classical methods is that the neural network method can handle data covering several variables which do not have to be specified in advance.

Aleme et al. (Aleme et al., 2009) developed a partial least-squares regression model for predicting the value of SG using distillation curves (ASTM-D86). They state that their method is effective in predicting the SG and content of ethanol in gasoline.

The standard techniques used to evaluate the quality of fuels, petroleum, and its fractions are tedious and time-consuming. They also require highly qualified personnel and costly equipment (de Oliveira et al., 2004). There is therefore a need for prediction methods which can determine gasoline properties from easy-to-determine gasoline parameters or properties, with an aim to understand which parameters most affect the properties. This study aims to develop a reliable predictive model for accurately predicting some gasoline properties, viz. SG, MON, RON and RVP.

2.11. Gas Compressibility Factor of Natural Gas

Generally, the volumetric properties of petroleum fluid are predicted from laboratory tests, empirically derived models, or thermodynamic models (Yan et al., 2013). Normally, high-temperature and high-pressure apparatuses are utilized for the experimental measurements (Chylinski et al., 2002). These measurements are expensive and time-consuming and it is impossible to measure properties for all possible compositions of natural gasses (Ahmed, 2006).

In addition to laboratory tests, equations of state and empirically derived models can predict the properties related to petroleum fluid. To determine the natural gas *z*-factor, empirical correlations are more rapid and simple than equations of state (involving a large number of parameters) which require longer computations and are complicated (Elsharkawy, 2004). Furthermore, the gas compressibility factor of all EoS models is implicit and consequently should be explained as a root of the EoS (Heidaryan et al., 2010a).

It is worth noting that in-spite of the abovementioned drawback, EoS has some advantages. For instance, it expends some accuracy in exchange for smoothing the model and for nice mathematical behaviour in terms of the derivatives of the mathematical function. The partial derivatives of compressibility lead to different expressions, including the entropy, enthalpy, and Gibbs free energy residuals, which are in turn utilized to estimate fugacity coefficients and then are used to describe phase equilibria (Kamari et al., 2013a).

As a definition, the gas compressibility factor of an ideal gas is the ratio of the molar volume occupied by a gas to the molar volume related to an ideal gas at a given/same temperature and pressure (Kumar, 2005). In other words, the compressibility factor of gasses is a dimensionless quality which is a function of pressure and temperature. According to the kinetic concept related to gasses (Ahmed, 2006), the volume of a molecule is insignificant and/or unimportant compared to the total bulk volume. Also, it is assumed that there are neither attractive nor repulsive forces among the gas molecules (Ahmed, 2006).

A mathematical equation called an equation of state shows a relationship among pressure (P), volume (V), and temperature (T) for a given quantity of moles of gas (n). Consequently, the abovementioned relationship is mathematically expressed by the following equation:

$$PV = n'RT \qquad (2.17)$$

where *P* represent the pressure, *V* denotes the volume, n' is number of moles, *R* stands for gas constant, and *T* indicates the temperature.

Gases that deviate from the ideal are known as real gasses. Ideal gas EoS shows low deviations from experimental data at atmospheric pressure (2-3% of average

absolute relative deviation) whereas, its application at high pressures is not recommended due to the high deviations which it causes (Ahmed, 2006). The deviation dramatically increases with an increase in temperature and pressure and the deviation depends on the gas composition.

To correlate the pressure, volume and temperature (PVT) parameters, various EoSs have already been reported for real gasses. In order to present the relationship among the parameters above (PVT variables), an equation can be formulated as follows:

$$PV = Zn'RT \tag{2.18}$$

where *P* represent the pressure, *V* denotes the volume, *Z* shows the gas compressibility factor *n*, is number of moles, *R* stands for gas constant, and *T* indicates the temperature.

Investigation of the compressibility factor for natural gasses of different compositions have shown that the *z*-factor can be employed with adequate accuracy for most engineering targets when they are defined in relation to the following two dimensionless properties (Ahmed, 2006), which are expressed as follows:

$$P_{pr} = \frac{P}{P_{pc}} \tag{2.19}$$

$$T_{pr} = \frac{T}{T_{pc}} \tag{2.20}$$

where P_{pr} denotes the pseudo-reduced pressure, T_{pr} expresses the pseudo-reduced temperature, T_{pc} stands for the pseudo-critical temperature and P_{pc} is the pseudo-critical pressure, which are expressed by following equations:

$$P_{pc} = \sum_{i=1}^{n} y_i P_{ci} \tag{2.21}$$

$$T_{pc} = \sum_{i=1}^{n} y_i T_{ci} \tag{2.22}$$

where P_{ci} denotes the critical pressure, T_{ci} stands for the critical temperature, and y_i expresses the mole fraction of component *i*.

For the determination of *z*-factor, Standing and Katz (Standing and Katz, 1942) expressed a generalized *z*-factor chart on the basis of the theory of pseudo-reduced properties such as pseudo-reduced pressure (P_{pr}) and pseudo-reduced temperature (T_{pr}). Equations of state are another method for estimating the gas compressibility factor. The van der Waals EoS model was one of the first to consider the molecule's volume and intermolecular forces (van der Waals, 2004 and Valderrama, 2003). It should be noted that determination of *z*-factor utilizing the van der Waals EoS is a development over utilizing the *z*-factor charts generalized by Standing and Katz. Nevertheless, there are several modified forms of the van der Waals EoS, which provide even more accuracy and development for prediction of *z*-factor (Al-Anazi et al., 2011). Some of the commonly used equations of state for calculation of *z*-factor are presented as follows: the Redlich and Kwong equation (Redlich and Kwong, 1949), the Soave-Redlich-Kwong equation (Soave, 1972), the Peng and Robinson equation (Peng and Robinson, 1976) and Lawal-Lake-Silberberg equation (Lawal, 1999).

In addition to equations of state, numerous empirically derived models for direct calculation of the gas compressibility factor have been reported over the years (2003; Beggs and Brill, 1973; Dranchuk and Kassem, 1975; Dranchuk et al., 1973; Gopal, 1977; Hall and Yarborough, 1973). For example, Hall and Yarborough in 1973 (Hall and Yarborough, 1973) developed a correlation that precisely denotes the gas compressibility factor chart presented by Standing and Katz. The developed equation is based on the Starling-Carnahan (Carnahan and Starling, 1969) EoS (Ahmed, 2006). The coefficients of the correlation were obtained by fitting them to data points extracted from the gas compressibility factor chart presented by Standing and Katz. Based on the Benedict-Webb-Rubin (Benedict et al., 1942) type of EOS, Dranchuk, Purvis, and Robinson in 1974 (Dranchuk et al., 1973) developed a correlation. They optimized the eight coefficients of the developed equations by fitting the equation to 1,500 data points extracted from the gas compressibility factor chart presented by Standing and Katz.

There are many empirically derived correlation that can be used for the determination of the plus properties of natural gas components. A review is given by the late Ali Danesh (Danesh, 1998). He recommended, as one of most reliable methods for this purpose, the correlation proposed by Twu (Twu, 1984) as follows:
Critical temperature.

$$T_c = T_c{}^o[(1 + 2f_T)/(1 - 2f_T)]^2 \quad (2.23)$$

where

$$f_T = \Delta SG_T\left[-0.362456/T_b^{\frac{1}{2}} + (0.0398285 - 0.948125/T_b^{\frac{1}{2}})\,\Delta SG_T\right] \quad (2.24)$$

where

$$\Delta SG_T = exp[5(SG^o - SG)] - 1 \quad (2.25)$$

Critical volume.

$$V_c = V_c{}^o[(1 + 2f_V)/(1 - 2f_V)]^2 \quad (2.26)$$

where

$$f_V = \Delta SG_V\,[0.466590/T_b^{\frac{1}{2}} + (-0.182421 + 3.01721/T_b^{\frac{1}{2}})\Delta SG_V] \quad (2.27)$$

where

$$\Delta SG_V = exp[4(SG^{o2} - SG^2)] - 1 \quad (2.28)$$

Critical pressure.

$$P_c = P_c{}^o(T_c/T_c{}^o)(V_c{}^o/V_c)\left[\frac{1+2f_P}{1-2f_P}\right]^2 \quad (2.29)$$

where

$$f_P = \Delta SG_P\left[(2.53262 - 46.1955/T_b^{\frac{1}{2}} - 0.00127885\,T_b) + (-11.4277 + 252.140/T_b^{\frac{1}{2}} + 0.00230535\,T_b)\Delta SG_P\right] \quad (2.30)$$

where

$$\Delta SG_P = exp[0.5(SG^o - SG)] - 1 \quad (2.31)$$

where in the equations above, the superscript "*o*" denotes correlations specific to the *n*-alkanes, T_b is the normal boiling point temperature (°R), *SG* stands for the specific gravity, and finally T_c, V_c and P_c express critical temperature, volume, and pressure, respectively.

CHAPTER 3

3. Data Management

Data mining is a broad term often utilized to describe the process of using database technology, modeling techniques, statistical analysis, and machine learning to analyze large amounts of data in an automated fashion to discover hidden patterns and predictive information in the data. By building highly complex and sophisticated statistical and mathematical models, organizations can gain new insight into their activities. More and more, organizations are using data mining to make proactive knowledge-driven decisions, and improving their organization's efficiency and effectiveness.

Despite the power of data mining tools, growth and user adoption of these applications has been low. Therefore, it can be said that the applicability, reliability and accuracy of any predictive model is normally associated with the validity of the employed data. In other words, the success of models developed on the basis of the artificial intelligence approach strongly depends on the comprehensiveness of the data employed, so that the database provided for model development should cover a comprehensive ranges of input and output variables.

3.1. Ranges of Data for Surfactant Retention

Normally, chemical EOR methods such as surfactant-based flooding techniques are applied to oil with an API gravity higher than 15 and viscosity in the range of 15-35cp and high-intermediate depths (Taber et al., 1997). As already mentioned, the retention of surfactant plays a key role in surfactant-based EOR methods. As a matter of fact, the test temperature, maximum effluent pH, reservoir rock type (carbonated or sandstone), co-solvent concentration, molecular weight of surfactant mixture, total acid number of the oil (TAN), absolute permeability, mobility ratio, salinity of polymer, and surfactant formulation all have considerable influence on the surfactant retention during surfactant-based flooding (Solairaj, 2011). In the implementation of the surfactant flooding method, the abovementioned parameters can affect retention and/or adsorption of surfactant in a porous media, as listed below (Yassin et al., 2013):

- The retention of surfactant dependents on several parameters including the acidity of the oil or TAN, chemical slurry formulation, reservoir temperature, types of electrolytes present in the solution, and also the type of reservoir rock.
- Alkalinity decreases adsorption of anionic surfactant on sand.
- By increasing the pH, the charge on the sand surface negatively increases and the rate of anionic surfactant adsorption decreases.
- With regard to aqueous phase stability and microemulsion phase behavior, an increase in temperature would affect surfactant retention for a given surfactant solution at certain conditions.
- An increase in the molecular weight of the surfactant results in an increase in the adsorption of surfactant.
- Adsorption of surfactant into a porous media is affected by the type and characteristics of the porous rock present.
- An increase in TAN results in a decrease in the retention of surfactant because by increasing TAN at high pH, the in-situ soap generated decreases the active sites present for synthetic surfactant adsorption, and consequently, reduces the retention of surfactant in porous rock (Solairaj, 2011; Solairaj et al., 2012; Zhang et al., 2006).
- An increasing the mobility ratio can cause an increase in retention of surfactant in porous media. This behavior is in agreement with an existing empirical correlation

presented by Solairaj et al. (Solairaj, 2011; Solairaj et al., 2012) for determination of surfactant retention during flooding into porous rocks.

In view of the issues mentioned above, it is necessary to select and collect all of the important parameters influencing the retention and/or adsorption of surfactant in porous media during surfactant flooding. Therefore, a reliable database containing 47 core flooding laboratory experiments (Solairaj, 2011; Solairaj et al., 2012) performed in the chemical EOR group of CPGE at the University of Texas at Austin is selected and used in this study. The available parameters in this database are the test temperature, maximum effluent pH, reservoir rock type, co-solvent concentration, average molecular weight of surfactant mixture, TAN, absolute permeability, mobility ratio, salinity of polymer, surfactant formulation, and the corresponding experimental/literature surfactant retention values. It should be mentioned that the surfactants employed in the laboratory experiments are composed of mixtures of several anionic surfactants including alkyl benzene sulfonates (ABS), internal olefin sulfonates (IOS), large hydrophobe Guerbetalkoxy carboxylates (GAC), alcohol propoxy sulfates (APS), and large hydrophobe Guerbetalkoxy sulfates (GAS). **Table 3.1** summarizes the maximum, minimum, average and standard deviation values for all parameters existing in the database for both sandstone and carbonated rocks.

Table 3. 1 Ranges and averages of the input/output data used for developing the new model for the prediction of surfactant retention in porous media.

Parameter	Unit	Min.	Avg.	Max.	SD	Type
K_{abs}	mD	115	1222.04	6400	1548.000	Input
TAN	mg KOH/g-oil	0.0	1.13	3.2	0.948	Input
Temperature	0_C	25	54.83	100	22.622	Input
Co-solvent	wt %	0	0.01	0.030	0.006	Input
Polymer Salinity	ppm	250	12779.79	41000	10202.000	Input
Max Effluent pH		6.0	9.34	11.0	1.768	Input
Mobility Ratio		0.01	0.31	1.40	0.282	Input
MW Surfactant		368.52	687.76	1320.00	263.000	Input
Retention	mg/g-Rock	0.040	0.18	0.370	0.101	Output

3.2. Variables for Estimation of Dew Point Pressure of Gas Condensate Reservoirs

The dew point pressure (DPP) in gas condensate reservoirs is strongly influenced by the properties of C_{7+} (molecular and specific gravity), reservoir temperature, and compositions of hydrocarbon and non-hydrocarbon components (A. González, 2003; Akbari and Jalali, 2007; Elsharkawy, 2002b; L. Nemeth, 1967; Nowroozi et al., 2009; Shokir, 2008). An extensive dataset is thus required if one wants to develop a good correlative model.

A comprehensive databank consisting of 562 experimental data (Nemeth, 1966) for DPP obtained from CVD tests was collected. The database comprises the values for DPP (P_d, Psia), molecular weight for heptane plus fractions (MW_{C7+}), reservoir temperature (T_R, °F), specific gravity for heptane plus fractions (SG_{C7+}), compositions of hydrocarbons including methane (C_1), ethane (C_2), propane (C_3), butanes (C_4), pentanes (C_5), hexanes (C_6), heptane-plus (C_{7+}), and compositions of non-hydrocarbons including nitrogen (N_2), carbon dioxide (CO_2), and hydrogen sulfide (H_2S). **Table 3.2** summarizes the ranges, averages, and units of the abovementioned parameters. As can be seen in **Table 3.2**, the parameters presented cover a wide range of DPP, reservoir temperature, *etc*.

Table 3. 2 Ranges, averages and units of the variables implemented for the development of the GEP-based model for the prediction of dew point pressures.

Property	Unit	Min.	Max.	Avg.
Dew-point pressure, DPP	Psia	1405	10790	4747.222
Reservoir temperature, T_R	°F	40	320	205.146
Molecular weight, MW_{C7+}	-	106	235	148.189
specific gravity, SG_{C7+}	-	0.7330	0.8681	0.788
Nitrogen, N_2	mole fraction	0.0000	0.4322	0.010
Carbon dioxide, CO_2	mole fraction	0.0000	0.9192	0.015
Hydrogen sulfide, H_2S	mole fraction	0.0000	0.2986	0.006
Methane, C_1	mole fraction	0.0349	0.9668	0.802
Ethane, C_2	mole fraction	0.0037	0.1513	0.057
Propane, C_3	mole fraction	0.0011	0.1090	0.030
Butanes, C_4	mole fraction	0.0017	0.2030	0.020
Pentanes, C_5	mole fraction	0.0006	0.0631	0.012
Hexanes, C_6	mole fraction	0.0004	0.0510	0.009
Heptane-plus, C_{7+}	mole fraction	0.0019	0.1356	0.037

3.3. Range of Data for PVT Properties of Reservoir Crude Oil

A comprehensive review on the previously published empirical methods developed for the prediction of oil formation volume factor and bubble point pressure since the early 1940s indicates the importance of PVT properties from the industry point of view (Arabloo et al., 2014). As a definition, OFVF is the reservoir oil required to produce one barrel (1 bbl) of oil at surface conditions (McCain, 1990). Additionally, in its original condition reservoir oil contains some natural gas in solution, consequently, the pressure at which this natural gas begins to come out of solution and form bubbles is identified as the P_b. As a result, P_b and OFVF are among the most vital properties for accurate calculation of hydrocarbon reservoir recoverable reserves, the oil-water flow ratio, reservoir capacity for production of oil, problems related to enhanced and improved oil recovery, and approximately all other issues associated with petroleum engineering computations (Bandyopadhyay and Sharma, 2011; Obanijesu and Araromi, 2008; Ostermann and Owolabi, 1983; Petrosky Jr and Farshad, 1993; Vazquez and Beggs, 1980). Therefore, presenting accurate and efficient methods for the determination of P_b and OFVF is extremely important.

As a result, in order to predict the PVT properties associated with petroleum reservoir fluid, empirically derived correlations of field measured data are utilized, such as reservoir temperature (T_R), reservoir pressure (P_R), crude oil API gravity (API) and gas specific gravity or gas relative density (γ_g) and solution gas oil ratio (Gharbi and Elsharkawy, 1997b). Reservoir temperature, crude oil API gravity, gas relative density and solution gas oil ratio, which are required variables for accurate estimation of P_b and OFVF, are found in the majority of published work (Al-Shammasi, 1999; Kartoatmodjo and Schmidt, 1994a; Macary and El-Batanoney, 1993; Petrosky Jr and Farshad, 1998; Standing, 1947a; Vazquez and Beggs, 1980) as follows:

$$P_b = f_1(T_R, \gamma_g, GOR, API) \quad (3.1)$$

$$P_b = f_2(T_R, \gamma_g, GOR, API) \quad (3.2)$$

Geographical and geological conditions associated with reservoir oil are very important in PVT empirical correlations because the chemical composition must be specified for any crude oil. In other words, obtaining consistent, accurate results by means of PVT empirical correlations for different crude oils having different chemical and

physical characteristics is difficult to accomplish (Mahmood and Al-Marhoun, 1996a). Hence, to account for regional characteristics, PVT empirical correlations need to utilize comprehensive datasets which cover a wide range of PVT properties from almost all regions of world. We can go so far as to say that the quality and reliability of predictive models for measuring the PVT and thermo-physical properties is contingent on the applicability of the applied database (Kamari et al., 2014b; Kamari et al., 2014c; Kamari et al., 2015d). Therefore, about 755 laboratory PVT datasets covering varied ranges of PVT experimental conditions from various geographical and geographic world's regions were utilized in this study to develop and test the models for accurately determination of P_b and OFVF.

The dataset used for developing the P_b and OFVF models comprise reservoir temperature (according to °F), oil formation volume factor at bubble point pressure (according to bbl/STB), crude oil API gravity and gas gravity and solution gas oil ratio at bubble point pressure (according to SCF/STB), which was collected from Moghaddam et al. (Moghadam et al., 2011), Obomanu and Okpobiri (Obomanu and Okpobiri, 1987), Bello et al. (Bello et al., 2008), Omar and Todd (Omar and Todd, 1993), Dokla and Osman (Dokla and Osman, 1992), Mahmood and Al-Marhoun (Mahmood and Al-Marhoun, 1996a), Ghetto et al. (De Ghetto and Villa, 1994), Al-Marhoun (Al-Marhoun, 1988), and Ostermann et al. (Ostermann and Owolabi, 1983). **Table 3.3** summarizes the values of minimum, maximum and average for reservoir temperature, gas oil ratio, bubble point pressure, gas gravity and oil formation volume factor. The table confirms that the databank collected in this study covers a wide range of PVT properties from volatile oils to heavy crudes.

Table 3. 3 The minimum, maximum and average values associated with the PVT properties in the databank utilized for the GEP models.

PVT Properties	Unit	Min.	Max.	Avg.	Type
Oil formation volume factor, B_{ob}	bbl/STB	1.02	2.92	1.40	Output
Bubble point pressure, P_b	psi	58.02	6613.82	1846.05	Output
Gas gravity, γ_g	-	0.52	3.44	1.12	Input
Initial solution gas oil ratio, R_{Si}	SCF/STB	7.08	3298.66	592.39	Input
Reservoir temperature, T_R	°F	74.00	360.93	207.17	Input
Oil gravity, API	-	6.00	56.80	34.36	Input

3.4. The Field Data for Reservoir Oil Viscosities

Laboratory tests conducted on the bottom hole cores or surface recombined samples of reservoirs normally arrive at the viscosities of reservoir oil isothermally at reservoir temperature and at different reservoir pressures (Obanijesu and Omidiora, 2009). In fact, three pressure regions, including above and below bubble point pressure as well as for dead oil (gas-free reservoir oil) are utilized (**Fig. 3.1**). Thus, a specific correlation should be developed for each pressure region in order to take into account differences in the nature of crudes and compositions (Hemmati-Sarapardeh et al., 2014b).

It has therefore been confirmed that the crude oil viscosities depend on the reservoir pressure and temperature. The viscosity at above bubble point pressure increases and below bubble point pressure reduces (Torabi et al., 2011).

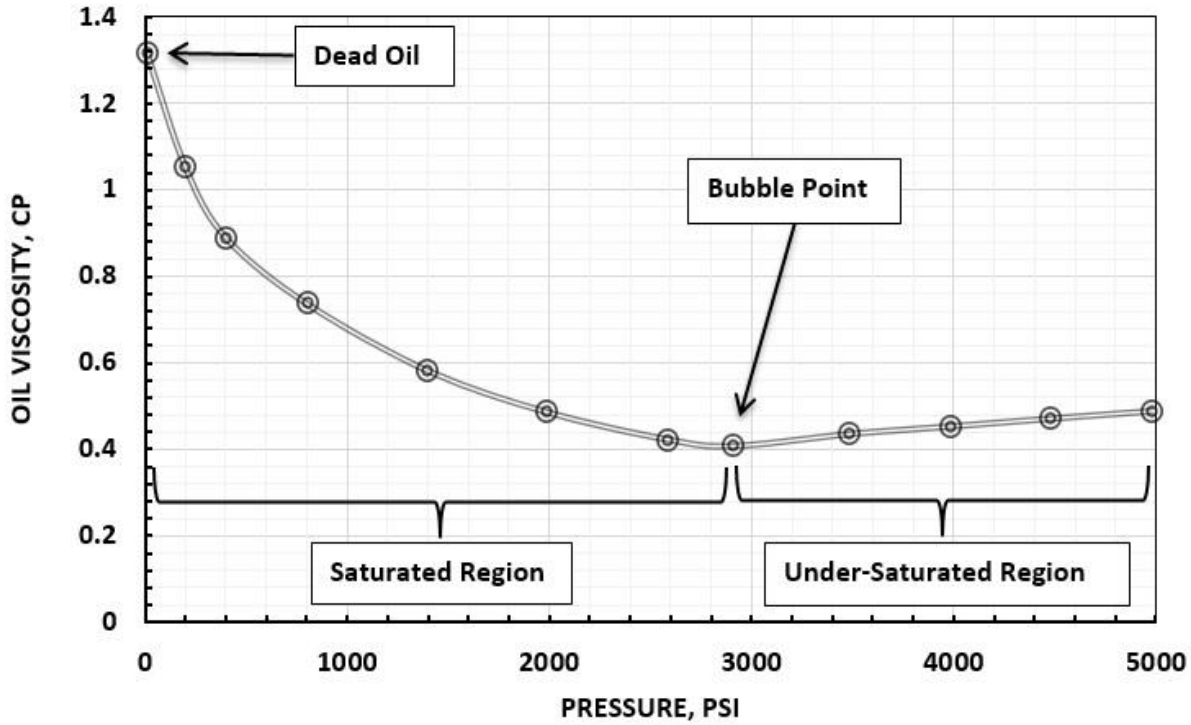

Fig. 3. 1 A representative trend plot of viscosity versus pressure illustrating regions related to dead, saturated, and under-saturated oil viscosities.

Various empirical correlations have been reported for dead, saturated and under-saturated crude oils over the years. In this study, several empirical correlations related to dead oils (Al-Khafaji et al., 1987; Alomair et al., 2011; Beal, 1946; Beggs and Robinson, 1975; Bennison, 1998; Egbogah and Ng, 1990; Elsharkawy and Alikhan, 1999; Glaso, 1980; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Kaye, 1985; Labedi, 1992; Naseri et al., 2005; Petrosky, 1990), saturated oils (Al-Khafaji et al., 1987; Beggs and Robinson, 1975; Bergman and Sutton, 2007; Chew and Connally Jr, 1959; Elsharkawy and Alikhan, 1999; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Khan et al., 1987; Labedi, 1992; Naseri et al., 2005; Petrosky, 1990) and under-saturated oils (Beal, 1946; Elsharkawy and Alikhan, 1999; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Khan et al., 1987; Labedi, 1992; Orbey and Sandler, 1993; Petrosky, 1990; Vazquez and Beggs, 1980) are employed as points of comparison with the models developed in this study.

Tables 2. 2-4 list the comparative correlations employed in this study as well as their ranges, data origin and formulas. To develop these types of correlations, the common reservoir fluid data including the oil gravity, solution gas oil ratio, bubble point pressure and reservoir temperature are applied. Consequently, the comparative correlations used confirm that reservoir temperature and crude oil API gravity are applied for the prediction of dead oil viscosity. Additionally, the dead oil viscosity and saturated pressure are the required parameters for the determination of saturated reservoir oil viscosity. The pressure in general, as well as pressure and viscosity at bubble point are the variables used for the estimation of under-saturated reservoir oil viscosity. The reservoir parameters used to develop the GEP-based models for the determination of dead, saturated and under-saturated crude oil viscosities are as follows:

$$\mu_{od} = f_1(T_R,\ API) \tag{3.3}$$

$$\mu_{ob} = f_2(\mu_{od},\ P) \tag{3.4}$$

$$\mu_o = f_3(\mu_{ob},\ P,\ P_b) \tag{3.5}$$

In the above equations, more than 1000 series were used of experimental PVT data drawn from Iranian oil fields, comprising the gravity of oil, reservoir temperature, solution gas oil ratio, and bubble point pressure. A rolling ball viscometer (Ruska, series 1602) has been implemented for accurately measuring the Iranian reservoir oil viscosities at different pressures above and below saturation pressure. **Tables 3.4-6** list the ranges of the applied variables used to develop the GEP-based models for the prediction of dead, saturated and under-saturated oil viscosities, respectively. It is worthwhile to note that the range of the data presented in **Tables 3.4-6** includes almost all of the PVT data available for Iranian oil fields.

Table 3. 4 Ranges and units of the applied variables for developing the dead oil viscosity model

Reservoir Property	Unit	Min.	Max.	Avg.	SD	Type
Temperature	°F	50.27	290.26	176.11	42.84	Input
Oil API gravity	-	17.30	43.56	29.32	7.00	Input
Dead oil viscosity	cP	0.55	69.50	7.41	11.44	Output

Table 3. 5. Ranges and units of the applied variables for developing the saturated oil viscosity model.

Reservoir Property	Unit	Min.	Max.	Avg.	SD	Type
Saturation Pressure	psi	158.09	5701.43	1705.64	7.50	Input
Dead oil viscosity	cP	0.55	37.18	4.55	5.05	Input
Saturated oil viscosity	cP	0.18	25.58	1.92	2.59	Output

Table 3. 6 Ranges and units of the applied variables for developing the under-saturated oil viscosity model.

Reservoir Property	Unit	Min.	Max.	Avg.	SD	Type
Bubble point Pressure	psi	729.53	5115.47	1135.39	7.81	Input
Bubble point viscosity	cP	0.18	18.16	1.62	2.19	Input
Pressure	MPa	5.03	86.18	25.38	11.63	Input
Under-saturated oil viscosity	cP	0.18	31.00	1.84	2.97	Output

3.5. Comprehensive Databank for Solution–Gas Oil Ratio

As mentioned above, the solution GOR plays a key role in PVT analysis related to petroleum engineering calculations. Solution GOR affects the OFVF, the viscosity compressibility of oil, and the determination of the in-situ total reservoir fluid rates. As a definition, solution GOR is the amount of gas dissolved in oil with regard to pressure. Here, it should be noted that reservoirs containing light oils have more dissolved gas than reservoirs with heavy oils. With an increase in pressure, solution GOR increases approximately linearly until the attainment of bubble point/saturation pressure (P_b); after which it is a constant and the oil is supposed to be under-saturated (**Fig. 3.2**). **Fig. 3.2** is a typical illustration of the trend of solution GOR versus pressure.

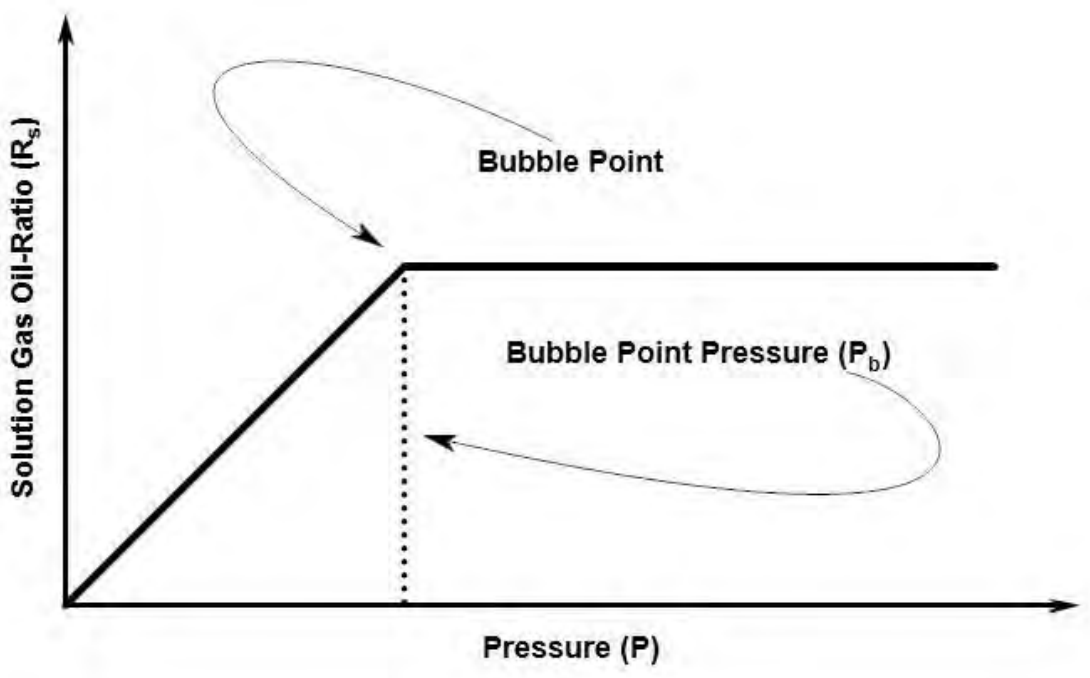

Fig. 3. 2 Typical trend of solution GOR versus pressure.

Most of the empirically derived methods reported in the open literature for the determination of reservoir fluid properties have been developed on the basis of data related to a specific region and limited PVT studies (Baniasadi et al., 2015). This is a drawback that can decrease their accuracy in predicting reservoir fluid properties at a

particular solution gas–oil ratio. This means that the empirical correlations used may lead to significant deviation when they are utilized for the estimation of reservoir fluid properties for other geographical locations. For this reason, it is important to collect a comprehensive databank covering a wide range of reservoir fluid properties for all regions in the world.

Therefore, a reliable and comprehensive databank (Abdul-Majeed et al., 1988; Al-Marhoun, 1988; Bello et al., 2008; Dokla and Osman, 1992; Ghetto et al., 1994; Mahmood and Al-Marhoun, 1996b; Moghadam et al., 2011; Obomanu and Okpobiri, 1987; Omar and Todd, 1993; Ostermann et al., 1983) comprising more than 1000 data series collected from various geographical locations including Asia, the Mediterranean Basin, North America, Africa, and the Middle East was compiled. The databank collected includes reservoir fluid properties, viz. solution GOR (R_s, SCF/BBL), bubble point pressure (P_b, psi), reservoir temperature (T_R, °F), and gas gravity, as well as oil gravity. A statistical description of the properties, including maximum, minimum, and average values is summarized in **Table 3. 7**.

Table 3. 7 Statistical analysis of reservoir fluid properties used for the estimation of solution gas–oil ratio.

Property	Unit	Minimum	Average	Maximum	Role
R_s	SCF/STB	7.08	515.32	3298.66	Output
γ_g	–	0.52	1.00	3.44	Input
T_R	°F	54.9	173.30	360.93	Input
API	–	6.00	33.41	56.8	Input
P_b	psi	58.01	1755.58	7127.01	Input

3.6. Iranian Crude Oil Data for Asphaltene Precipitation

According to Rassamdana et al. (Rassamdana et al., 1996) and Rassamdana and Sahimi (Rassamdana and Sahimi, 1996), the solvent to oil dilution ratio, temperature,

and solvent molecular weight, are the most important parameters for use in asphaltene scaling equations for prediction of asphaltene precipitation (W_t). To forecast the phase behaviour of asphaltene precipitation, an asphaltenic crude (Ashoori et al., 2010) with oil density of 0.934 g/cc taken from one of the southwestern reservoirs in Iran is used to pursue our modeling target in this study.

The SARA analysis of asphaltenic crude oil is as follows: saturates=29.3 wt. %, aromatics=35.2 wt. %, resins=27.2 wt. %, and asphaltenes=8.3 wt. %. The temperatures tested for measuring the amount of asphaltene precipitated are 30, 50, and 70 °C, which have been undertaken at atmospheric pressure. Furthermore, at different dilution ratios, three asphaltene precipitants including *n*-pentane, *n*-hexane, and *n*-heptane are implemented. **Table 3. 8** summarizes the ranges of parameters applied for estimating the amount of asphaltene precipitation in Iranian crude oil studied (Ashoori et al., 2010).

Table 3. 8 Ranges of effective variables in predicting asphaltene precipitation.

Property	Unit	Minimum	Average	Maximum	Role
Asphaltene precipitation	wt %	0.5	4.78	10.4	Output
MW	–	72.15	68.18	100.21	Input
T	°C	30	50	70	Input
R_v	mL/g	0.67	7.61	20	Input

3.7. Database for Wax Disappearance Temperature

As mentioned earlier, in order to develop reliable models, a reliable database, which covers a wide range of phase behavior related parameters of deposition, needs to be collected. Hence, the experimental data related to wax disappearance temperature (T, K) as a function of molar mass (M, g/mol) and pressure (P, MPa) has been collected from different sources in the literature (Daridon et al., 2002; Ji et al., 2004; Metivaud et al., 1999; Milhet et al., 2005; Robles et al., 1996; Vafaie-Sefti et al., 2000). The experimental wax disappearance temperature data are measured at 0.1-100.3 MPa for the mixtures including C_6–C_{16}, C_6–C_{17}, C_{16}–C_{18}, C_{16}–C_{20}, C_{15}–C_{19}, C_6–C_{16}–C_{17}, C_7–C_{36}, C_{14}–C_{15}, C_{14}–C_{16}, C_{14}–C_{15}–C_{16}, C_{16}–C_{17}–C_{18}, C_{17}–C_{19}, C_{18}–C_{19}–C_{20}, C_{19}–C_{20}–C_{21}, and C_{13}–C_{24}. Ranges (*i.e.* maximum,

minimum and average) of WDT, pressure and molar mass parameters are summarized in **Table 3. 9**.

Table 3. 9 Ranges of inputs/output parameters applied for predicting wax disappearance temperature.

Parameter	Min.	Avg.	Max.	Type
Pressure, MPa	0.1	39.0	100.3	Input
Molar Mass, g/mol	191.0	220.7	293.9	Input
Wax Disappearance Temperature, K	270.6	293.8	321.1	Output

3.8. Distributed Data for Hydrocarbon–plus (C_{7+}) Properties

Specific gravity, boiling point temperature, and molecular weight, which are the basic distributed properties of oil and gas, are normally applied for the determination and characterization of the heptane–plus components of crude oil, and gas-condensate samples (Riazi, 1997). To this end, 62 sets or samples (Whitson et al., 1989) including around 801 data points of heptane-plus properties are used in this study. The databank contains the distributed properties of the specific gravity, boiling point temperature, and molecular weight as well as the weight fractions, bulk molecular weight (MW_b), and bulk specific gravity (SG_b).

The cumulative weight fraction (CX_w) was calculated from weight fractions of each sample and added to the database for estimating the distributed properties of heptane–plus components crude oil, and gas-condensate samples. Ranges (*i.e.* maximum, minimum and average) of T_b, MW, SG, as well as MW_b, SG_b, and CX_w are listed in **Table 3. 10**. The bulk molecular weight, bulk specific gravity, and cumulative weight fraction are considered as the input variables for the determination of distributed properties of heptane-plus components of crude oil, and gas-condensate samples.

Table 3. 10 Ranges of data used for the prediction of C_{7+} properties using the models developed in this study.

Parameter	Min.	Avg.	Max.	Type
Cumulative weight fraction	0.0139	0.44	1	Input
Average molecular weight (g/mole)	118.9	201.91	348.2	Input
Average specific gravity	0.7597	3.91	213.6	Input
Boiling temperature (°R)	631	924.42	1829.8	Output
Specific gravity	0.704	0.82	0.989	Output
Molecular weight (g/mole)	15	196.00	675	Output

3.9. Database for Vaporization Enthalpy

Vaporization enthalpy data are normally determined at the normal T_b through an appropriate technique and are then calculated at the required temperature. There are two classes of empirically derived methods for determining the vaporization enthalpy. The first class of correlations relate vaporization enthalpy, at the normal boiling point temperature, to the critical properties and their normal boiling point (Reid et al., 1987). The next class of correlations relate vaporization enthalpy to the specific gravity, molecular weight, and the normal boiling point temperature (Mohammadi and Richon, 2007). As a result, the selection of the most appropriate input/predictor variables used for building the LSSVM models plays a significant role. Therefore, to accurately predict the vaporization enthalpy (ΔH^{vap}, *kJ/g-mol*) of pure hydrocarbon and petroleum fractions, the parameters which most influence the property, viz. the boiling point temperature (T_b, *K*), specific gravity, and molecular weight (M, *g/g-mol*) are gathered from literature (Fang et al., 2003).

The collected data covers an extensive range of the vaporization enthalpy, from 19.0 to 80.1 *kJ/g-mol* for both petroleum fractions and pure hydrocarbon components. Distribution of the collected data in terms of minimum, and maximum, as well as averages are summarized in **Table 3. 11**, with the input variables being T_b, S, and M, and the output parameter, ΔH^{vap}.

Table 3. 11 Distribution of the data used in this study for forecasting the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

Parameter	Min.	Avg.	Max.	Type
T_b, K	231.1	451.9	722.8	Input
S	0.5	0.7	0.8	Input
M, g/g-mol	44.1	160.2	422.8	Input
ΔH^{vap}, kJ/g-mol	19.0	42.1	80.1	Output

3.10. Ranges of Gasoline Properties Data

Distillation curves obtained using standard American Society for Testing and Materials (ASTM) distillation tests, in terms of both percent volume and boiling points can provide comprehensive information about the intrinsic components in gasoline and petroleum fractions (Albahri, 2014). These parameters have significant influence on the SG, MON, RON and RVP properties of gasoline and petroleum fractions and also provide a reliable correlation with the experimental/literature-reported data.

In this study, ASTM D86 distillation temperatures at various volume percentages (5% to 95%, as well as initial boiling point (IBP) and final boiling point (FBP), commonly extracted from a standard ASTM D86 distillation apparatus, are used to predict the properties of gasoline, viz. SG, MON, RON and RVP. Depending on the gasoline property, the dataset may include additional intrinsic chemical structures like saturates, olefins and aromatic components, in particular for MON and RON properties (Albahri, 2014). The numbers of data points for SG, MON, RON and RVP are 178, 178, 178 and 362.

In order to develop a reliable predictive model for gasoline properties, experimental data were gathered and collated from literature (Albahri, 2014; Healy et al., 1959). The databases generated have values of SG ranging from 0.6849 to 0.9248, MON values ranging from 34 to 107, RON values ranging from 37 to 97, and RVP values ranging from 0.007 to 4.55 bars. These four distinct databases were used for the different properties. Each of the four databases were randomly separated into two sub-data sets

comprising the “Training” and “Test” sets. In the development of a reliable and accurate LSSVM model for predicting the gasoline properties, viz. SG, MON, RON, and RVP, eighty percent of the main databank was randomly selected for the “Training” set and the remaining twenty percent assigned for the testing part.

3.11. The Literature Data for the Gas Compressibility Factor

A large dataset covering wide ranges of pressures and temperatures for estimating the *z*-factors of gasses was collected from the literature (Buxton and Campbell, 1967; Elsharkawy and Foda, 1998; McLeod, 1968; Robinson Jr and Jacoby, 1965; Simon and Briggs, 1964; Whitson and Torp, 1981; Wichert and Aziz, 1972). The minimum, maximum and average values of the data are reported in **Table 3. 12**. As can be observed in **Table 3. 12**, the data points include an extensive range of pse1udo-reduced temperatures, pressures, and compositions. As mentioned above, capability and consistency of a correlation for estimation of a specific parameter like the gas compressibility factor of gasses relies on the inclusiveness of the dataset used for its development. Hence, the method proposed in this work is expected to be consistent for the estimation a range of samples.

Table 3. 12 Range and corresponding statistical parameters of the input/output data utilized in the development of the model for the prediction of gas compressibility factor.

Property	**Min.**	**Max.**	**Average**
Pressure ,psi	154	7026	2820
Reservoir temperature, ˚F	40	300	147
Methane	17.27	97.48	71.18
Ethane	0	28.67	3.86
Propane	0	13.16	1.44
Iso-Butane	0	2.23	0.21
N-Butane	0	3.10	0.36
Iso-Pentane	0	2.85	0.18
n-Pentane	0	0.79	0.10
Hexane	0	2.68	0.20
Heptane plus	0	8.17	0.64
Mw C7+	0	150	50
SG C7+	0	0.90	0.31
Hydrogen sulfide	0	73.85	13.92
Carbon dioxide	0	54.46	6.00
Nitrogen	0	25.15	1.83
Tpr	0.97	1.96	1.46
Ppr	0.17	10.19	3.75
z-factor	0.40	1.241	0.86

CHAPTER 4

4. Model development

The development of predictive models plays a key role in the accurate estimation of thermo-physical properties, petroleum reservoir properties, enhanced oil recovery processes, production operations, drilling technology, and fluid flow in porous media. Based on the recent progress in the development of predictive models, the petroleum industry has become aware of the immense potential offered by intelligent systems.

The daily working life of petroleum professionals consists of highly complex and dynamic problems and high-stakes decisions. During the last two decades, the petroleum industry worldwide has seen a rapid increase in the number of artificial intelligence applications. This upsurge in the number of applications of artificial intelligence (AI) is due to the availability of human expertise and the publication of a large number of case studies. Artificial intelligence is the science and engineering of making intelligent machines. AI is devoted to designing ways to make computers perform tasks that were previously thought to require human intelligence.

4.1. Artificial Intelligence Technology

As previously mentioned, reliable and powerful predictive means are required by the petroleum industry for the accurate estimation of petroleum reservoir fluid properties. This is occasioned by the unwieldy and costly nature of the current methods. To this end, a number of efficient intelligent algorithms have been developed and employed, with the aim of replacing the old methodologies. These intelligent approaches have not previously been applied for the prediction of petroleum reservoir fluid properties.

Within the industry, there are many different petroleum fluid properties requiring the prediction of thermodynamic phase behavior. These include: solution gas oil ratio, reservoir oil viscosities, bubble and dew point pressures, oil formation volume factor, asphaltene scaling, wax disappearance temperature and natural gas deviation factor.

Artificial intelligence methods have in recent years gained popularity in solving complex nonlinear problems, and have shown the potential to be applied in reservoir modeling and characterization. These predictive models can be applied to solve complex practical problems relating to underground hydrocarbon reservoirs, which cannot be physically inspected and measured.

The models can be effectively applied to reservoir engineering, production operations, drilling technology, fluid flow in porous media, well-testing and well-logging. Furthermore, models developed on the basis of artificial intelligence methods can perform prediction and generalization rapidly once trained.

4.1.1. Least Square Support Vector Machine

The machine-learning community regards the SVM computational strategy to be a reliable means to analyze data, solve regression and classification problems, and recognize patterns (Eslamimanesh et al., 2012a; Suykens and Vandewalle, 1999). As a consequence, the SVM approach was introduced to solve, in particular, classification problems, utilizing hyper-planes for defining the decision borders between the data related to its various classes (Suykens and Vandewalle, 1999). Based on the initial SVM algorithm equation, primary function f(x) is defined as below (Farasat et al., 2013; Shokrollahi et al., 2013; Suykens et al., 2002b):

$$f(x) = w^T \varphi(\mathrm{x}) + b \tag{4.1}$$

where w^T , $\varphi(x)$, x, and b are the transposed output layer vector, the feature map, a vector of dimension n, and the bias, respectively. To obtain *w* and *b*, the following equation has been presented as a cost function (Suykens et al., 2002b):

$$Cost\ function = \frac{1}{2} w^T + c\sum_{k=1}^{N} (\xi_k - \xi_k^*) \tag{4.2}$$

To satisfy constraints:

$$\begin{cases} y_k - w^T \varphi(x_k) - b \le \varepsilon + \xi_k, & k = 1,2,....,N \\ w^T \varphi(x_k) + b - y_k \le \varepsilon + \xi_k^*, & k = 1,2,....,N \\ \xi_k, \xi_k^* \ge 0, & k = 1,2,....,N \end{cases} \tag{4.3}$$

where x_k *and* y_k stand for k^{th} input variable data, and k^{th} output variable data, respectively. The ε stands for the established accuracy of the function approximation. The ξ_k and ξ_k^* stands for slack variables. Here it is valuable to mentioned that if we consider a low value of ε for developing a an accurate model, some data points may be out-domain of the ε accuracy. Therefore, the application of the slack variables is required to determine the permitted margin of inaccuracy. As a matter of fact, the *c* in Eq. (4.2) is recognized as the adjustable parameter of the SVM algorithm, which controls the error difference from the wanted ε. To reach a minimization of the cost function, utilization of the Lagrangian is needed (Hemmati-Sarapardeh et al., 2013c; Suykens et al., 2002b):

$$L(a, a^*) = -\frac{1}{2}\sum_{k,l=1}^{N} (a_k - a_k^*)(a_l - a_l^*) K(x_k, x_l) - \varepsilon\sum_{k=1}^{N} (a_k - a_k^*) + \sum_{k=1}^{N} y_k (a_k - a_k^*) \tag{4.4}$$

$$\sum_{k=1}^{N} (a_k - a_k^*) = 0\ , a_k, a_k^* \in [0, c] \tag{4.4a}$$

$$K(x_k, x_l) = \varphi(x_k)^T \varphi(x_l), \quad k = 1,2,\ldots,N \tag{4.4b}$$

where a_k and a_k*are identified as Lagrangin multipliers. In the end, the ultimate formula for the SVM algorithm is expressed as below (Hemmati-Sarapardeh et al., 2013c):

$$f(x)=\sum_{k=1}^{N}(a_k-a_k^*)K(x,x_k)+b \tag{4.5}$$

For the determination of $a_k, a_k^*, and\ b$, a quadratic programming problem needs to be solved. To this end, Suykens and Vandewalle (Pelckmans et al., 2002; Suykens and Vandewalle, 1999) used the least-square modification (LSSVM) to improve the SVM algorithm. To introduce the LSSVM algorithm, Suykens and Vandewalle (Pelckmans et al., 2002; Suykens and Vandewalle, 1999) reformulated the SVM as below (Arabloo et al., 2013; Hemmati-Sarapardeh et al., 2013c; Rafiee-Taghanaki et al., 2013; Suykens et al., 2002b):

$$Cost\ function=\frac{1}{2}w^Tw+\frac{1}{2}\gamma\sum_{k=1}^{N}e_k^2 \tag{4.6}$$

Subjected to the following constraint constraints (for k=1,...,N):

$$y_k=w^T\varphi(x_k)+b+e_k \tag{4.7}$$

where γ and e_k are the adjustable parameter related to the LSSVM approach and the deviation variable, respectively. The following equation (Lagrangian) is expressed to solve the problem (Hemmati-Sarapardeh et al., 2013c):

$$L(w,b,e,a)=\frac{1}{2}w^Tw+\frac{1}{2}\gamma\sum_{k=1}^{N}e_k^2-\sum_{k=1}^{N}a_k\left(w^T\varphi(x_k)+b+e_k-y_k\right) \tag{4.8}$$

As a result, for solving the problem, the derivatives of Eq. (4.8) should be considered equal to zero. To this end, below equations are expressed:

$$\begin{cases}\dfrac{\partial L}{\partial w}=0\ \Rightarrow w=\sum_{k=1}^{N}a_k\varphi(x_k)\\ \dfrac{\partial L}{\partial b}=0\ \Rightarrow \sum_{k=1}^{N}a_k=0\\ \dfrac{\partial L}{\partial e_k}=0\ \Rightarrow a_k=\gamma e_k,\ k=1,2,...,N\\ \dfrac{\partial L}{\partial a_k}=0\ \Rightarrow w^T\varphi(x_k)+b+e_k-y_k=0\ \ k=1,2,...,N\end{cases} \tag{4.9}$$

Eq. (4.9) indicates that there are 2N+2 equations and 2N+2 unknown parameters $(a_k, e_k, w, and\ b)$. Hence, the parameters of LSSVM are acquired by solving the system of equations presented in Eq. (4.9) (Suykens et al., 2002b). As mentioned already, γ is one of the adjustable parameter of LSSVM algorithm. Meanwhile, as either of the LSSVM and SVM are kernel-based methods, the parameters of the kernel functions should be considered in the same manner as for other tuning parameters. The RBF kernel function is formulated as follows (Farasat et al., 2013; Fazavi et al., 2013; Taghanaki et al., 2013):

$$K(x, x_k) = \exp(-\| x_k - x \|^2 / \sigma^2) \tag{4.10}$$

where σ^2 is recognized as an adjustable parameter of the LSSVM algorithm. Consequently, σ^2 and γ are two adjustable parameters of LSSVM methodology with the RBF kernel function, which should be tuned with a reliable optimization technique (Shokrollahi et al., 2013; Suykens et al., 2002b).

In the LSSVM approach developed in this study, the mean square error (MSE) has been used as follows (Arabloo et al., 2013; Farasat et al., 2013; Shokrollahi et al., 2013):

$$MSE = \frac{\sum_{i=1}^{n} (Z_{rep./pred_i} - Z_{\exp_i})^2}{n} \tag{4.11}$$

where Z is the reservoir fluid properties, subscripts *rep./pred*. and *exp*. express the values estimated by the LSSVM model developed in this study, and actual data, respectively, and *n* stands for the number of samples from the initial population. In the present work, the original LSSVM approach introduced by Suykens and Vandewalle (Suykens and Vandewalle, 1999) is employed for the determination of reservoir fluid properties through a performance evaluation analysis. The MATLAB code developed in this study for the least square support vector machine algorithm is presented in **Appendix A**.

4.1.2. Artificial Neural Network

An artificial neural network (ANN) is an advanced model to process and classify information which simulates the biological neural network in the human brain and is

based on the mathematical systemitization of processes which happen in the brain (Al-Bulushi et al., 2007; Hegeman et al., 2007; Zabihi et al., 2011). Multilayer-perceptron (MLP) denotes the advanced, and well researched type of ANN in regression and classification, applying a feed forward, supervised and hetero-associative model (Rafiq et al., 2001).

The MLP and many other neural networks use an algorithm named back-propagation. By using back-propagation, the input data is repetitively imported to the neural network. With each step that outputs are generated, the values will compare to measured outputs and consequently, error is calculated. This error is then provided back (back-propagated) to the ANN and used to regulate the weights such that the error declines with each iteration and the neural model produces accurate outputs which are close to the desired targets. This procedure is recognized as "training" (Arulampalam and Bouzerdoum, 2003).

In artificial neural networks, information handling is executed in numerous simple separate processors, which are called neurons. When inputs are imported to the neuron, they are multiplied by the related weight of each link. Then the bias (threshold) of the neuron is supplied for the summation of weighted inputs. Each neuron uses a transfer function which is applied on inputs to produce outputs. In the artificial neural network approach, the data from the input neurons are propagated through the network via weighted interconnections (Mohammadi and Richon, 2010). Every *i* neuron in a *k* layer is connected to every neuron in adjacent layers. The activation function of the exponential sigmoid function has traditionally been utilized to develop ANNs (Eslamimanesh et al., 2011a) as following:

$$f(x) = \frac{1}{1+e^{-x}} \tag{4.12}$$

where *x* stands for parameter of activation function. A bias term, *b*, is associated with every interconnection in order to introduce a supplementary degree of freedom. The expression of the weighted sum, *S*, to the *i*th neuron in the *k*th layer (k ≥ 2) is (Mohammadi and Richon, 2010)

$$S_{k,i} = \sum_{j=1}^{N_{k-1}} \left[(w_{k-1,j,i} I_{k-1,j}) + b_{k,i} \right] \tag{4.13}$$

where *w* is the weight parameter between each neuron-neuron interconnection. Using this feed-forward network with activation function, the output, *O*, of the *i* neuron within the hidden *k* layer is (Mohammadi and Richon, 2010)

$$O_{k,i} = \frac{1}{1+e^{-\left(\sum_{j=1}^{N_{k-1}}\left[\left(w_{k-1,j,i}I_{k-1,j}\right)+b_{k,i}\right]\right)}} = \frac{1}{1+e^{-S_{k,i}}} \tag{4.14}$$

The MATLAB code developed in this study for the artificial neural network algorithm is presented in **Appendix B**.

4.1.3. Decision Tree

The decision tree constructs models with a tree structure for the purpose of regression or classification. It divides a dataset into smaller subsections while a related decision tree is incrementally developed.. The outcome is a tree with decision nodes and leaf nodes. A decision node has two or more divisions, each signifying values for the feature tested. A leaf node represents a decision on the target. The top decision node in a tree relates to the best predictor called a root node.

Decision trees can process both categorical and numerical data (Erdogan et al., 2001; Heinze et al., 1995; Sethi and Chatterjee, 1977). There are three kinds of decision tree: CRT, CHAID and Exhaustive CHAID, and Quest. The procedures for the three types follow the following phases: start tree building by allocating the nodes to classes, stop tree building; reach the optimum tree selection and perform cross-validation. CRT undertakes tree "pruning" before creating the optimum tree selection, while the CHAID method implements statistical tests at each step of splitting (Alkhasawneh et al., 2014; Chandra et al., 2010; Osei-Bryson, 2004).

The CRT (Classification and Regression Tree) is a recursive subdividing technique used for both regression and classification. The CRT is built by division of data sets into the subgroups using all of the predictor variables to repeatedly create two child nodes, using the whole data set. The best predictor is chosen using a diversity of measures. The objective is to yield subdivisions of the data which are as similar as possible to the target variable (Laughton et al., 2006; Tan et al., 2006; Wang et al., 2015).

The CHAID (Chi-Square Automatic Interaction Detector) technique is established based on the X^2-test of association. A CHAID algorithm is a decision tree that is created by repetitively dividing subgroups of the space into two or more child nodes. To control the best splitting at any node, any acceptable couple of groups of the predictor variables are combined until there is no statistically important alteration in the couple with respect to the target variable. This CHAID method works logically with communications between the independent variables that are directly accessible from an analysis of the tree. The final nodes recognize subsections defined by diverse groups of independent variables (Gandomi et al., 2013; Tan et al., 2006). There is no assurance that the original CHAID procedure will discover the best split for all those inspected because it uses the last split tested. The Exhaustive CHAID procedure tries to solve this issue by continuing to combine groups, regardless of significance level, until only two groups stay for each predictor. It then uses the split with the major importance value rather than the last one tried. The Exhaustive CHAID needs additional computer time (Alkhasawneh et al., 2014; Tan et al., 2006).

The QUEST (Quick-Unbiased-Efficient Statistical Tree) is a binary split decision tree process for classification and data analysis. The QUEST can be used with univariant or linear grouping splits. An exceptional characteristic is that its attribute selection technique has insignificant bias. If all the attributes are uninformative with respect to the class attribute, then each has approximately the identical chance of being selected to split a node time (Alkhasawneh et al., 2014; Tan et al., 2006). Finally, here it is worthwhile to note that a regression tree is similar to a classification tree, except that the Y variable takes ordered values and a regression model is fitted to each node to give the predicted values of Y (Loh, 2011).

The MATLAB code developed in this study for the decision tree algorithm is presented in **Appendix C**.

4.1.4. Gene Expression Programming

As mentioned already, the GEP mathematical algorithm is applied in this study to develop reliable empirical models for the determination of petroleum reservoir fluid properties. The GEP (Ferreira, 2001) approach is a new and modified version of a genetic

algorithm combined with genetic programming, which is implemented for solving regression and classification problems by employing populations of individuals, choosing them in keeping with fitness, and presenting genetic variation utilizing one and/or more genetic operator (Mitchell, 1998). The nature of individuals is the fundamental difference between GA, GP and GEP algorithms (Ferreira, 2001). The individuals in GA, GP and GEP algorithms are the chromosomes or linear strings of fixed length, parse trees or nonlinear entities of different shapes and sizes, and the chromosomes or genome and or linear strings of fixed length which are subsequently presented as nonlinear entities of different shapes and sizes, respectively (Ferreira, 2001).

In the GEP (Ferreira, 2001) algorithm, the structure of the genes allows encoding of any program for effective evolution and development of solutions (Ferreira, 2006). The GEP (Ferreira, 2001) mathematical algorithm employs two elements, namely, the expression tree (ET) and the chromosome (the chromosome has the role of encoder for the candidate solution which is translated into an expression tree). Each genetic chromosome involves terminals with constants and variables and functions structured in one and/or more genes of equal length (Teodorescu and Sherwood, 2008). The constants are produced by the GEP algorithm in a range selected by the employer while the functions and variables are recognized as input data. Additionally, the gene consists of a tail made only of terminals, and a head made of functions, in addition to terminals including variables and constants. The head length (h) is recognized as an input parameter for the GEP mathematical algorithm while the tail length (t) is expressed as follows:

$$t = h(n-1)+1 \tag{4.15}$$

where *t* stands for the tail length of the gene, *h* shows the head length, and *n* is the largest arity of the functions utilized in the gene's head. To better view of GEP procedure, **Fig. 4. 1** is an example of a two-gene chromosome composed of four functions including -, ∗, / and Q, and three terminals including x, y and z, together with its decoded ET and the corresponding mathematical expression which is formulated as ($\sqrt{x/z}$) - (x*y).

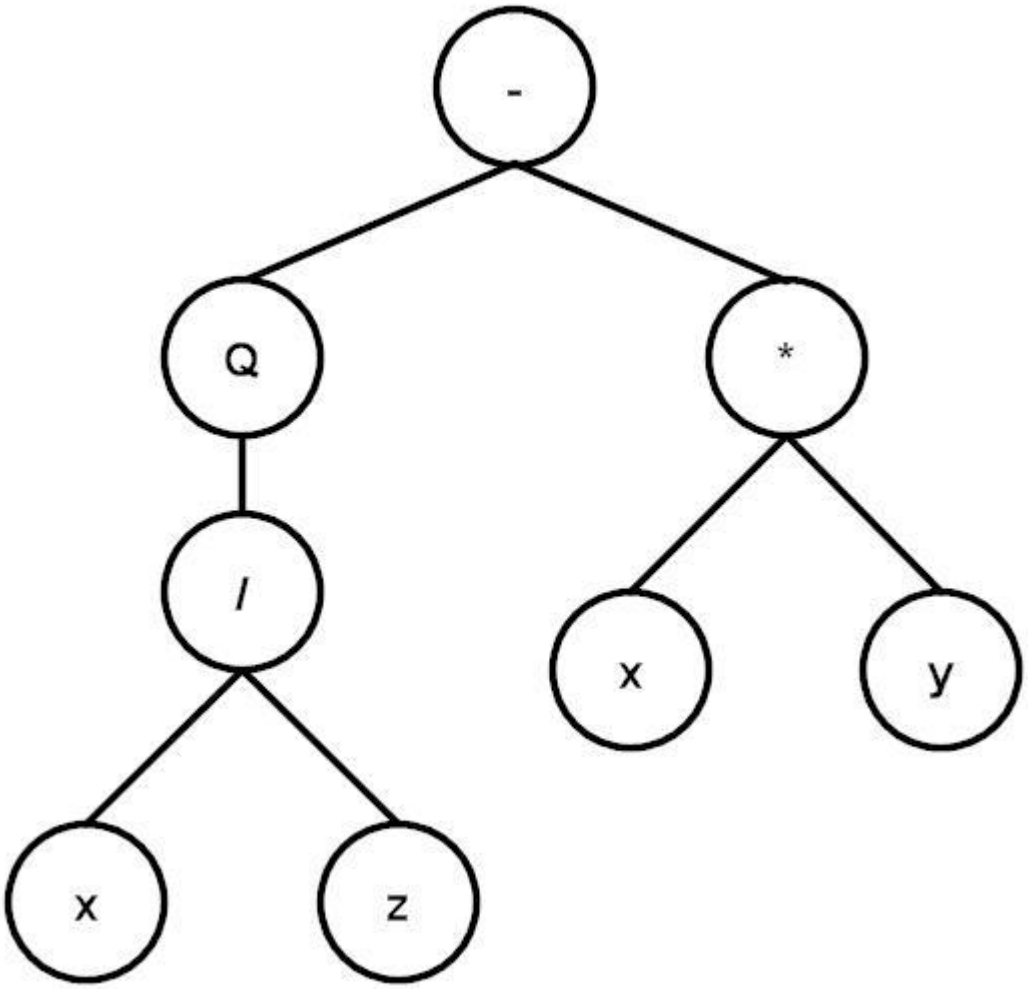

Fig. 4. 1 A typical two-gene chromosome with its corresponding mathematical expression; Q is the square root function.

The procedure presented by Ferreira (Ferreira, 2006) has been followed in this study to develop empirical models using a GEP algorithm in order to predict the petroleum reservoir fluid properties as follows (Ferreira, 2006):

I. The population individuals initializing, which is based on counting the random made chromosomes of a certain number of individuals by setting various stated correlations;
II. The population individuals fitting, considering fitness functions (cases);
III. The population individuals selecting, in keeping with their fitness in order to replicate with modifications;

IV. The new population individuals are dealt with by implementing the same procedure, with modifications, as for the confrontation of the selection environment and the genomes expression, selection, and duplication;

V. The stages above are followed for a certain number of generations or until an optimum solution has been established (convergence of the algorithm in keeping with the criteria defined).

In order to present capable and reliable empirical models for the calculation of petroleum reservoir fluid properties, various input variables have been considered. Higher valued input variables can suppress the impact of the smaller ones during the training phase of a mathematical algorithm, such as in the GEP method. To overcome such an obstacle, and in order to make the GEP algorithm fit for the prediction of petroleum reservoir fluid properties, all data points should be adequately processed and well scaled prior to input to the GEP. Although normalizing the data points is not a necessity in the estimation process by GP-based methods, better results are generally acquired by normalizing the parameters (Alavi et al., 2010). Thus, all data points related to the input and output variables are normalized as follows:

$$r_n = \left(\frac{r}{1.5 \times r_{\max}} \right) \times 0.8 + 0.1 \tag{4.16}$$

where r_n stands for the data points normalized, r indicates the actual data and r_{max} is the maximum value of the data (Srinivas et al., 2010). In the next step, the normalized data points were returned to their original values at the end of modeling process.

The MATLAB code developed in this study for the gene expression programming algorithm is presented in **Appendix D**.

4.1.5. Adaptive neuro-fuzzy inference system

The adaptive neuro-fuzzy inference system proposed by Jang in 1993 (Jang, 1993) is viewed as a smart hybrid methodology composed of, and/or combining both fuzzy logic and artificial neural networks. The production of reasonable results by means of the implementation of the simple fuzzy inference system (FIS) is directly associated with expert-knowledge rules. The absence of such rules is recognized as a major disadvantage of this intelligent system.

The definition of intelligent rules and appropriate membership functions can help to avoid imperfect results. The ANFIS combines the strengths of fuzzy logic and artificial neural networks in order to develop models on the basis of an FIS with optimized rules and membership functions. The ANFIS methodology relies on the assumption that ANN trains the whole network and FIS structure to pursue the fuzzy logic representation. It is worth mentioning that the structures of rules and functions in models developed based on the ANFIS method are similar to those which have been proposed by the concept of fuzzy logic (Rahimzadeh Kivi et al., 2013).

The MATLAB code developed in this study for the adaptive neuro-fuzzy inference system algorithm is presented in **Appendix E**.

4.2. Structural Based Models

There are two main molecular structures based methods namely the group contribution (GC) and quantitative structure–property relationship (QSPR) strategies. As a result, molecular descriptors are one of the most important ingredients in developing a GC and/or QSPR model. Moreover, the molecular descriptors are the final result of a logical and mathematical procedure in technical terms, which transform chemical structure information encoded within a symbolic representation of a molecule into a useful number or the result of some standardized experiment (Gharagheizi et al., 2013). The optimized molecular structures are a necessity to calculate molecular descriptors. In order to calculate the molecular descriptors, the optimized molecular structures must be loaded into Dragon software (Eslamimanesh et al., 2011b). It is capable of calculating over 3000 descriptors from several diverse classes. The descriptors obtained were carefully analyzed and those which were not able to be calculated for some compounds were completely neglected.

In order to develop a reliable group contribution and/or QSPR model, the chemical structures of all the components used in a database should be examined thoroughly to find out the most efficient sub-structures for the estimation of thermos-physical properties. In other words, having defined the compounds present in the database, the chemical structures of all of the studied compounds have been analyzed to recognize the functional groups.

To represent the GC and/or QSPR models, a sequential search mathematical strategy is applied to reduce the number of molecular descriptors to several tens of descriptors. The sequential search strategy implements an extensive search throughout the feature subsets (Dudek et al., 2006). As a first step, a single feature that leads to the best estimation is selected. Next, sequentially, each feature is individually added to the current subset and the errors of resulting models are quantified. The feature that is the best in decreasing the error is incorporated into the subset. Thus, in each step, a single best feature is added, resulting in a sequence of nested subsets of features. The procedure stops when a specified number of features are selected. When this feature is to be selected as the one that improves the quality of the model, the procedure is stopped. In other words, the basic idea is to replace each variable one at a time with all the remaining ones and see whether a better model is obtained. The disadvantage of sequential forward selection (SFS) is that if several features collectively are good predictors, but alone each is a poor prediction, none of the features may be chosen (Dudek et al., 2006). However, it should be mentioned that the SFS with percentage of average absolute relative deviation as an objective function is successfully used for selection of variables.

4.3. Intelligent Optimization Algorithms

4.3.1. Coupled Simulated Annealing

To avoid local optima, simulated annealing methodology has been utilized (Atiqullah and Rao, 1993; Fabian, 1997; Vasan and Raju, 2009) as an optimization strategy. The idea is based on allowing moves which lead to solutions of worse quality than the present solution for avoiding the problem of local optima. The coupled simulated annealing (CSA) optimization technique, as an improved version of simulated annealing (SA), has been proposed to avoid the problem of local optima. The innovative principles of CSA methodology was reported by Suykens et al.(Suykens et al., 2001). In order to avoid the problem of local optima in non-convex complications, they applied coupling to local optimization developments. Additionally, Xavier et al. (Xavier-de-Souza et al., 2010) used CSA as a reliable optimization technique for the improvement of accuracy in the final solution of their problem. Further, the coupled optimization techniques like CSA could be

more effective if the communication of a coupling strategy is decreased to minimum (Koch, 2005).

The below formula defines the acceptance possibility of function A with coupling term ρ (Chamkalani et al., 2013a):

$$A_{\theta}(\rho, x_i \to y_i) = \frac{\exp\left(\frac{-E(y_i)}{T_k^{ac}}\right)}{\exp\left(\frac{-E(y_i)}{T_k^{ac}}\right) + \rho} \tag{4.17}$$

with $A_{\theta}(\rho, x_i \to y_i)$ the acceptance possibility for every $x_i \Theta$, $y_i \in \gamma$ and $\gamma \gamma \Theta$.

Therefore, γ expresses the set of all probable states and the set $\Theta = \{x_i\}_{i=1}^{q}$ is indicated as the set of current states of *q* minimizers. Moreover, the variance σ^2 of *A* is as follows (Chamkalani et al., 2013a):

$$\frac{1}{q}\sum_{\forall x_i \in \Theta} A_{\Theta}^2 - \frac{1}{q^2} \tag{4.18}$$

Consequently, the coupling term ρ is presented as follows (Chamkalani et al., 2013a):

$$\rho = \sum_{x_{j \in \Theta}} \exp\left(\frac{-E(y_i)}{T_k^{ac}}\right) \tag{4.19}$$

4.3.2. Genetic Algorithm

The genetic algorithm proposed by Holland (Holland, 1975) is capable of exploiting the information on an initially unknown search space to bias the subsequent search into promising subspaces. On the basis of the Darwinian principle of 'survival of the fittest', the genetic algorithm can explore the finest coordinate in the particular space after a series of repetitive calculations. As a result, artificial mutation, crossover and selection operators are the major ingredients of this searching process. To operate the

algorithm, an initial population, containing an already defined number of solutions under the title of individuals or chromosomes, is generated to switch the process on. The next step is to encode the components of populations into bit-string so-called chromosomes. Then, the nobility of the strings, normally named fitness, is evaluated in association with some functions representing the constraints of the issue.

4.3.3. Particle Swarm Optimization

The particle swarm optimization algorithm is proposed for solving the problems in continuous search space. This optimization technique was founded on an image of social interaction and communication (*e.g.* bird flocking and fish training) (Coulibaly and Baldwin, 2005; Eberhart and Kennedy, 1995). The particle swarm optimization algorithm employs social rules for searching in the design space by systematizing the trajectories of a series of independent particles. The location of each particle, which indicates a particular solution associated with the problem, is applied in order to calculate the tuned extent of the fitness function (Zendehboudi et al., 2013a). In the PSO approach, each particle has a different position during the optimization process, as its position is continuously altered to discover the solution in space.

On the basis of velocity and location, every particle moves around at each iteration. The cost function for each particle is assessed to rank its current position. As the particle velocity builds quickly (Coulibaly and Baldwin, 2005), Shi and Eberhart (Shi and Eberhart, 1998) proposed the concept of inertia weight (ω) to the classical form of the particle swarm optimization for decreasing the velocity of particles. Afterward, the particles velocities are updated stochastically as follows (Coulibaly and Baldwin, 2005; Shi and Eberhart, 1998):

$$V_i^{t+1} = \omega V_i^t \; C_1 r_1^t \; (P_i^t - X_i^t) + \; C_2 r_2^t \left(P_g^t - X_i^t\right) \tag{4.20}$$

$$X_i^{t+1} = X_i^t + V_i^{t+1} \tag{4.21}$$

where V_i^t stands for the velocity at iteration of t, X_i^t shows the particle location at iteration of t, ω is the inertia weight, r denotes the random number, P_g^t expresses the best ever particle position of the particle i, and P_i^t indicates the global best position in the swarm until iteration of t.

4.4. Computational Procedure

To develop our intelligent models, various input variables associated with each property have been selected, while the petroleum reservoir fluid properties are regarded as output variables. Three sub-data sets, the "Training" set, "Validating" set and the "Test" set, have been considered on the main databank in order to develop and check the models constructed by different artificial intelligence approaches. Here it should be noted that only two sub-data viz. training and test sets are considered for some analyses. Routinely, the "Training" set is employed for developing the main structure of the model, and the "Validating" set as well as the "Test (prediction)" set are utilized for visualizing the accuracy, capability, and reliability of the model obtained (Arabloo et al., 2013; Farasat et al., 2013; Shokrollahi et al., 2013). Here, it should be mentioned that the allocation process of the data is random.

The *K*-means clustering technique (Kamari et al., 2014e) has been utilized in the present study for establishing the relationship among different variables to establish whether they have relevance. The *K*-means clustering is recognized as a method which divides *n* observations into *k* clusters in which each observation belongs to the cluster with the nearest mean. In other word, the *K*-means clustering approach helps to assign and divide the datasets. However, the dataset may have to be grouped by a different method before the *K*-means clustering is utilized.

To this end, the *K*-Fold cross validation method is applied in this study. The approach considers k=10, then it partitions the dataset into 10 subsets. Holding one subset as a test/validation set, the model will be trained with the remaining 9 subsets. Afterward, the same procedure is repeated 10 times to find the best data division. In other words, the dataset is partitioned into subsets D_1, D_2, …, D_{10}. In the first iteration, D_1 is taken as test/validation set and $\{D_2, D_3, ..., D_{10}\}$ as training sets. In the second iteration, D_2 is considered as the test/validation set and $\{D_1, D_3, D_4, ..., D_{10}\}$ as training sets. Finally, the average of testing and validation accuracy is calculated over 10 runs.

In this study, we selected different data divisions (80:10:10, 80:20, *etc.*) to study and predict the petroleum reservoir fluid properties. As a result, the value/percentage of assigned data for training sets must be balanced and reasonable to avoid an over-fitting problem, and also to achieve an accurate and tested prediction. If the value/percentage

of assigned data is high for a training set, then an over-fitting problem may occur for the predictive model. Moreover, if the value/percentage of assigned data is low, then we cannot develop a strong and reliable model for prediction targets. The results obtained in this study show that the model developed with the division of 80:20 (80% of data points for training and 20% for test sets) is the most appropriate and reliable because of its balanced accuracy in the testing and validation phases.

As noted earlier, small value input parameters may be affected by higher values during the training phase of model development by intelligent approaches. To overcome this problem, the available data points should be normalized to achieve a good prediction by intelligent approaches. Therefore, all available data points related to input/output variables are normalized as below:

$$X' = \frac{X - \mu}{\eta} \tag{4.22}$$

where X' denotes the initial value or actual data, X expresses the normalized values for the actual data, μ stands for the mean and finally η is the standard deviation. In this method, each of the input/output variables is normalized so that the mean and standard deviation of normal variable are 0 and 1, respectively (Karambeigi et al., 2011). It should be mentioned that normalizing the data has no impact on the results obtained because, finally, all normalized data points will be returned to their original values.

4.5. Methods for Performance Evaluation

4.5.1. Leverage Approach

The Leverage technique is composed of statistical analysis comprising residual errors and the Hat matrix that consider the actual values of data and the estimated values of the petroleum reservoir fluid properties (Eslamimanesh et al., 2012b; Mohammadi et al., 2012a; Mohammadi et al., 2012c). In fact, the use of a mathematical model is the main application criterion of the Leverage algorithm. The Hat matrix embedded in the Leverage technique is presented as follows (Eslamimanesh et al., 2012b; Gharagheizi et al., 2012b; Goodall, 1993; Gramatica, 2007; Mohammadi et al., 2012a; Rousseeuw and Leroy, 2005):

$$H = X(X^t X)^{-1} X^t \quad (4.23)$$

where *t* stands for the transpose matrix and *X* refers to a matrix containing *k* columns and *N* rows. The Hat values, which represent the viable region of the case under study, are characterized by the diagonal elements of the H matrix. Moreover, the outliers are usually detected on the basis of H values obtained from Eq. (4.23). The H indices and standard residual values are well described in the Williams plot. In general, a warning leverage (H*) is set to be 3p/n, where p is equal to the number of model coefficients plus one and the number of training data points is symbolized with n. If the leverage is 3, it means the data points are accepted with a standard deviation of ±3 with respect to the average (mean) value. If H [0, H*] and R [-3, 3] are the intervals in which the main part of the data are placed, the statistical accuracy of the technique is demonstrated in the defined domain in terms of predictive performance. It is important to note that acceptably high leverage is attributed to the condition where H is equal to or greater than H* and R is between -3 and 3. The data points in the intervals of R < -3 or 3 < R are recognized as the suspect data, known as poor high leverage. The presence of the outliers in computation and analysis may cause considerable error in the model output, leading to false decisions.

4.5.2. Variables Relevancy Analysis

A sensitivity analysis is performed in order to show the degree of reliability of the petroleum reservoir fluid properties selected as input variables (*e.g.* saturation pressure, reservoir temperature, gas specific gravity, and API gravity) on the output variables estimated by the developed intelligent models. Hence, the relevancy factor (r) (Chen et al., 2014) is utilized in this study to measure the degree of effect of each input variable used for the determination of petroleum reservoir fluid properties.

In terms of the relevancy factor approach, an input variable has a higher influence on the output parameter if the calculated absolute value of r between the input and output variables is greater than the r values for other input variables. As a consequence, the positive or negative influence of input variables (*e.g.* saturation pressure, reservoir temperature, gas specific gravity, and API gravity) on the outputs is not determined by an absolute value of r. Consequently, the following equation is used to calculate the r values by means of the relevancy analysis (Hosseinzadeh and Hemmati-Sarapardeh, 2014):

$$r(Inp_k, \mu_g) = \frac{\sum_{i=1}^{n}(Inp_{k,i} - \overline{Inp_k})(\mu_i - \overline{\mu})}{\sqrt{\sum_{i=1}^{n}(Inp_{k,i} - \overline{Inp_k})^2 \sum_{i=1}^{n}(\mu_i - \overline{\mu})^2}} \quad (5)$$

where $Inp_{k,i}$ stands for *i*th value of the *k*th input variables and $\overline{Inp_k}$ denotes the average value of the *k*th input variables (*e.g.* bubble point pressure, reservoir temperature, gas gravity, and oil gravity), μ_i indicates the *i*th value of the outputs determined by the developed intelligent models, and $\overline{\mu}$ is the average value of the outputs determined the developed intelligent models.

CHAPTER 5

5. Results & Discussion

In this chapter, models based on the least square support vector machine, artificial neural network, decision tree, and adaptive neuro-fuzzy inference system, are applied in order to gauge various petroleum reservoir fluid properties. Then, different graphical plots are sketched in order to analyse the applicability domain of the values predicted by these models in relation to different ranges of petroleum reservoir fluid properties. Furthermore, different error parameters viz. average absolute relative deviation, R-squared, average relative deviation, standard deviation, and root mean square error, are employed to statistically evaluate the accuracy of the models that have been developed and applied. This data is presented in **Appendix F**.

The effects of the input variables on the accuracy of prediction of the petroleum reservoir fluid properties are investigated and discussed by means of the relevancy analysis approach.

Finally, a list of the limitations/drawbacks of the deterministic methods discussed in this thesis are highlighted. Additionally, the importance of this study for the advancement of modeling of petroleum reservoir fluid properties is discussed in terms of fluid flow in porous media, enhanced oil recovery, PVT analysis of the reservoir fluid, and production technology.

5.1. New Equation for Surfactant Retention

To develop a reliable model for the determination of surfactant retention, the GEP algorithm has been employed. After an acceptable numbers of generations were considered in the development of the model, a neutral gene was added to the model in order to increase the accuracy. Furthermore, various mathematical functions and basic arithmetic operators were implemented to obtain the optimal model with the highest possible accuracy. Additionally, a function according to mean absolute error and R^2 was selected to compute the overall fitness associated with the evolved programs. The program was run until there was no longer a significant improvement in the accuracy and performance of the several models developed with the functions employed. The final form obtained can be expressed as follows:

$$R = A - B + 0.22481 \tag{5.1}$$

$$A = C_{co-solvent} + 0.001\left[\left(MW_{Sur}{}^{\frac{1}{2}}\right)T^2(9.3489 - pH)\right]^{\frac{1}{3}} - \left(\frac{0.0022395T}{0.1K_{abs} - pH}\right) \tag{5.2}$$

$$B = (10TAN - 5.2855) \times 0.001\left[(100MR - 1000C_{co-solvent} + 30.723)^{\frac{1}{2}} - \frac{622.68}{S_{PD}}\right] \tag{5.3}$$

where R denotes the retention of surfactant, K_{abs} the absolute permeability, TAN is the total acid number of the oil, T denotes reservoir temperature, $C_{co\text{-}solvents}$ is the co-solvent concentration, S_{PD} the salinity of the polymer drive, pH is the maximum effluent pH, MR is the value of mobility ratio, and MW_{Sur} is the average molecular weight of the surfactant solution.

The performance of the newly developed model for the prediction of surfactant retention in porous media in terms of the statistical model validation parameters are given in **Table 5. 1**. As can be seen in this Table, the R^2 in the testing phase is 0.9464. Furthermore, E_a or average absolute relative deviation in the testing stage is reported as 9.66 %. These results indicate that the newly developed model predicts the surfactant retention values with an acceptable accuracy.

Table 5. 1 Statistics error parameters of the developed model for prediction of the surfactant retention.

Performance	E_a %	E_r %	SD	RMSE	R^2
Total	16.62	-3.6473	0.0580	0.0275	0.9246
Training	18.50	-4.7640	0.0562	0.0294	0.9213
Testing	9.66	0.4840	0.0144	0.0187	0.9464

Fig. 5. 1 shows a point-to-point comparison between values obtained by the new model and the literature-reported surfactant retention data. As can be seen in this Figure, the points are in good agreement. **Fig. 5. 2** shows a comparison between predicted and real surfactant retention in porous media. As is clear from the figure, approximately all data corresponding to the training and testing stages straddle the Y=X line, indicating that there is a good fit between the newly developed model predictions and the real data taken from the literature.

Fig. 5. 3 provides more statistical information for the model performance evaluation and depicts the magnitudes of relative error percentage versus the real data employed in both the training and testing phases. Based on this study, a small margin of error is noticed. Therefore, it can be concluded that the newly developed model is efficient in estimating surfactant retention in porous media during the application of the chemical flooding recovery method. The accuracy of the model is sufficient for its utilization in the design of EOR processes in petroleum reservoir disciplines.

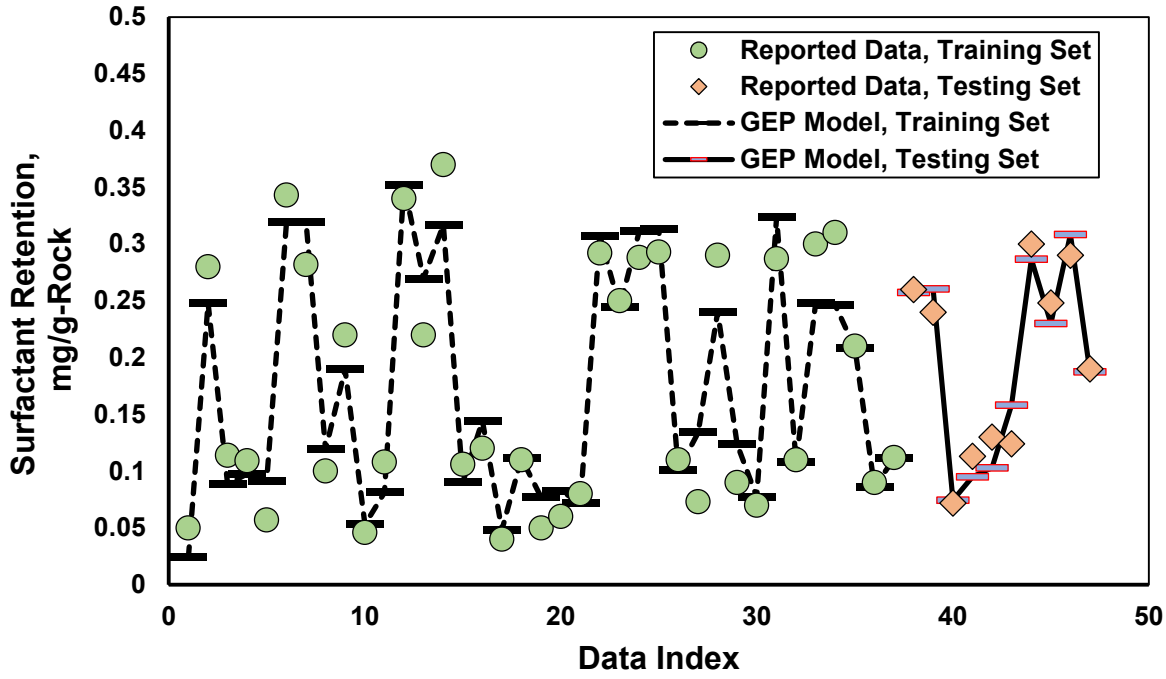

Fig. 5. 1 Point-to-point comparison between the results of the developed empirical correlation and literature-reported values of the surfactant retention in porous media.

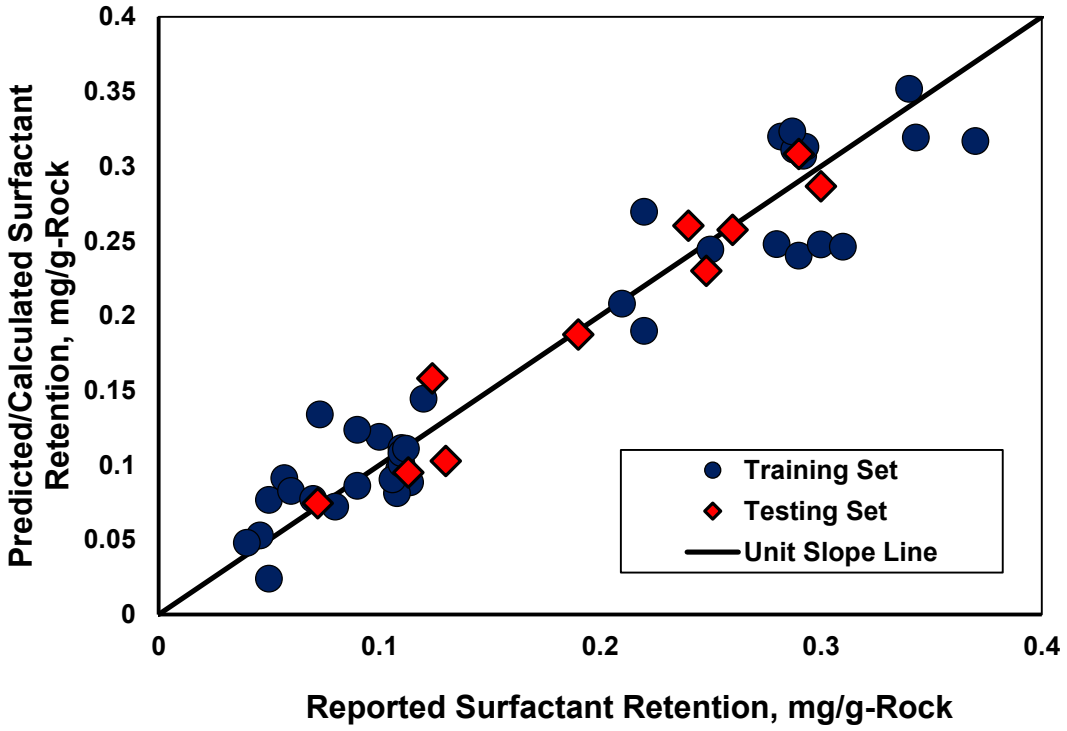

Fig. 5. 2 Crossplot for predicted values by the new empirical correlation and literature-reported values of the surfactant retention in porous media.

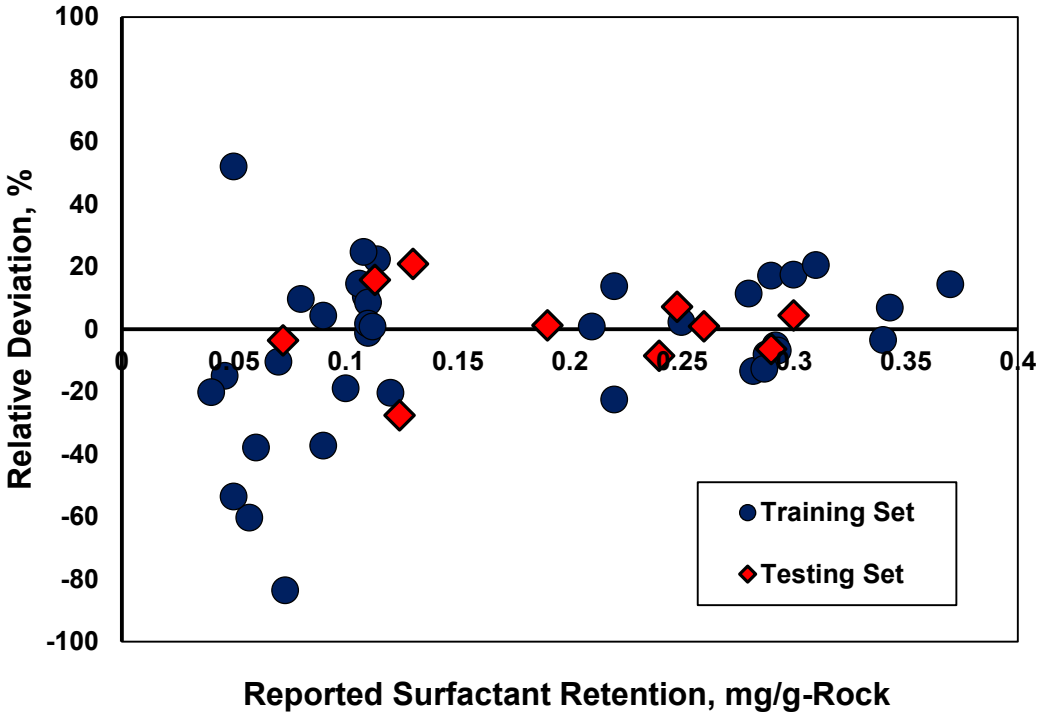

Fig. 5. 3 Relative deviations of the surfactant retention in porous media obtained by the newly proposed model from the database values.

Figs. 5. 4 and **5. 5** display the performance of the model developed in terms of R^2 and average absolute percent relative error (AAPRE). This model is for prediction of surfactant retention in comparison with predicted values by the intelligent artificial method called LSSVM (Yassin et al., 2013). For the testing stage, a previously published empirical correlation, namely Solairaj et al. (Solairaj, 2011; Solairaj et al., 2012) is used. **Figs. 5. 4** and **5. 5** graphically illustrate the calculated R-squared error and AAPRE for the all methods investigated in this study. From these figures it can be concluded that the new model performs better for the prediction of surfactant retention compared with LSSVM (Yassin et al., 2013) and Solairaj et al.'s correlation (Solairaj, 2011; Solairaj et al., 2012) based on the available database. Moreover, unlike LSSVM, the new model can provide a symbolic equation for correlation and determination of surfactant retention in porous media, which can be employed in reservoir engineering science and in issues associated with EOR processes.

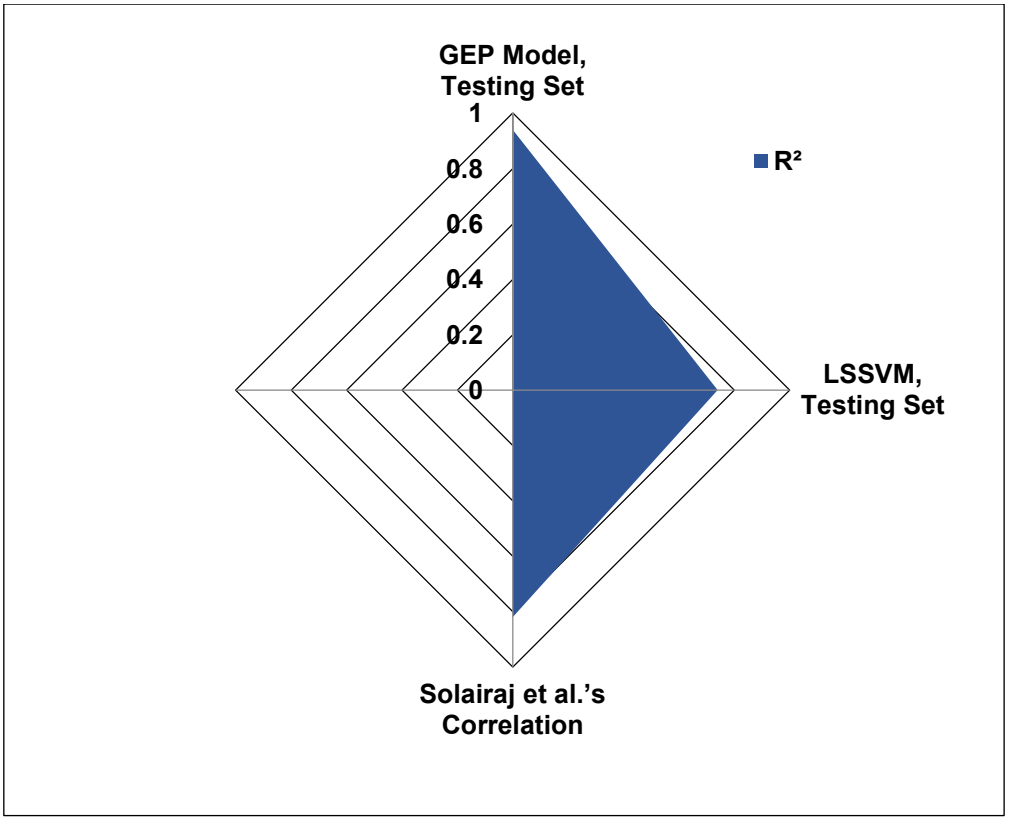

Fig. 5. 4 Comparison between the R^2 obtained by the newly developed correlation in the testing phase, the LSSVM values (Yassin et al., 2013) in the testing stage and Solairaj et al.'s correlation (Solairaj, 2011; Solairaj et al., 2012) for predicting the surfactant retention in porous media.

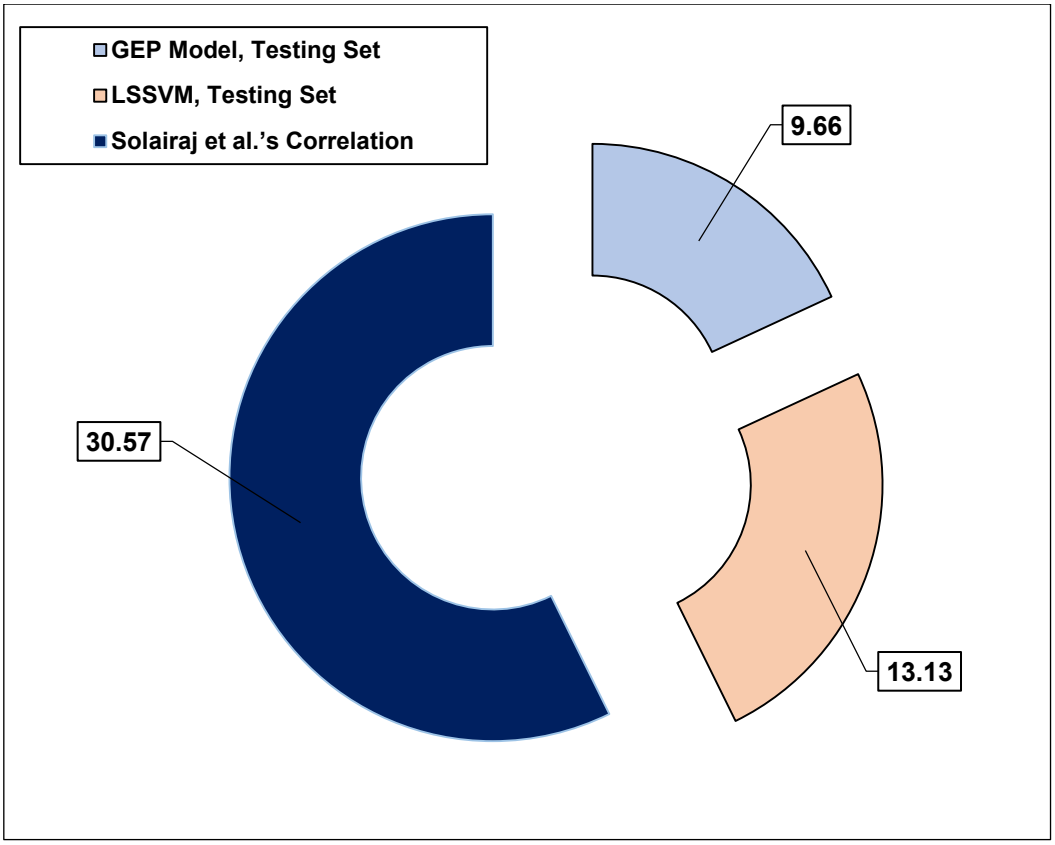

Fig. 5. 5 Comparison between the average absolute relative deviations (AAPRE) obtained by the newly developed correlation in the testing phase, the LSSVM values (Yassin et al., 2013) in the testing stage and Solairaj et al.'s correlation (Solairaj, 2011; Solairaj et al., 2012) for predicting the surfactant retention in porous media.

From these results it can be concluded that the method introduced in this study can result in excellent generalization and can be advantageously employed for the estimation of surfactant retention in porous media during surfactant/chemical flooding. In addition, the significance of the new method is its use of readily existing reservoir sample data. It can therefore be widely applied in situations where experimentally measured records are not available. Moreover, the method can be employed in reservoir engineering software developed for the simulation of EOR methods, in particular surfactant-based flooding, and can provide precise performance in predicting the retention of surfactant with regard to stated parameters.

5.2. New Equation for Dew Point Pressure of Gas Condensate Reservoirs

5.2.1. The GEP-based model

As indicated earlier, there are a number of different factors which affect the prediction capability of GEP-based models. Hence, many computation runs were performed on a trial and error basis to find optimum accuracy and simplicity. During the development of the GEP-based model for the prediction of DPP in gas condensate reservoirs, it was noticed that an increase in the number of permitted genes in the applied individual, and the maximum depth of ET, have an effect on the domain size of the solution space. The complexity of the evolved function rises and there is an increase in the run-time of the process.

The accuracy of models developed based on the GEP evolutionary approach is normally increased in relation to an increase in the number of genes and the depth of ET. Consequently, we applied two genes, AARD and R-squared as fitness functions, and a function set including *, +, -, ln and /. Specific gravity and molecular weight for heptane plus fractions, reservoir temperature, compositions of hydrocarbons including methane, ethane, propane, butanes, pentanes, hexanes, heptane-plus, and compositions of non-hydrocarbons including nitrogen, carbon dioxide, hydrogen sulfide were used as input variables. The optimum GEP-based model obtained for the estimation of DPP in gas condensate reservoirs is as follows:

$$P_d = A + B \tag{5.4}$$

$$A = \frac{13.145 - 4.942\, z_{C_1} + 1961.7\, z_{C_2} - 6212.71\, z_{C_4} + 39335.07\, (z_{C_4})^2 + 2097\, z_{C_5} - 3451.17\, z_{C_6} + 201.93\, z_{H_2S} - 0.065224 T_R}{0.0031904\, T_R + 0.094398} \tag{5.5}$$

$$B = \frac{1367.4 + 9.98\, z_{C_1} MW_{C_{7+}} - 1697.6\, z_{C_3} - 5096.8\, z_{C_7} + 358.09 \ln(z_{C_{7+}}) + 933.35\, z_{CO_2} + 1909.7\, z_{N_2}}{1.0214 - SG_{C_{7+}}} \tag{5.6}$$

where P_d (Psia) denotes the DPP in gas condensate reservoirs, MW_{C7+} is the molecular weight of the heptane-plus fraction, SG_{C7+} is the specific gravity of heptane plus fraction, T_R is reservoir temperature (°F) and $z_{C_1}, z_{C_2}, z_{C_3}, z_{C_4}, z_{C_5}, z_{C_6} z_{C_{7+}}, z_{CO_2}, z_{N_2}, z_{H_2S}$ are compositions of methane, ethane, propane, butanes, pentanes, hexanes, heptane-plus, carbon dioxide, nitrogen, and hydrogen sulfide, respectively.

5.2.2. Capability and precision evaluation of the GEP-based Model

In order to evaluate the capability of the model in predicting DPP, we used some important error functions including AARD, R-squared, average percent relative error, root mean square error and standard deviation. Furthermore, a graph that provides a cross-plot of actual data against the predicted values, using Eq. (5.4), as well as a relative error distribution plot, are presented. **Table 5. 2** summarizes the calculated errors for data predicted by the GEP-based model developed for DPP in gas condensate reservoirs.

The proposed model based on the GEP approach has an overall AARD of 7.8 % and an R-squared equal to 0.89. These values are acceptable for the proposed model taking into account the dataset used in its development. **Fig. 5. 6** shows the cross-plot of the actual data against the predicted values using Eq. (5.4), as well as the relative error distribution plot. As can be seen in **Fig. 5. 6**, there is good agreement between the actual data and values calculated using the GEP-based model proposed in this study. There is an acceptable match, as observed in the top panel of **Fig. 5. 6** between the reported and estimated DPP values using the GEP model. Additionally, the bottom panel of **Fig. 5. 6** illustrates satisfactory distribution of relative deviation around the zero line.

Table 5. 2 The overall performance of the model developed, for the training and testing phases, as well as comparison with other models in terms of statistical error analysis.

Method	AARD %	APRE %	SD	RMSE	R^2
Elsharkawy's model (Elsharkawy, 2002a)	15.3	-9.667	0.201	891.075	0.759
Shokir's model (Shokir, 2008)	11.0	-0.728	0.151	704.177	0.818
Nemeth and Kennedy's model (Nemeth and Kennedy, 1967)	8.6	4.794	0.095	611.369	0.886
This study, overall	7.8	0.871	0.088	549.219	0.890
This study, training set	7.8	0.759	0.078	524.082	0.891
This study, testing set	8.0	1.321	0.040	640.348	0.882

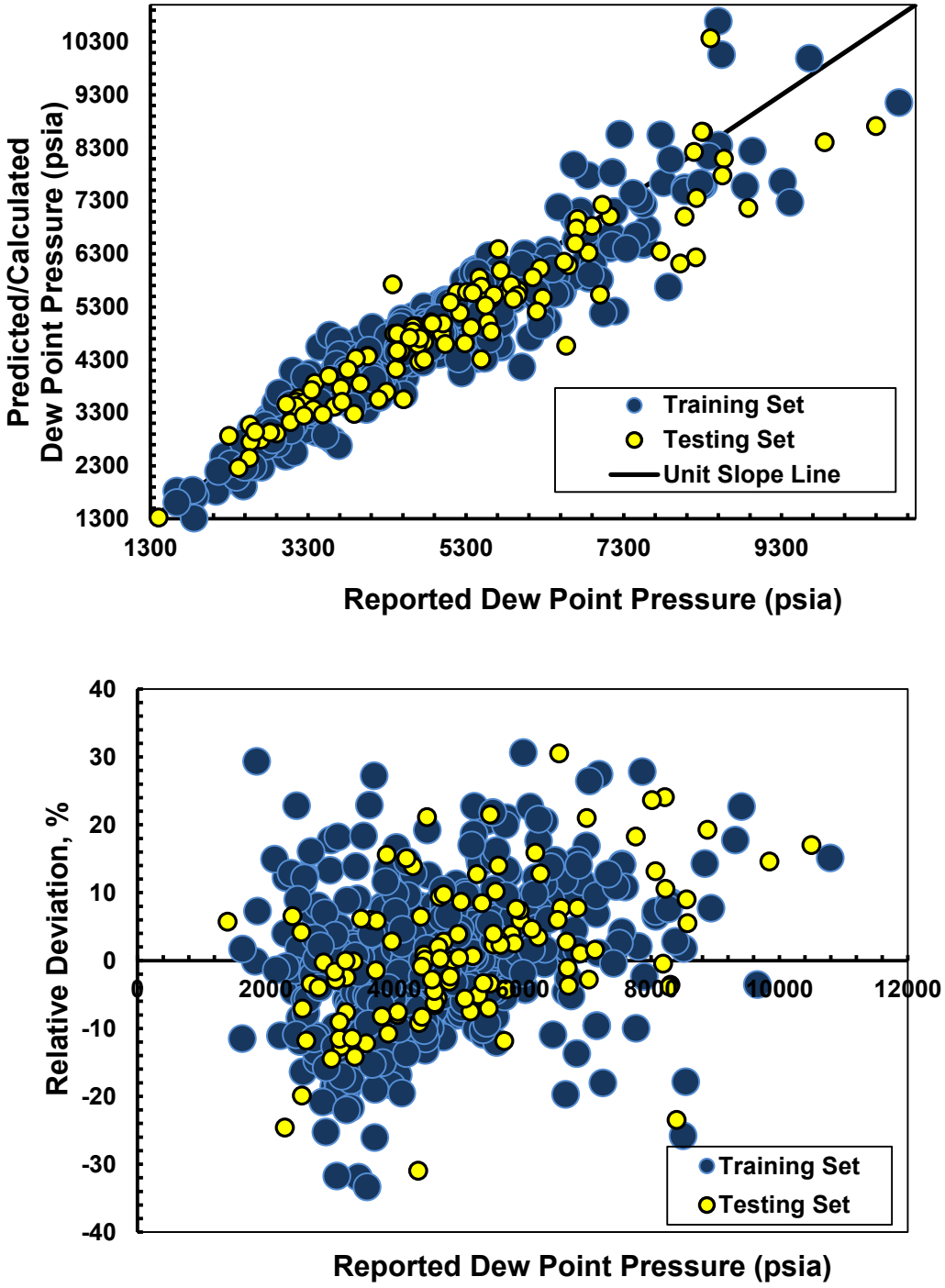

Fig. 5. 6 Comparison between the data calculated by Eq. (5.4) and actual magnitudes of dew point pressures with regard to line Y=X and residual error percentage.

A performance of Eq. (5.4) was further evaluated by comparing the results obtained with three well-known empirically derived correlations, viz. Elsharkawy (Elsharkawy, 2002a), Shokir (Shokir, 2008), and Nemeth and Kennedy (Nemeth and

Kennedy, 1967). **Table 5. 2** also provides a systematic comparison between the actual DPP data, the output values of the Elsharkawy (Elsharkawy, 2002a), Shokir (Shokir, 2008), and Nemeth and Kennedy (Nemeth and Kennedy, 1967) methods, as well as the results of the GEP-based model. Moreover, **Figs. 5. 7-9** illustrate the results obtained by Nemeth and Kennedy (Nemeth and Kennedy, 1967), Shokir (Shokir, 2008), and Elsharkawy's (Elsharkawy, 2002a) methods, respectively.

As can be seen in these figures, the agreement between the values predicted by these comparative methods and the actual data of DPP is not as good as with the GEP-based model. **Fig. 5. 10** shows the calculated AARD for the comparative methods as well as for the GEP-based model. It can be observed that the GEP-based model has better performance than the other correlations with respect to the calculated AARD values. The AARDs for the Elsharkawy (Elsharkawy, 2002a), Shokir (Shokir, 2008), and Nemeth and Kennedy (Nemeth and Kennedy, 1967) methods are 15.3, 11.0 and 8.6, respectively, while it is 7.8 % for Eq. (5.4).

Fig. 5. 11 shows a further comparison of the DPP values predicted using Eq. (5.4) with actual data, as well as the calculated DPP values using the comparative techniques, taking into account the influence of reservoir temperature on DPP. This figure also shows that the values predicted using Eq. (5.4) are closer to the actual data compared to the other methods. Furthermore, it illustrates that DPP has an incremental trend with an increase in temperature.

The comparative analysis confirms that the GEP-based model is able to calculate the desired parameter (*i.e.* the DPP in gas condensate reservoirs) with greater accuracy and consistency. The model also has a smaller number of adjustable, resulting in the model optimization and development of GEP-based models being faster, less laborious, and less costly. In other words, the GEP-based model for the estimation of DPP in gas condensate reservoirs can be considered as more simple-to-use than the comparable methods.

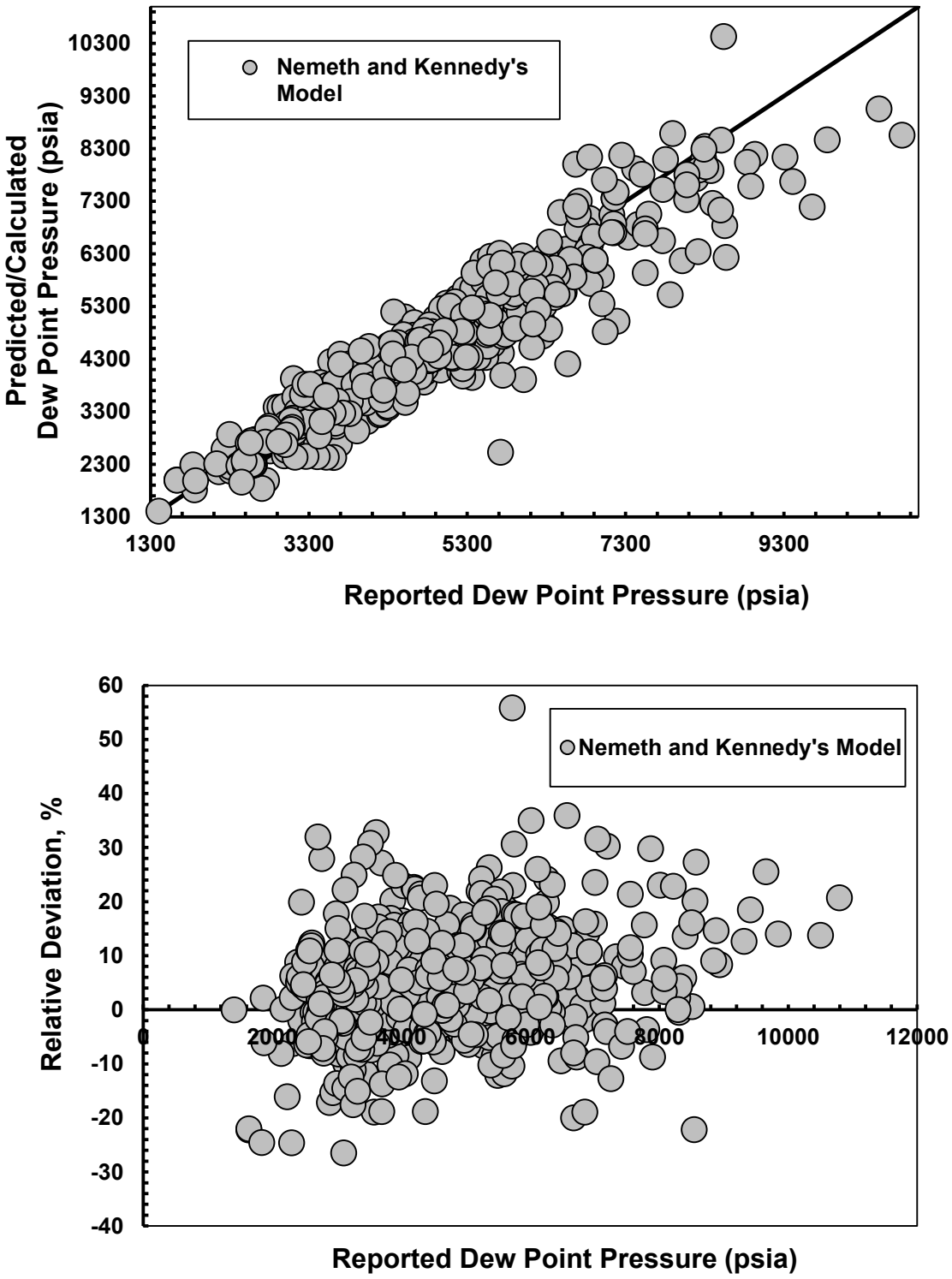

Fig. 5. 7 Comparison between the data calculated by the Nemeth and Kennedy's model (Nemeth and Kennedy, 1967) and actual magnitudes of dew point pressures with regard to line Y=X and residual error percentage.

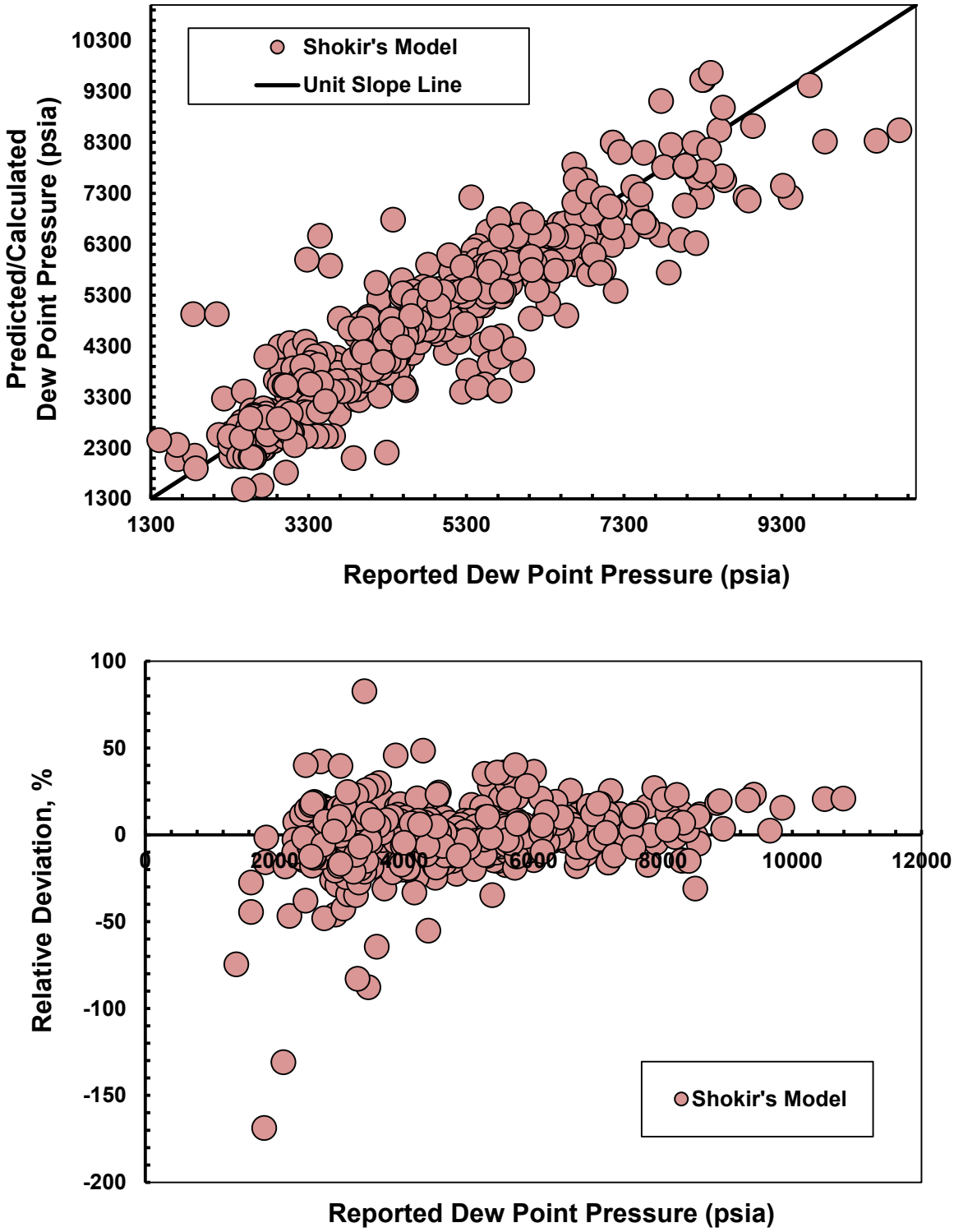

Fig. 5. 8 Comparison between the data calculated by the Shokir's model (Shokir, 2008) and actual magnitudes of dew point pressures with regard to line Y=X and residual error percentage.

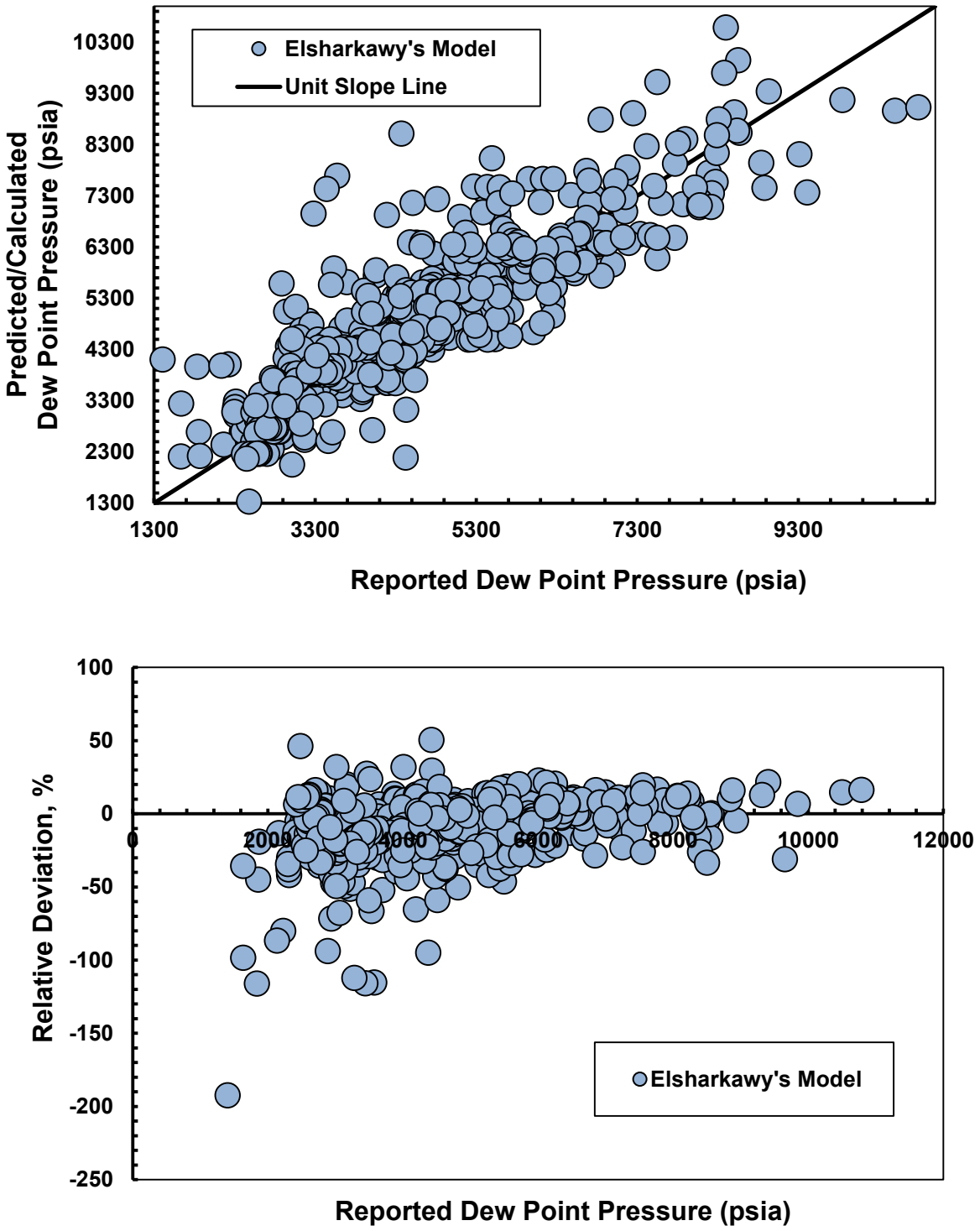

Fig. 5. 9 Comparison between the data calculated by the Elsharkawy's model (Elsharkawy, 2002a) and actual magnitudes of dew point pressures with regard to line Y=X and residual error percentage.

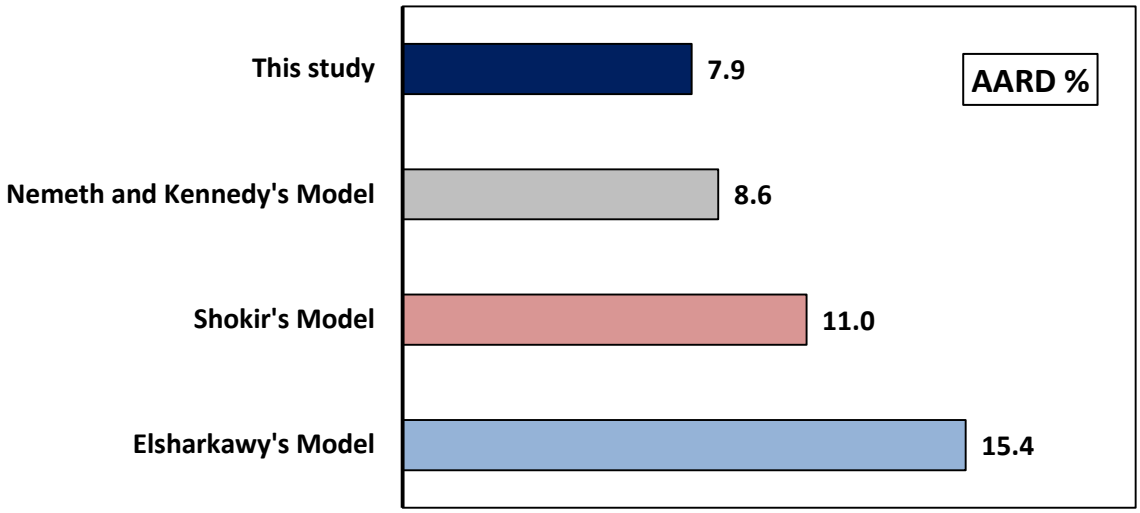

Fig. 5. 10 The performance of Eq. (5.4 and the comparative methods in predicting dew point pressures with respect to AARD.

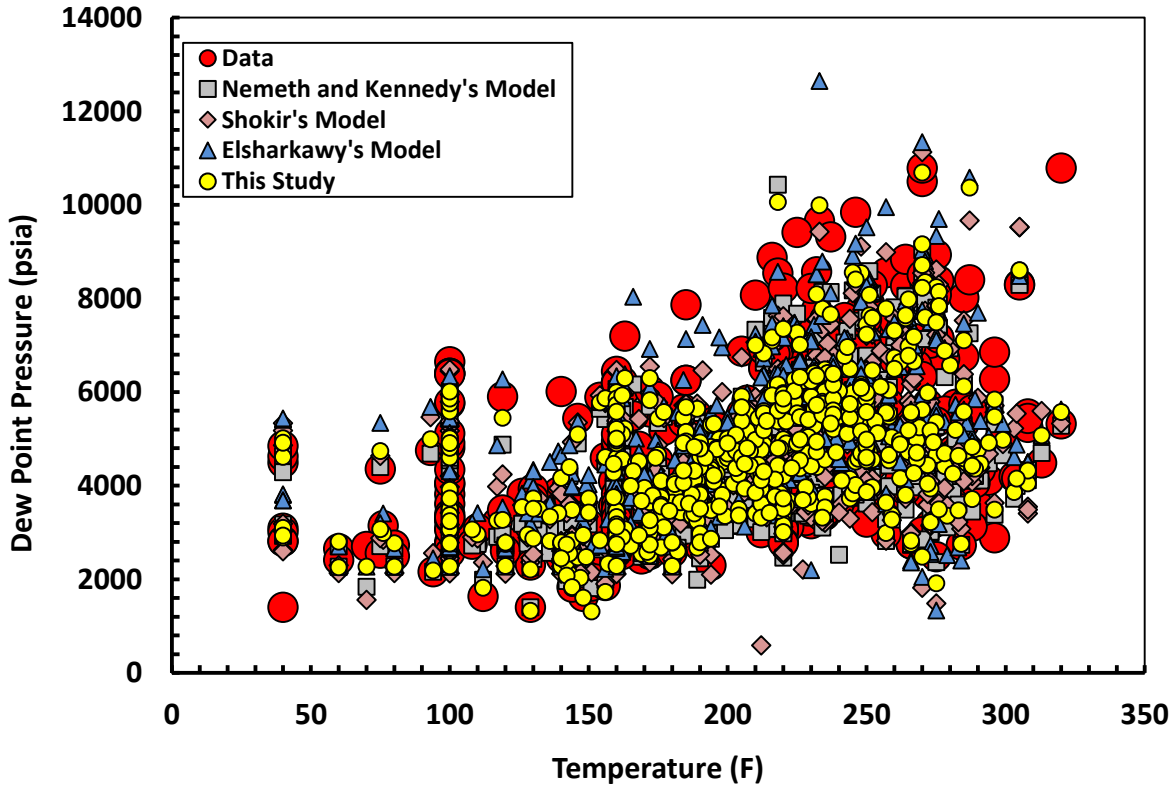

Fig. 5. 11 The performance of Eq. (5.4), and the comparative methods in predicting dew point pressures with respect to the study of the influence of temperature on dew point pressure trend.

5.2.3. Detection of the outlier DPP data points available in the databank

The Leverage scheme was used to detect outlier DPP data points existing in the assembled database. To find doubtful DPP data points, the Williams' plot is drawn by calculation of the Hat values, as illustrated in **Fig. 5. 12**. The GEP-based model proposed in this study for the prediction of DPP in gas condensate reservoirs displays a higher performance statistically, as the main portions of the data are located within the domains of 0 < H<0.07473 and −3 <R<3. It can be observed, based on the Leverage scheme, as illustrated in **Fig. 5. 12,** that there are only 13 data points (among a total of 562 data points) outside of the acceptable range of the proposed technique. Further details on outlier detection are available in the literature (Goodall, 1993; Gramatica, 2007; Mohammadi et al., 2012a).

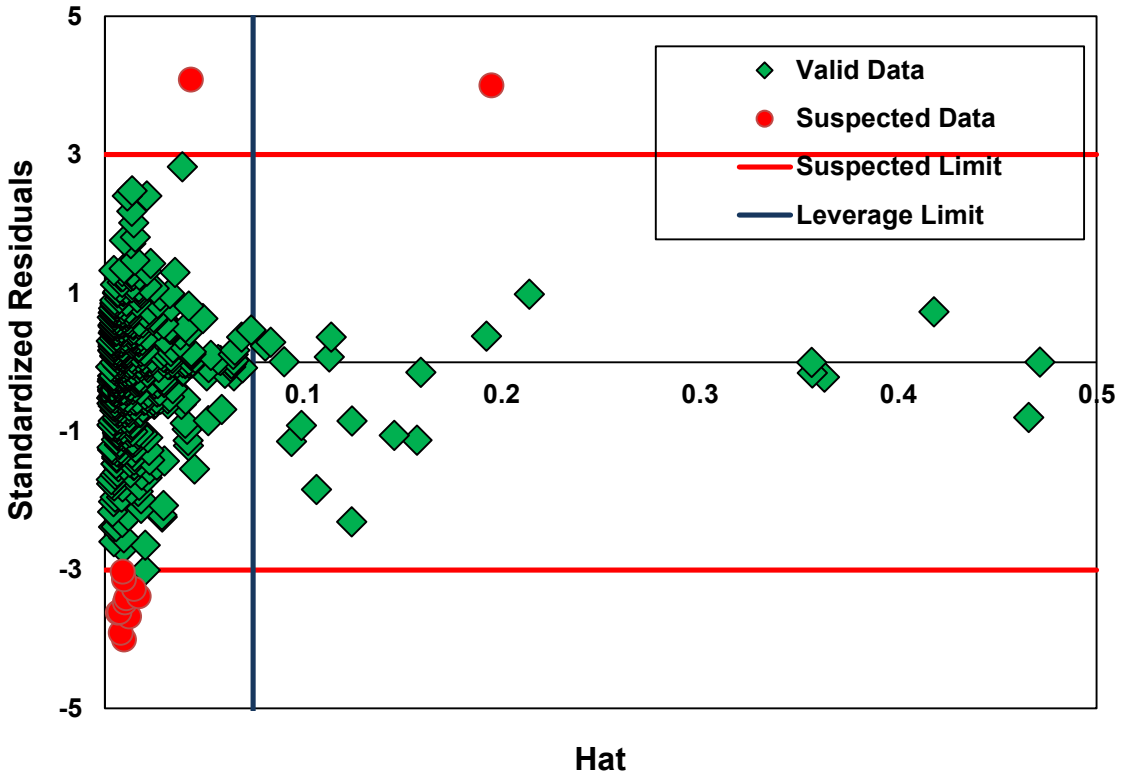

Fig. 5. 12 Determination of doubtful data points in the dew point pressure dataset, collected during development of the GEP-based scheme, using Hat values.

5.3. New Equations for Bubble Point Pressure and Oil Formation Volume Factor

5.3.1. Performance Evaluation

Previous research has proved that the performance of an empirical correlation of bubble point pressure is more complicated than for oil formation volume factor. As a result, the empirical correlations for bubble point pressure are less vigorous and precise than those for the estimation of oil formation volume factor (Arabloo et al., 2014). Hence, in order to obtain more robust, reliable and accurate empirical models for both bubble point pressure and oil formation volume factor, the GEP algorithm computational procedure has been utilized because the computational phases associated with the GEP approach (Ferreira, 2006) define the required parameters, which therefore give the most precise model for both bubble point pressure and oil formation volume factor based on the variables of solution gas oil ratio, gas gravity, oil API gravity and reservoir temperature.

In order to increase the capability and accuracy of the GEP model, compared to existing empirical correlations, neutral genes can be added to the model (this is a newly developed equation) after reaching a reasonable and appropriate number of generations required to show that both models relate to OFVF and P_b. properties. Moreover, the function called the average absolute percent relative error and correlation coefficient (R^2) was chosen to compute the overall fitness of the evolved programs.

The run of program was continued until there was no longer important improvement in the accuracy and capability of the several proposed models with the various functions utilized. The final equations for both bubble point pressure and oil formation volume factor properties are obtained as follows:

$$B_{ob} = 1 - 0.000081623\,\gamma_g\left[\sqrt{API\,T_R} + \frac{R_{Si}-4.846}{\sqrt{\gamma_g}}\right]\left(0.37658\,\gamma_g - (API - T_R)^{0.3652}\right) \qquad \text{Eq. (5.7)}$$

$$P_b = \frac{87.3067 R_{Si}\,T_R\,|\gamma_g - 2.95787| + 7639.17}{947.493\,\gamma_g + \exp(0.000641267\,API\,T_R) + API\,T_R + 3.59953\,\gamma_g\,R_{Si}} \qquad \text{Eq. (5.8)}$$

where P_b denotes bubble point pressure (psi), B_{ob} stands for OFVF at bubble point pressure (bbl/STB), T_R expresses the reservoir temperature (°F), *API* is the crude oil API

gravity, γ_g indicates the gas gravity and R_{Si} shows the solution gas oil ratio at bubble point pressure (SCF/STB).

Having developed the GEP models for the prediction of B_{oi} and P_b properties, error analyses were performed to evaluate the prediction performance of those models. These include statistical error analysis in which R^2, AAPRE, average percent relative error (APRE), root mean square (RMSE) and graphical error analysis in which a parity diagram and relative error distribution is plotted. **Table 5. 3** summarizes the statistical error parameters calculated for the B_{oi} model developed in this study.

The results obtained indicate a R^2=0.93 and an AAPRE=3.62. The error values acquired reveal that the newly developed GEP model has predicted the oil formation volume factor values with an acceptable and reliable accuracy. **Fig. 5. 13** illustrates the parity diagram and a comparison between the calculated and actual values related to oil formation volume factor data. As is clear from the figure, most of the data points are almost placed on the line of Y=X, illustrating there is agreement between the model predictions and the actual oil formation volume factor data gathered from the literature. For the purpose of illustrating the capability and performance of the GEP model in predicting oil formation volume factor, a relative error percentage distribution plot is also drawn in **Fig. 5. 14**., which reveals a small level of error on the basis of Eq. (5.7).

Additionally, the capability of the model for evaluation of the oil formation volume factor data sourced from the literature has been compared with the results of several of the most widely-utilized empirical correlations. Thirteen empirical correlations are used, viz. Arabloo et al. (Arabloo et al., 2014) model, Al-Shammasi (Al-Shammasi, 1999) model, Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994a) model, Frashad et. al (Frashad et al., 1996) model, Al-Marhoun (Al-Marhoun, 1992) model, Standing (Standing, 1947a) model, Petrosky and Farshad (Petrosky Jr and Farshad, 1998) model, Omar and Todd (Omar and Todd, 1993) model, Dindoruk and Christman (Dindoruk and Christman, 2004) model, Vazquez and Beggs (Vazquez and Beggs, 1980) model, Macary and El-Batanony (Macary and El-Batanoney, 1993) model, Abdul-Majeed (Abdul-Majeed et al., 1988) model, and Labedi (Labedi, 1990) model.

A summary of the comparative study, which was performed using statistical error parameters, is listed in **Table 5. 3**. These results confirm that Eq. (5.7) has performed

better for the calculation of oil formation volume factor compared to the studied empirical correlations. An acceptable AAPRE of the predicted values is generated from the actual B_{oi} data. A graphical comparison of AAPRE results is shown in **Fig. 5. 15**. The bar plots drawn in **Fig. 5. 15** illustrates that the model developed for the prediction of oil formation volume factor is acceptable in comparison with values calculated by other empirical correlations.

Table 5. 3 Summarized statistical error parameters including AAPRE, APRE, RMSE and R2 for the newly developed model for oil formation volume factor as well as the empirical correlation results from the actual data.

Method	**AAPRE, %**	**APRE, %**	**RMSE**	**R^2**
Present study (Eq. (5.7))	2.17	0.18	0.07	0.93
Arabloo et al. (Arabloo et al., 2014) model	2.24	−0.04	0.07	0.94
Al-Shammasi (Al-Shammasi, 1999) model	2.59	−0.92	0.07	0.93
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994a) model	2.92	−0.30	0.07	0.93
Frashad et.al (Frashad et al., 1996) model	2.94	0.39	0.07	0.93
Al-Marhoun (Al-Marhoun, 1992) model	3.09	−0.38	0.08	0.93
Standing (Standing, 1947a) model	3.36	−1.98	0.08	0.93
Petrosky and Farshad (Petrosky Jr and Farshad, 1998) model	3.46	−2.35	0.08	0.93
Omar and Todd (Omar and Todd, 1993) model	5.03	2.08	0.12	0.85
Dindoruk and Christman (Dindoruk and Christman, 2004) model	5.52	−2.94	0.14	0.83
Vazquez and Beggs (Vazquez and Beggs, 1980) model	5.59	3.01	0.13	0.82
Macary and El-Batanony (Macary and El-Batanoney, 1993) model	9.11	−8.44	0.19	0.85
Abdul-Majeed (Abdul-Majeed et al., 1988) model	27.77	−27.73	0.4	0.83
Labedi (Labedi, 1990) model	37.68	−37.64	0.68	0.93

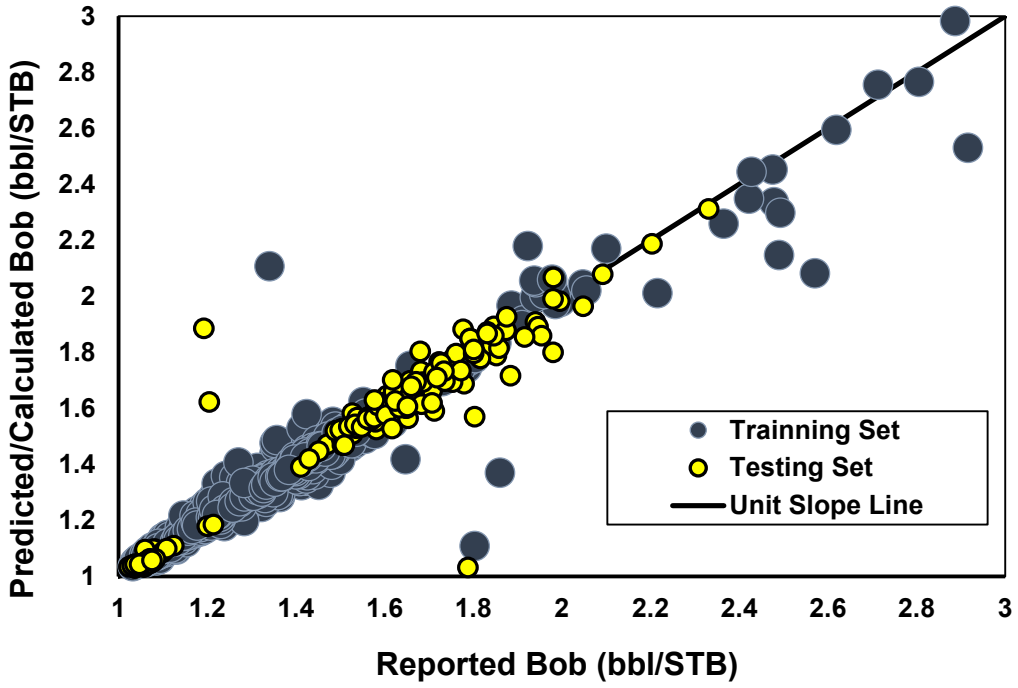

Fig. 5. 13 Parity diagram for the predicted values by the new empirical model and the literature-reported values of the oil formation volume factor.

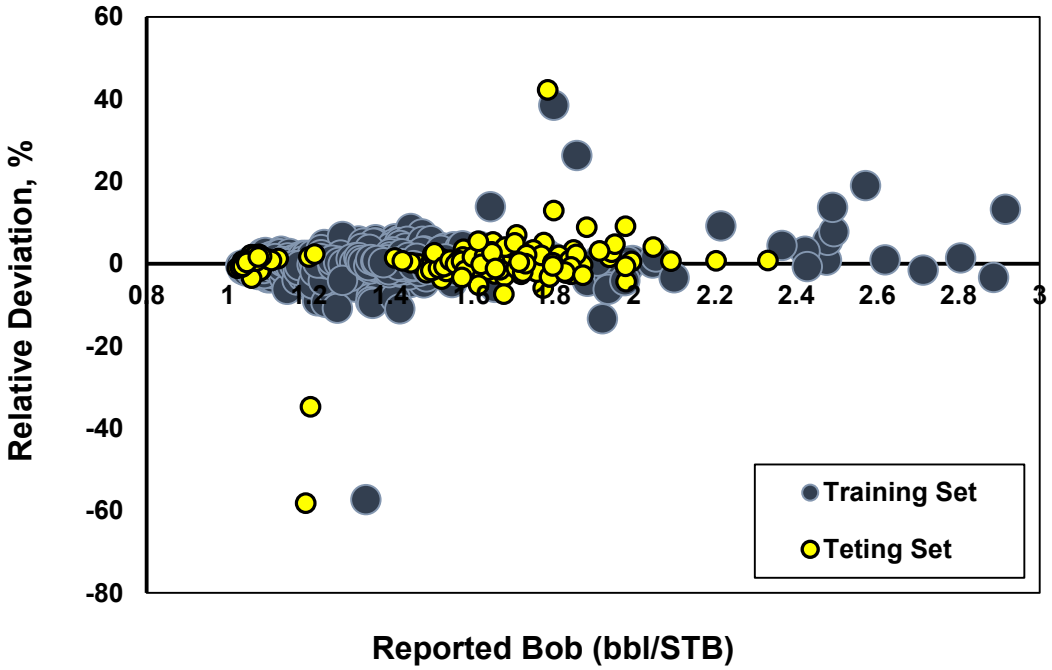

Fig. 5. 14 Relative deviations plot of the oil formation volume factor values obtained by the newly proposed model from the available dataset.

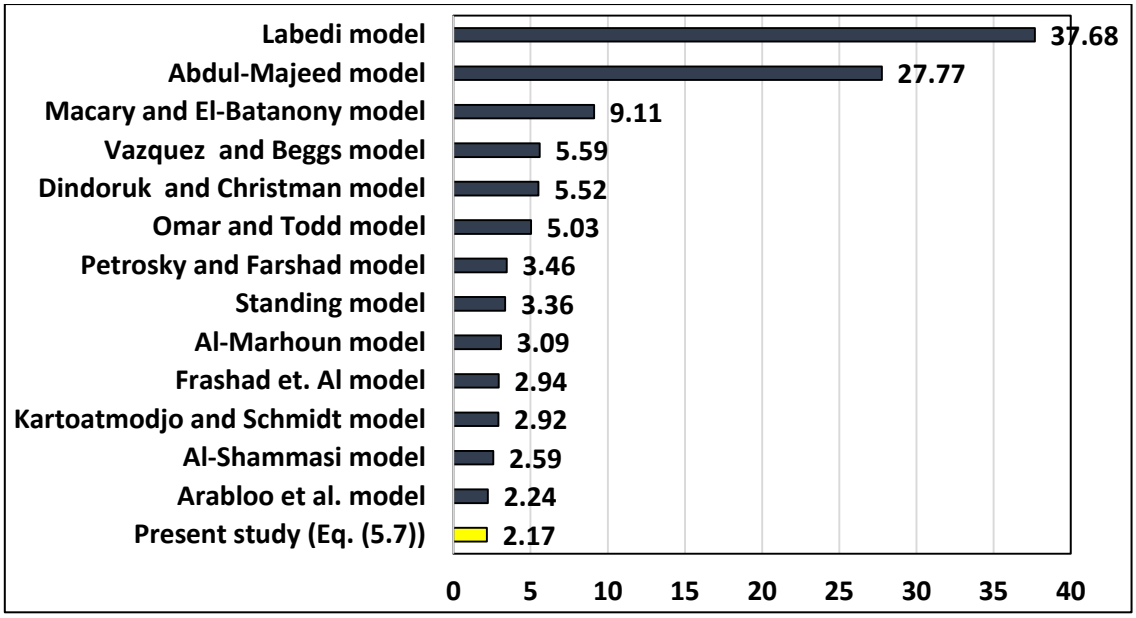

Fig. 5. 15 Comparison between the AAPRE values obtained by the new model, and the corresponding empirical correlations, for the prediction of oil formation volume factor

Table 5. 4 lists the statistical error parameters calculated for the proposed P_b model. The table reports that the values obtained for AAPRE, APRE, RMSE and R^2 are 15.3%, 2.23%, 468.11 and 0.88, respectively. These amounts of error confirm the level of accuracy of the new model for the prediction of bubble point pressure. **Figure 5. 16** provides a diagram of the parity between the calculated and reported values of bubble point pressure data, which illustrates a high level of agreement between the newly developed model predictions and the actual bubble point pressure data. **Figure 5. 17** represents the relative percentage error distribution plot for the P_b model. This figure indicates the small existing error range and the low scatter around the zero error line, for the bubble point pressure data and APRE obtained for Eq. (5.8). Furthermore, several widely-utilized empirical correlations available in the literature are provided relating to bubble point pressure property, including Arabloo et al. (Arabloo et al., 2014) model, Al-Shammasi (Al-Shammasi, 1999) model, Lasater (Lasater, 1958) model, Dindoruk and Christman (Dindoruk and Christman, 2004) model, Valko and Mcain (Valko and McCain Jr, 2003) model, Frashad et al. (Frashad et al., 1996) model, Velarde et al. (Velarde et al., 1997) model, Al-Marhoun (Al-Marhoun, 1988) model, Standing (Standing, 1947a) model, Vazquez and Beggs (Vazquez and Beggs, 1980) model, Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994a) model, Macary and El-Batanony (Macary and El-

Batanoney, 1993) model, Petrosky and Farshad (Petrosky Jr and Farshad, 1998) model, Yi (Yi, 2000) model, Omar and Todd (Omar and Todd, 1993) model, and the Ikiensikimama and Ogboja (Ikiensikimama and Ogboja, 2009) model. **Table 5. 4** summarizes statistical error parameters calculated for the aforementioned correlations as well as for the P_b model developed in this study. As can be seen in the table, the P_b model has better performance in comparison with the studied methods. To highlight relative performance, the AAPRE obtained for all methods is shown in **Figure 5. 18**.

Figure 5. 18 confirms that Eq. (5.8) is more capable and accurate than the other correlations for the prediction of bubble point pressure. In view of the above results and discussion, it can be confirmed that the models proposed for the evaluation of reservoir oil PVT properties, including oil formation volume factor and bubble point pressure, perform in terms of universality, reliability and accuracy, and they can be implemented in various applications in petroleum engineering, such as in the development of new software.

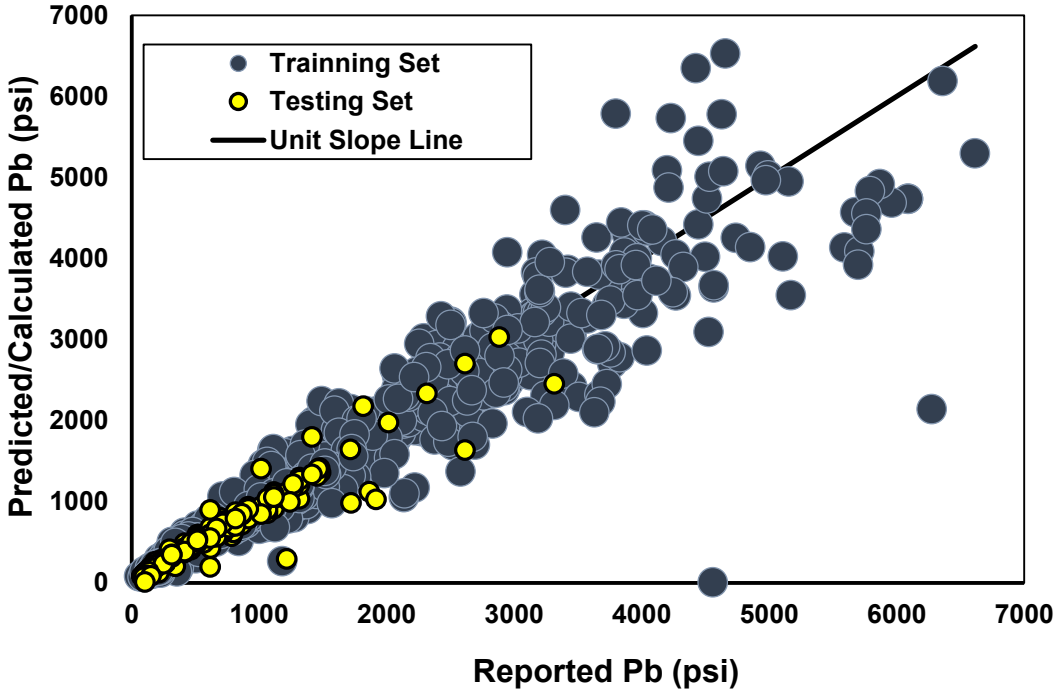

Fig. 5. 16 Parity diagram for the predicted values by the new empirical model and the literature-reported values of bubble point pressure.

Table 5. 4 Summarized statistical error parameters, including AAPRE, APRE, RMSE and R2, for the new model and the empirical correlation results from the actual data, for bubble point pressure.

Method	**AAPRE, %**	**APRE, %**	**RMSE**	**R^2**
Present study (Eq. (5.8))	15.3	2.23	468.11	0.88
Arabloo et al. (Arabloo et al., 2014) model	18.9	3.2	501.7	0.86
Al-Shammasi (Al-Shammasi, 1999) model	20.8	−7.6	478.7	0.87
Lasater (Lasater, 1958) model	25.5	−8.6	481.5	0.87
Dindoruk and Christman (Dindoruk and Christman, 2004) model	25.6	−2.8	510.8	0.86
Valko and Mcain (Valko and McCain Jr, 2003) model	25.7	0.1	584.2	0.82
Frashad et al. (Frashad et al., 1996) model	25.9	−8.7	507.4	0.85
Velarde et al. (Velarde et al., 1997) model	26.9	−2.1	596.6	0.82
Al-Marhoun (Al-Marhoun, 1988) model	27.9	−4.5	550.8	0.84
Standing (Standing, 1947a) model	28.7	−16.4	588.4	0.85
Vazquez and Beggs (Vazquez and Beggs, 1980) model	32.3	−24.7	693.9	0.87
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994a) model	35.6	−27.2	819.7	0.84
Macary and El-Batanony (Macary and El-Batanoney, 1993) model	52.9	−38.3	596.4	0.85
Petrosky and Farshad (Petrosky Jr and Farshad, 1998) model	90.7	58.7	840	0.85
Yi (Yi, 2000) model	94	94	2115.2	0.77
Omar and Todd (Omar and Todd, 1993) model	361.5	−356.0	11387.4	0.03
Ikiensikimama and Ogboja (Ikiensikimama and Ogboja, 2009) model	555.5	−555.5	5175.9	0.40

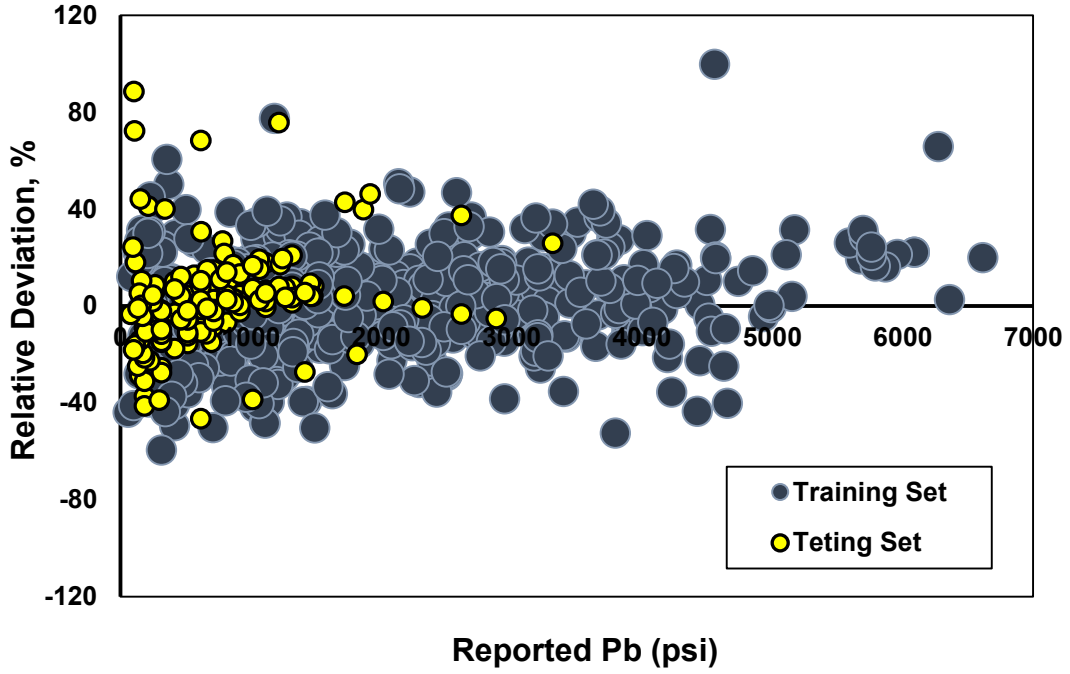

Fig. 5. 17 Relative deviation plot of the bubble point pressure values obtained by the newly proposed model from the available dataset.

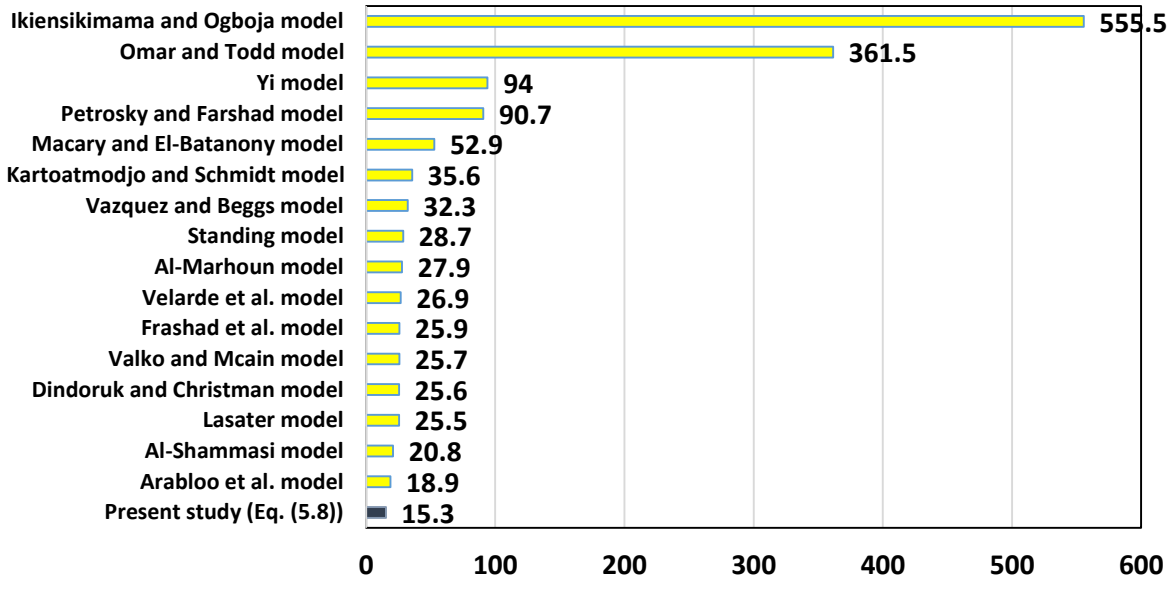

Fig. 5. 18 Comparison of the AAPRE values obtained by the new model, and the empirical correlations, for the prediction of bubble point pressure.

5.3.2. Detection of Outlier Data Points Existing in the Dataset

The accuracy performance and capability of any predictive model will be improved by means of the detection of the outlier data point(s) existing in the datasets because the results will probably be sensitive to such data points. The outlier data points are the individual datum (or groups of data) that may differ from the bulk of the data existing in a databank (Gramatica, 2007; Mohammadi et al., 2012a; Mohammadi et al., 2012c; Rousseeuw and Leroy, 2005). Hence, evaluating the datasets related to both oil formation volume factor and bubble point pressure is a requisite, since uncertainties affect the accuracy performance and capability of predictive models developed.

The Leverage value statistical technique is applied in this study for the detection of outlier data points existing in the datasets associated with both oil formation volume factor and bubble point pressure properties (Goodall, 1993; Gramatica, 2007). The detection of the suspended data or outliers is undertaken by drawing the Williams plot based on the H values calculated (Mohammadi et al., 2012a; Mohammadi et al., 2012c). For more information about the Leverage approach, a detailed definition related to the computational procedure and also equations of this technique can be obtained elsewhere (Mohammadi et al., 2012a; Mohammadi et al., 2012c).

Figs. 5. 19 and **5. 20** illustrate the Williams plots for the predicted values of oil formation volume factor and bubble point pressure using the newly developed GEP models. As can be seen in these figures, the existence of the majority of data points, in the ranges $0 \leq H \leq 0.01984$ and $-3 \leq R \leq 3$ for both oil formation volume factor and bubble point pressure models, confirms that the applied models are statistically valid and correct. As a matter of fact, **Figs. 5. 19** and **5. 20** show that there are 12 data points for the oil formation volume factor model and 13 data points for the bubble point pressure model, that are outside of the applicability domain of the GEP models and are accounted as outliers whose values may be doubtful, compared with the corresponding actual data.

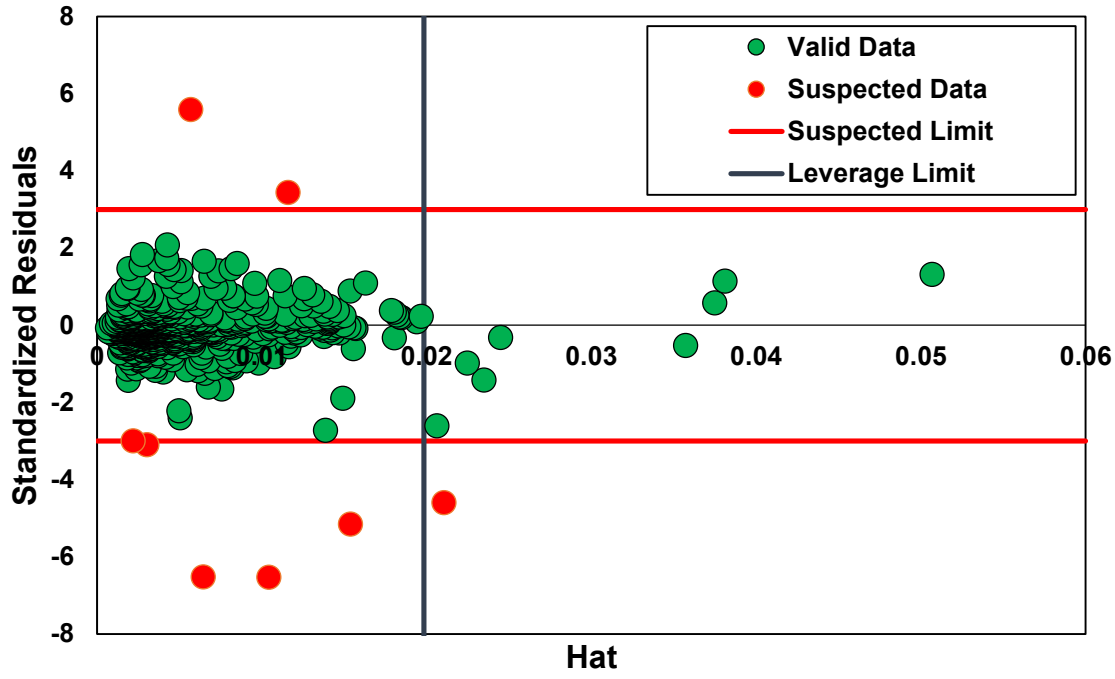

Fig. 5. 19 Detection of the probable outlier and doubtful data of oil formation volume factor and the applicability domain of the proposed GEP model.

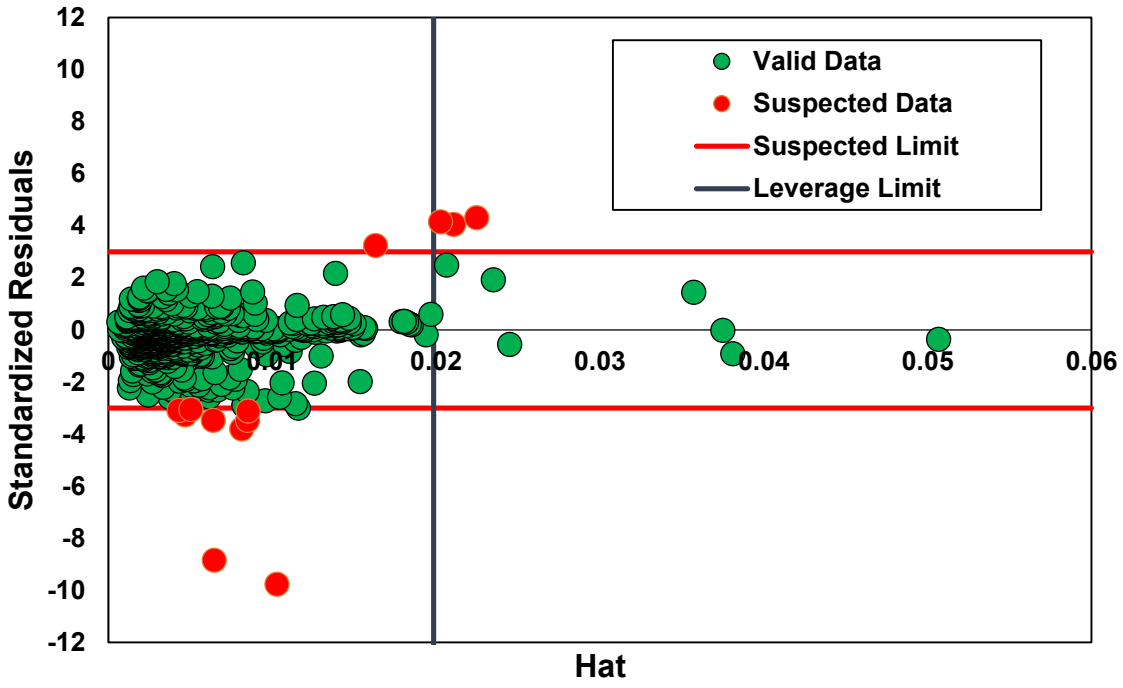

Fig. 5. 20 Detection of the probable outlier and doubtful data of bubble point pressure and the applicability domain of the proposed GEP model.

5.4. New Equations for Reservoir Oil Viscosities

5.4.1. Dead Oil Viscosity

In the GEP-based symbolic regression approach, the most effective input variables on the target parameter (dead oil viscosity) are selected automatically from the set of all of the independent predictors. But in this study the algorithm was fixed on two variables including oil gravity and reservoir temperature. Certain previously published models (Al-Khafaji et al., 1987; Alomair et al., 2011; Beal, 1946; Beggs and Robinson, 1975; Bennison, 1998; Egbogah and Ng, 1990; Elsharkawy and Alikhan, 1999; Glaso, 1980; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Kaye, 1985; Labedi, 1992; Naseri et al., 2005; Petrosky, 1990), used these two parameters as effective variables for the determination of dead oil viscosity. Consequently, for the development of the GEP algorithm, the best GEP-based model using the independent predictors (reservoir temperature and oil gravity) obtained after the GEP evolution process, is given as follows:

$$\mu_{od} = \frac{614.82\ API \times T - 63529.0\ T + 2.0359 \times 10^{7}}{T \times API^{3} - 482088} \qquad (5.9)$$

where μ_{od} denotes the dead oil viscosity (cp), T represents reservoir temperature (°F) and finally API stands for dead oil gravity.

Table 5. 5 (top section) summarizes the statistical error factors, including the average absolute relative deviation, root mean square error and R-squared error, for the calculation of dead oil viscosities. The table reports reasonable values for the abovementioned error parameters (*e.g.,* AARD=17.29%, RMSE=1.82, and R^2=0.97).

To illustrate the accuracy of the performance of the new model, a comparison between the values of dead oil viscosity using Eq. (5.9) and the values in the applied dataset, is demonstrated by means of a parity diagram or crossplot and a relative error distribution (**Fig. 5. 21**). Furthermore, to make visual the accuracy and capability of the model for dead oil viscosity, a broad comparative study was conducted applying the most reliable empirically derived correlations available (Al-Khafaji et al., 1987; Alomair et al., 2011; Beal, 1946; Beggs and Robinson, 1975; Bennison, 1998; Egbogah and Ng, 1990; Elsharkawy and Alikhan, 1999; Glaso, 1980; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Kaye, 1985; Labedi, 1992; Naseri et al., 2005; Petrosky, 1990).

Table 5. 5 Summarized statistical error factors including AARD, RMSE and R^2 for the developed models associated with dead, saturated and under-saturated oil viscosities.

Developed model	AARD [a] %	RMSE [d]	R^2 [e]
Dead oil viscosity model			
Training	17.40	1.95	0.974
Testing	16.86	1.16	0.900
Overall	17.29	1.82	0.973
Saturated oil viscosity model			
Training	14.06	0.53	0.966
Testing	11.53	0.26	0.961
Overall	13.55	0.49	0.966
Under-saturated oil viscosity model			
Training	1.51	0.10	0.999
Testing	1.29	0.06	0.998
Overall	1.47	0.09	0.999

Table 5. 6 reveals the results of the comparative study on the estimation of dead oil viscosities. As is clear from the table, in relation to all of the error factors investigated, the model for dead oil viscosities is more accurate than the empirically derived correlations. Additionally, Naseri et al. (Naseri et al., 2005) is shown to have developed the second best model for the determination of dead oil viscosity of Iranian crudes (AARD equal to 27.5 % as shown in **Fig. 5. 22**.

To represent the smoothness of the developd model for dead oil viscosity, a trend analysis is performed illustrating the changes of viscosity versus input variables. To this end, a 3D plot of changes of dead oil viscosity, oil API gravity and reservoir temperatures is presented. **Fig. 5. 23** shows that dead oil viscosity decreases with an increase in temperature and oil API gravity. This figure confirms that the model has captured the reported general trend (Ahmed, 2006) of dead oil viscosity versus oil API gravity and temperature.

Table 5. 6 Comparison of summarized statistical error factors including AARD, RMSE and R^2 for dead oil viscosity and empirical correlations to the actual data.

Method	AARD %	R^2	RMSE
Beal (Beal, 1946)	891.2	0.1088	83.14
Beggs and Robinson (Beggs and Robinson, 1975)	216.7	0.0376	245.05
Glaso (Glaso, 1980)	33.4	0.9270	3.84
Kaye (Kaye, 1985)	52.0	0.2268	10.19
Al-Khafaji (Al-Khafaji et al., 1987)	29.9	0.7283	6.04
Petrosky (Petrosky, 1990)	41.6	0.8695	4.18
Egbogah and Ng (Egbogah and Ng, 1990)	55.6	0.9208	3.26
Labedi (Labedi, 1992)	177.9	0.3910	14.95
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994b)	36.8	0.9065	4.50
Bennison (Bennison, 1998)	70.9	0.6689	12.25
Elsharkawy and Alikhan (Elsharkawy and Alikhan, 1999)	72.9	0.9065	13.25
Hossain *et al*. (Hossain et al., 2005)	68.9	0.6000	16.24
Naseri *et al*. (Naseri et al., 2005)	27.5	0.8233	3.88
Alomair *et al*. (Alomair et al., 2011)	72.4	0.8275	6.36
This study	17.2	0.9730	1.82

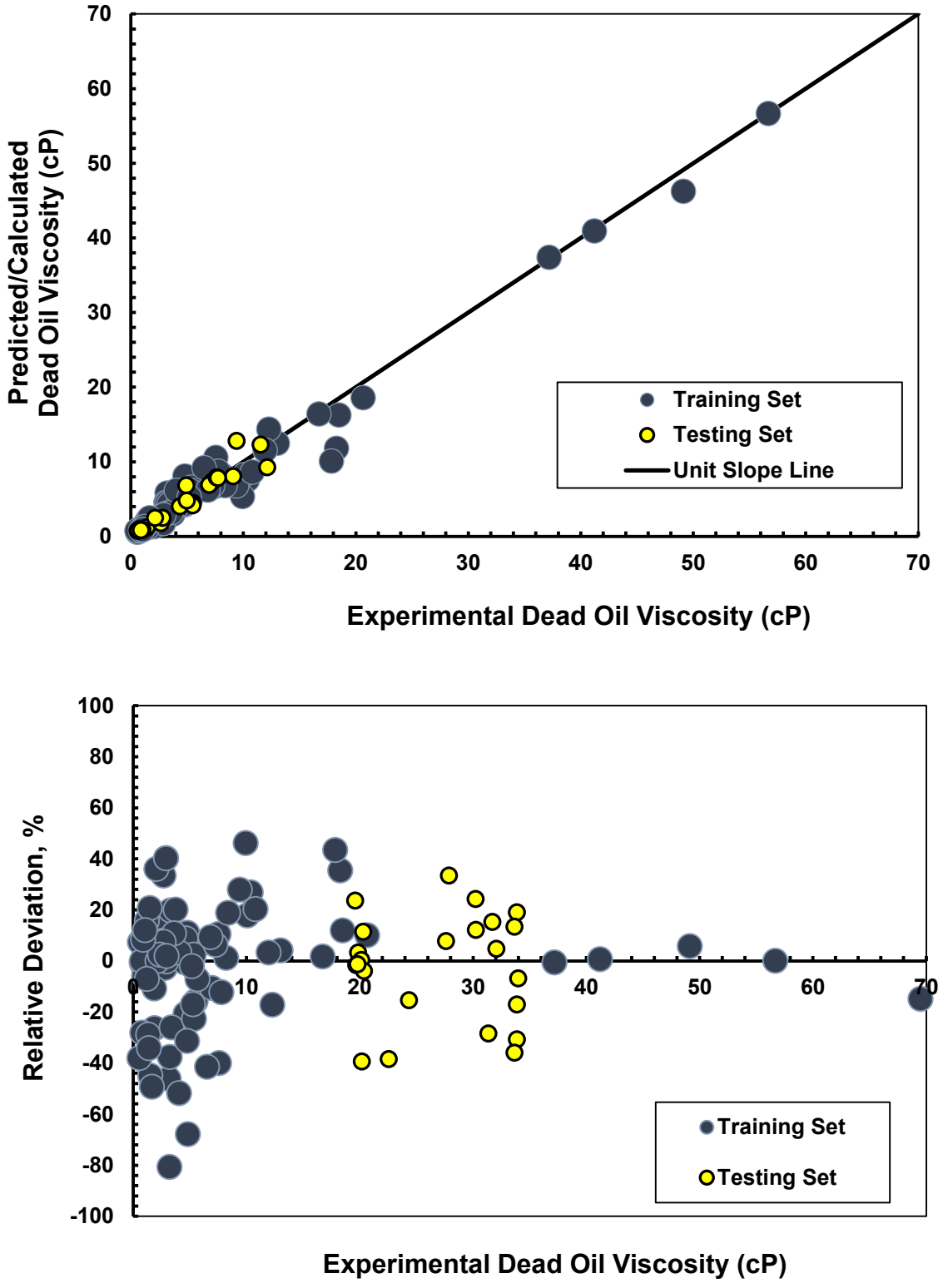

Fig. 5. 21 Parity diagram and relative deviation distribution plot for the developed dead oil viscosity model.

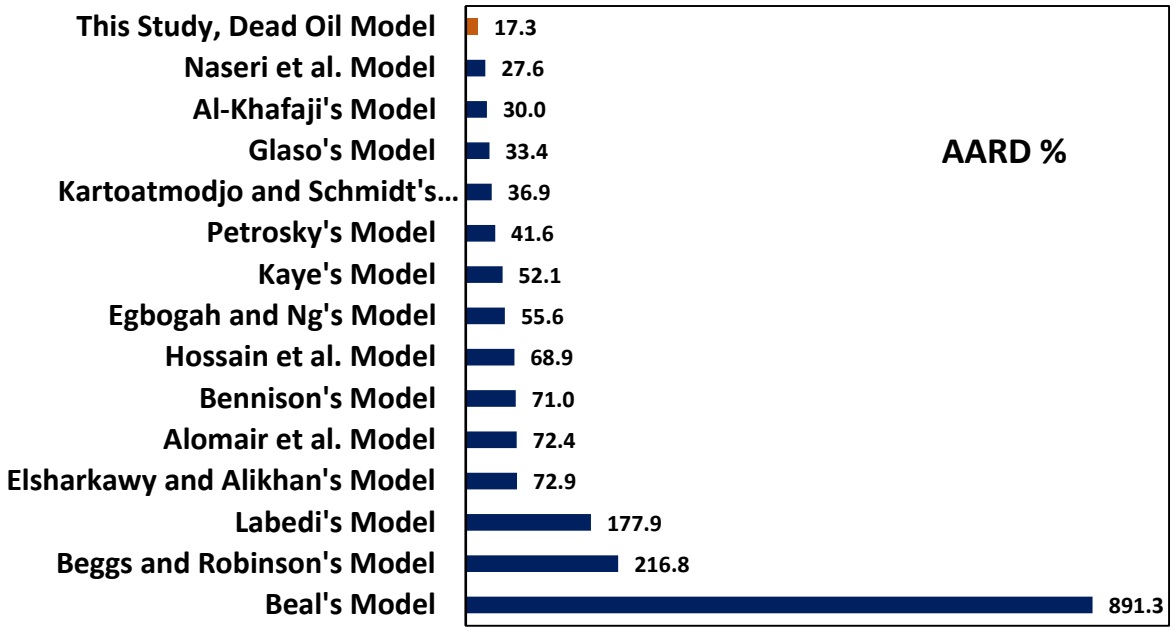

Fig. 5. 22 Comparison between the AARD % values for the determination of dead oil viscosity obtained by the model developed in this work and comparative methods.

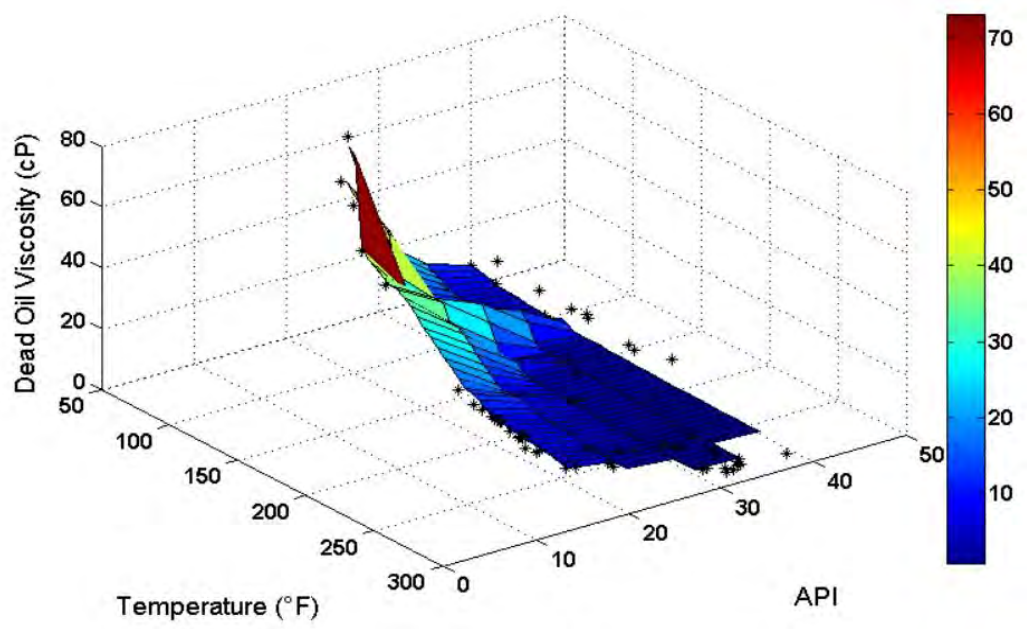

Fig. 5. 23 3D plot of change of dead oil viscosity versus change in temperature and oil API gravity.

5.4.2. Saturated Oil Viscosity

In order to develop a reliable model for the prediction of the saturated oil viscosity of Iranian crudes using the GEP methodology, pressure at bubble point and dead oil

viscosity are considered as the most effective input variables or predictors, recommended by published researchers (Al-Khafaji et al., 1987; Beggs and Robinson, 1975; Bergman and Sutton, 2007; Chew and Connally Jr, 1959; Elsharkawy and Alikhan, 1999; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Khan et al., 1987; Labedi, 1992; Naseri et al., 2005; Petrosky, 1990). As a result, the run of the encoded GEP algorithm was continued until there were no longer important improvements in the efficiency, simplicity and accuracy of the equations obtained. Consequently, the final model for the determination of saturated oil viscosity is obtained as follows:

$$\mu_{ob} = \frac{3.5752\,\mu_{od} + 1.9812}{145.0377P_b - 0.1379\,\mu_{od} + 9.4204} + 0.71308^{\mu_{od}{}^{7.3}} + 0.083617 \qquad (5.10)$$

where μ_{ob} is saturated oil viscosity (cp), μ_{od} expresses the dead oil viscosity (cp), P_b stands for bubble point pressure (psi).

The middle section of **Table 5. 5** lists the statistical error parameters calculated for Eq. (5.10). The equation proposed in this study for the determination of saturated oil viscosity has an acceptable accuracy with an AARD %, RMSE and R^2 equal to 13.55, 0.49 and 0.96, respectively. **Fig. 5. 24** illustrates the analysis of the predicted saturated oil viscosity versus the actual data points. The top view is a crossplot and the bottom view demonstrates the relative error distribution plot of saturated oil viscosity. As can be seen, there is good agreement between the predicted and actual saturated oil viscosity data. Moreover, the results obtained by using the model are compared with well-known empirical correlations available in open literature (Al-Khafaji et al., 1987; Beggs and Robinson, 1975; Bergman and Sutton, 2007; Chew and Connally Jr, 1959; Elsharkawy and Alikhan, 1999; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Khan et al., 1987; Labedi, 1992; Naseri et al., 2005; Petrosky, 1990).

Table 5. 7 Comparison of summarized statistical error factors for saturated oil viscosity, including AARD, RMSE and $R^{2,}$ for the developed model and the empirical correlation results from the data obtained.

Method	AARD %	R^2	RMSE
Chew and Connally I (Chew and Connally Jr, 1959)	29.6	0.8238	1.18
Chew and Connally II (Chew and Connally Jr, 1959)	39.6	0.7999	1.46
Chew and Connally III (Chew and Connally Jr, 1959)	Out of Range	0	Out of Range
Beggs and Robinson (Beggs and Robinson, 1975)	31.9	0.5320	1.35
Al-Khafaji (Al-Khafaji et al., 1987)	20.7	0.7992	1.32
Khan *et al*. (Khan et al., 1987)	Out of Range	0	Out of Range
Petrosky (Petrosky, 1990)	27.3	0.7399	1.14
Labedi (Labedi, 1992)	248.6	0.4283	3.70
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994b)	25.7	0.7895	1.11
Elsharkawy and Alikhan (Elsharkawy and Alikhan, 1999)	24.4	0.6576	1.20
Hossain *et al*. (Hossain et al., 2005)	Out of Range	0	Out of Range
Naseri *et al*. (Naseri et al., 2005)	52.3	0.5672	1.38
Bergman and Sutton (Bergman and Sutton, 2007)	26.7	0.7339	1.15
This study	13.5	0.9660	0.49

Table 5. 7 summarizes the calculated AARD %, RMSE and R-squared for the model developed in this study along with the comparative methods studied. The table clearly points out that Eq. (5.10) has the highest accuracy with regard to all of the error factors investigated. Additionally, **Fig. 5. 25** shows the AARD % analysis of all studied models, which confirms a low deviation result using Eq. (5.10) in comparison with the other methods. As a result, an increase to the pressure at bubble point leads to a reduction of saturated oil viscosities (Hemmati-Sarapardeh et al., 2014b). **Fig. 5. 26** confirms this change of saturated reservoir oil viscosities.

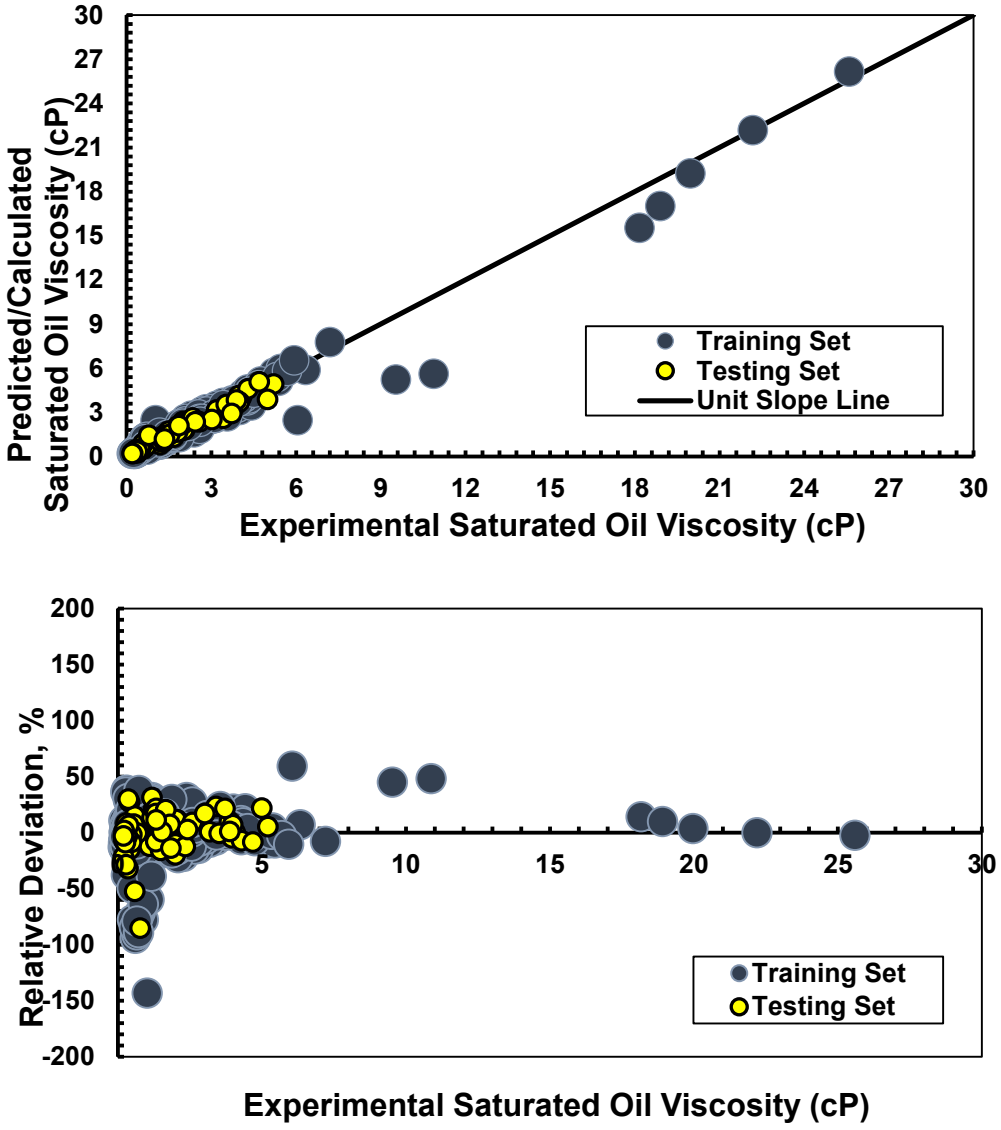

Fig. 5. 24 Parity diagram and relative deviation distribution plot for the developed saturated oil viscosity model.

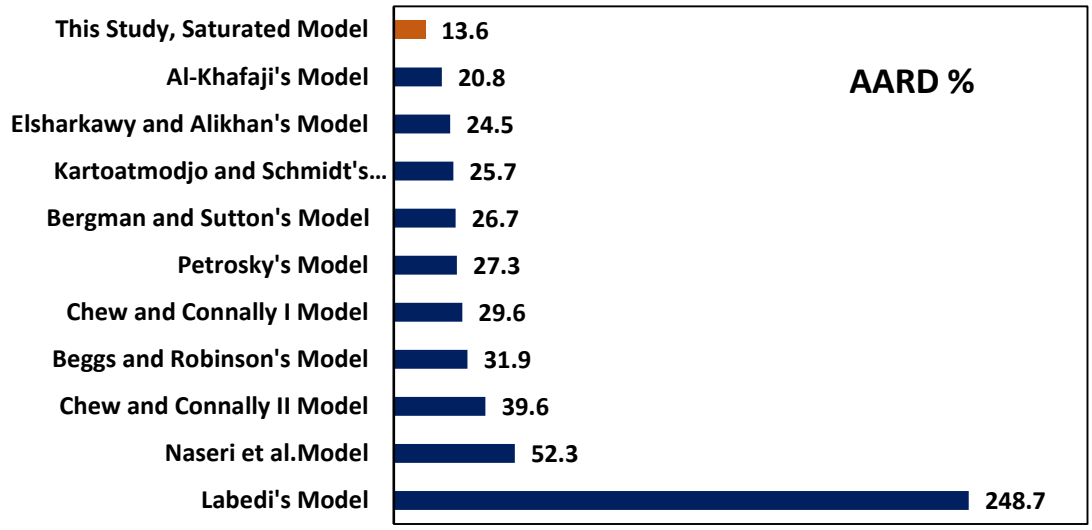

Fig. 5. 25 Comparison between the AARD % values obtained for the determination of saturated oil viscosity and those obtained in the comparative methods. The AARDs obtained by means of the Chew and Connally III (Chew and Connally Jr, 1959), Khan et al. (Khan et al., 1987), and Hossain et al. (Hossain et al., 2005) methods are out of range.

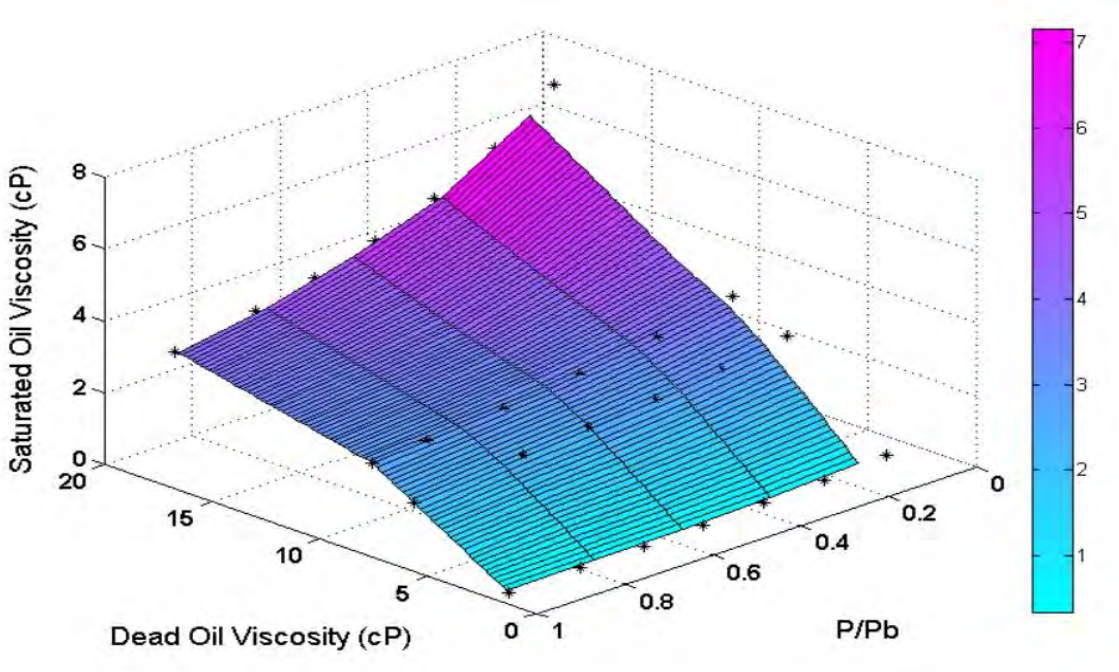

Fig. 5. 26 3D plot of change of saturated oil viscosity versus change in the pressure at bubble point and dead oil viscosity.

5.4.3. Under-Saturated Oil Viscosity

Previously published research (Beal, 1946; Elsharkawy and Alikhan, 1999; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Khan et al., 1987; Labedi, 1992; Orbey and Sandler, 1993; Petrosky, 1990; Vazquez and Beggs, 1980) recommend the pressure, saturated oil viscosity and bubble point pressure as the most effective reservoir parameters for the estimation of under-saturated oil viscosity. Therefore, these variables are selected as required parameters to develop a reliable model using the GEP approach. The final model for the determination of under-saturated reservoir oil viscosity using the GEP algorithm is obtained as follows:

$$\mu_o = \frac{0.01115\,P}{P_b} + \frac{1.1989\times10^{-8}\ (P\times\mu_{ob})^{\,2} + 7.9372\times10^{-4}\times P\times\mu_{ob} + 10.926\,\mu_{ob}}{0.001\,P_b + 10.712} \tag{5.11}$$

where μ_o represent reservoir oil viscosity (cp), P denotes pressure (psi), P_b stands for pressure at bubble point (psi) and finally μ_{ob} is saturated oil viscosity (cp).

Table 5. 5 (bottom section) shows the statistical error factors calculated for the developed model applied to under-saturated oil viscosities. The table shows that the obtained AARD %, RMSE, and R^2 are 1.47, 0.09 and 0.99, respectively. From these values of deviation it can be concluded that the model is reliable for the calculation of under-saturated reservoir oil viscosity. **Fig. 5. 27** provides crossplot and relative deviation distribution plots for the predicted under-saturated oil viscosities against the obtained data. As can be seen in the figure, the data points are almost on the line of Y=X, illustrating that there is agreement between the GEP-based model calculations and the actual data on under-saturated reservoir oil viscosities. Moreover, the low distribution of data points is observed in the bottom section of **Fig. 5. 27**.

Additionally, comparison of the values estimated by Eq. (5.11) with the previously published models (Beal, 1946; Elsharkawy and Alikhan, 1999; Hossain et al., 2005; Kartoatmodjo and Schmidt, 1994b; Khan et al., 1987; Labedi, 1992; Orbey and Sandler, 1993; Petrosky, 1990; Vazquez and Beggs, 1980) for the determination of under-saturated oil viscosities, clearly shows the accuracy of the model compared to other comparative methods (**Table 5. 8** and **Fig. 5. 28**).

Finally, **Fig. 5. 29** demonstrates the behaviour of the GEP-based model with changes of under-saturated oil viscosity against decreasing and/or increasing saturated

oil viscosity and pressure, ultimately to a pressure at bubble point. As can be seen in the figure, there is a smoothness for the model developed in predicting under-saturated oil viscosity in terms of trend analysis.

Table 5. 8 Comparison of statistical error factors for under-saturated oil viscosity, including AARD, RMSE and R2, of the new model and the empirical correlation results from data.

Method	**AARD %**	**R^2**	**RMSE**
Beal (Beal, 1946)	1.8	0.9978	0.133
Vazquez and Beggs (Vazquez and Beggs, 1980)	5.2	0.9713	0.575
Khan *et al*. (Khan et al., 1987)	3.2	0.9881	0.298
Petrosky (Petrosky, 1990)	6.7	0.8301	0.940
Labedi (Labedi, 1992)	1078.4	0.0969	43.557
Orbey and Sandler (Orbey and Sandler, 1993)	1.9	0.9624	0.495
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994b)	3.6	0.9986	0.107
Elsharkawy and Alikhan (Elsharkawy and Alikhan, 1999)	1.8	0.9372	0.618
Hossain *et al*. (Hossain et al., 2005)	4.7	0.9948	0.225
This study	1.4	0.9990	0.09

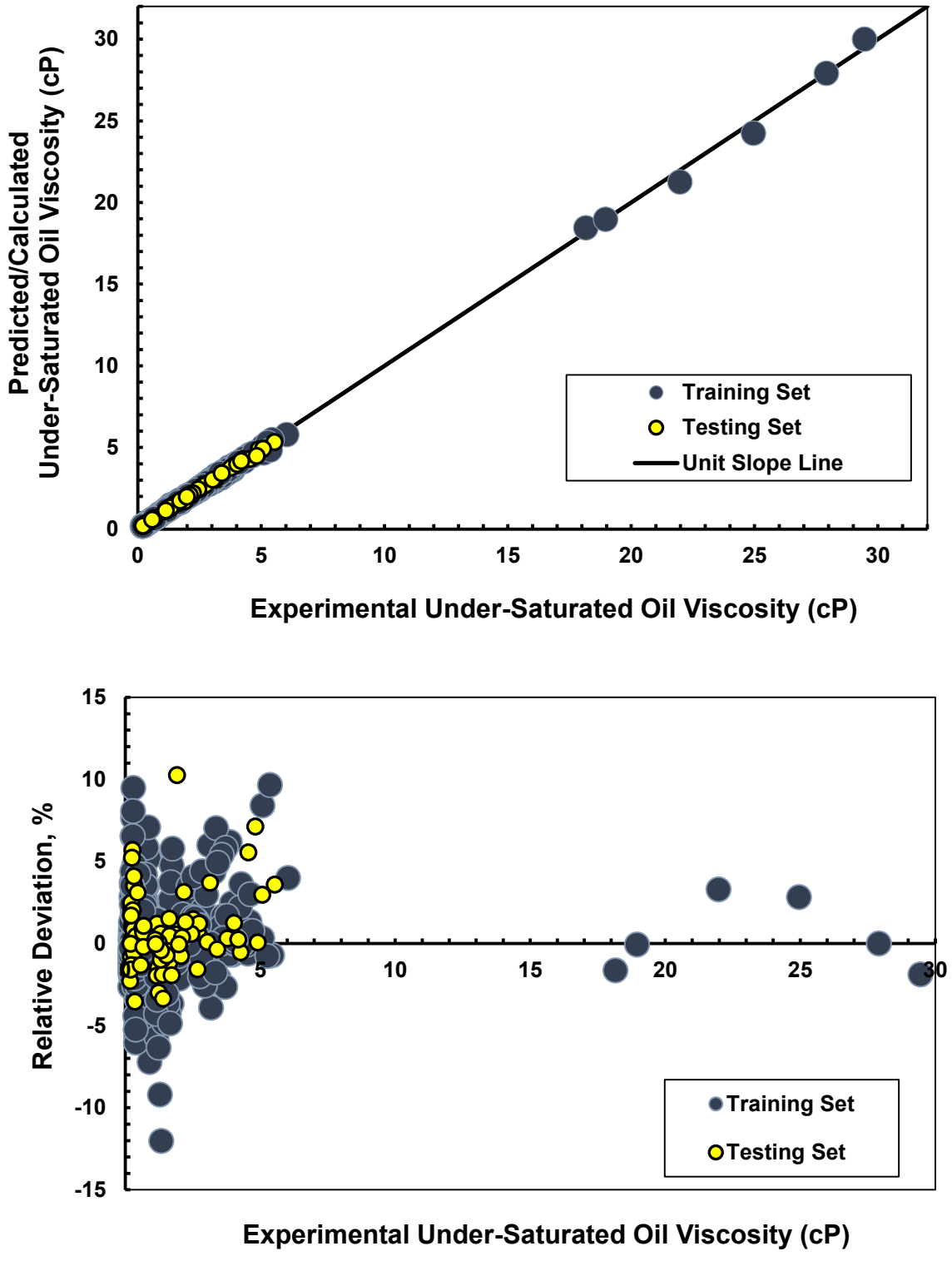

Fig. 5. 27 Parity diagram and relative deviation distribution plot for the developed under-saturated oil viscosity model.

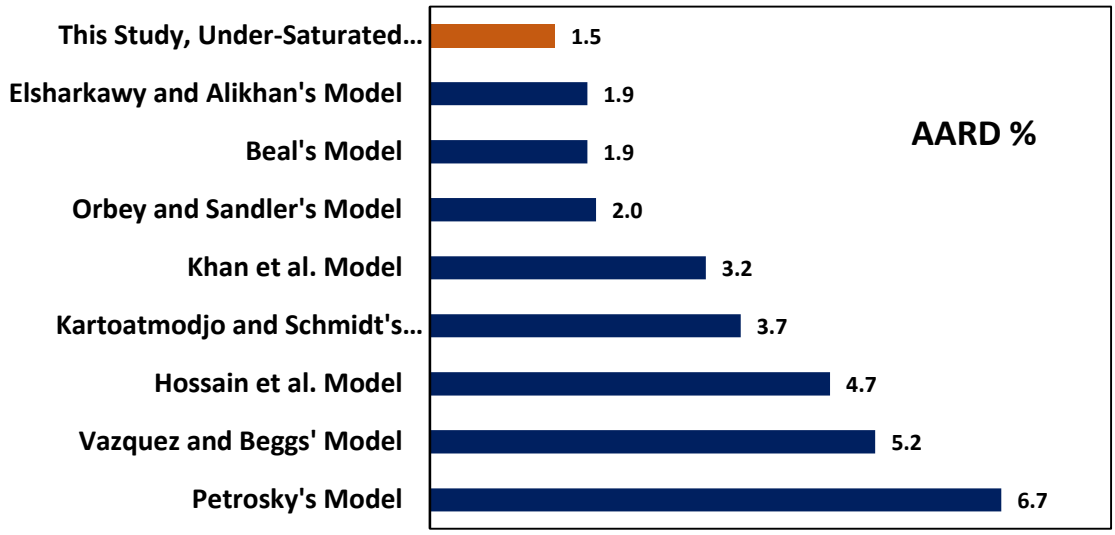

Fig. 5. 28 Comparison of the AARD % values obtained for the determination of under-saturated oil viscosity with comparative methods; the obtained AARD % for Labedi (Labedi, 1992) method is 1078.4 %.

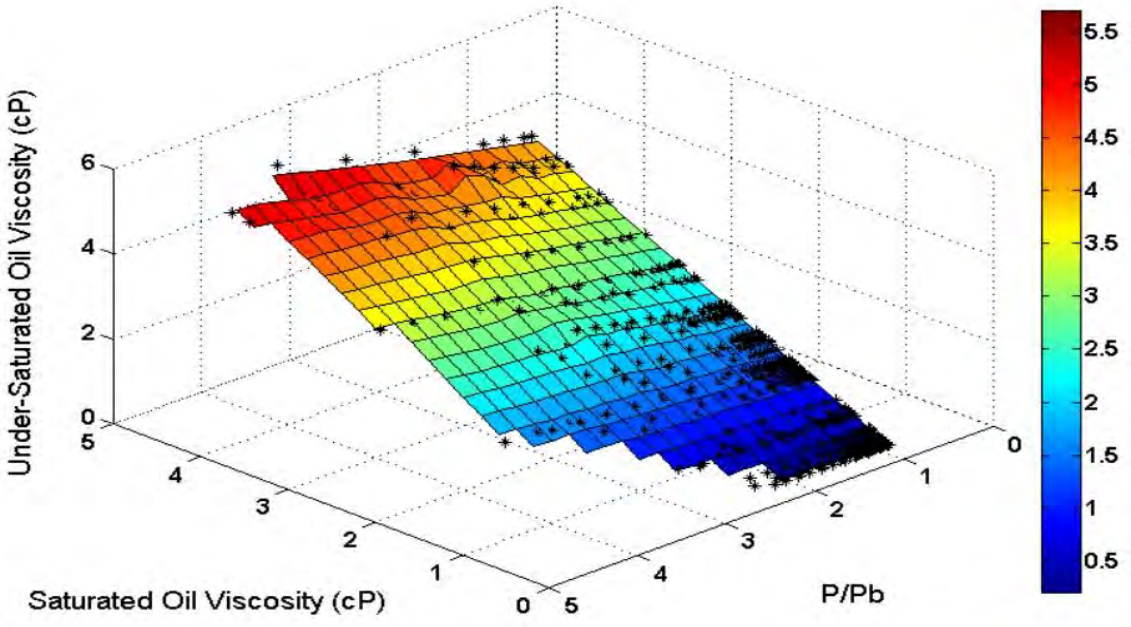

Fig. 5. 29 3D plot of change of under-saturated oil viscosity versus change in pressure up to the pressure at bubble point and saturated oil viscosity.

5.4.4. Identification of Outlier Data Points

Normally, there are some outlier data points in databases associated with dead, saturated and under-saturated reservoir oil viscosities. The resulting uncertainties in experimental data may lead to a high deviation in predicting the reservoir oil viscosities. Accordingly, the Leverage value statistics technique is employed in this study for detecting outlier data points available in the reservoir oil viscosities databanks (Goodall, 1993; Gramatica, 2007). As a consequence, the Williams plot in the Leverage analysis is implemented to show outlier data points on the basis of the H values calculated (Mohammadi et al., 2012a; Mohammadi et al., 2012c).

Figs. 5. 30-32 demonstrates the Williams plots for the calculated values of dead, saturated, and under-saturated reservoir oil viscosities applying the GEP-based models. It is evident from the figures that the majority of data points are in the ranges $0 \leq H \leq 0.08035$ and $-3 \leq R \leq 3$ for the dead oil viscosity model, $0 \leq H \leq 0.02233$ and $-3 \leq R \leq 3$ for the saturated oil viscosity, and $0 \leq H \leq 0.02298$ and $-3 \leq R \leq 3$for the under-saturated oil viscosity. This confirms that the equations presented in this study are statistically valid and correct in calculating those viscosities. In addition, **Figs. 5. 30-32** illustrates that there are only three data points for the developed dead oil viscosity model, five data points for the developed saturated oil viscosity, and finally, eight data points for the developed under-saturated oil viscosity, which is outside the applicability domain of the GEP-based models and are therefore probably accounted as outliers whose values may be doubtful.

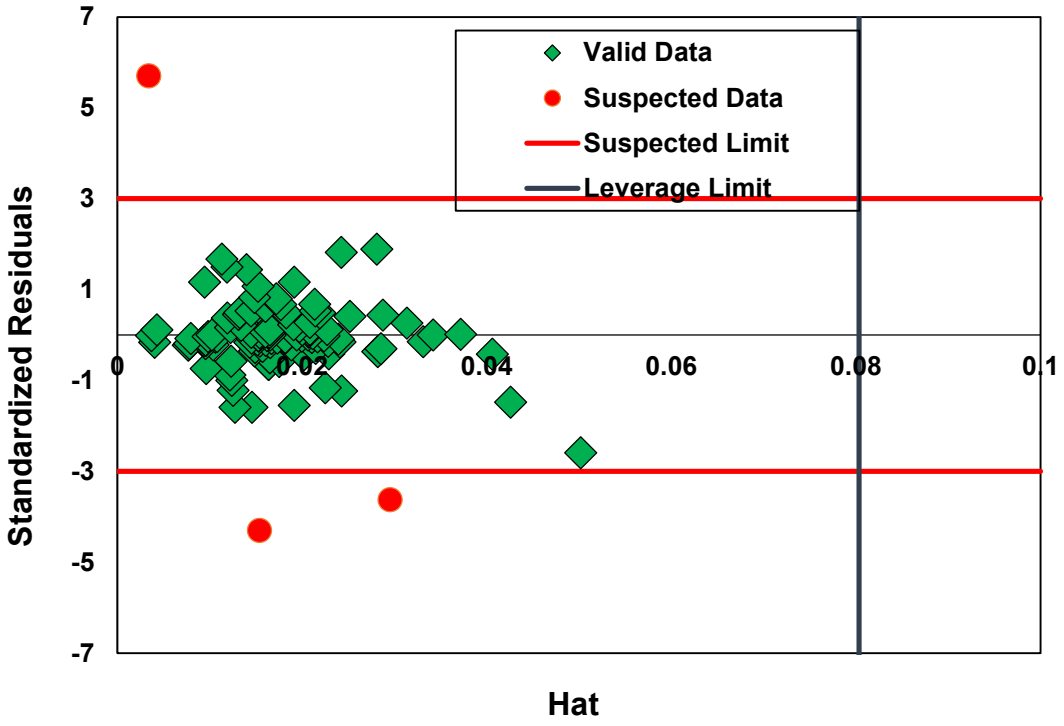

Fig. 5. 30 Detection of outlier data points existing in the dead oil viscosity dataset during development of the model using the Leverage approach.

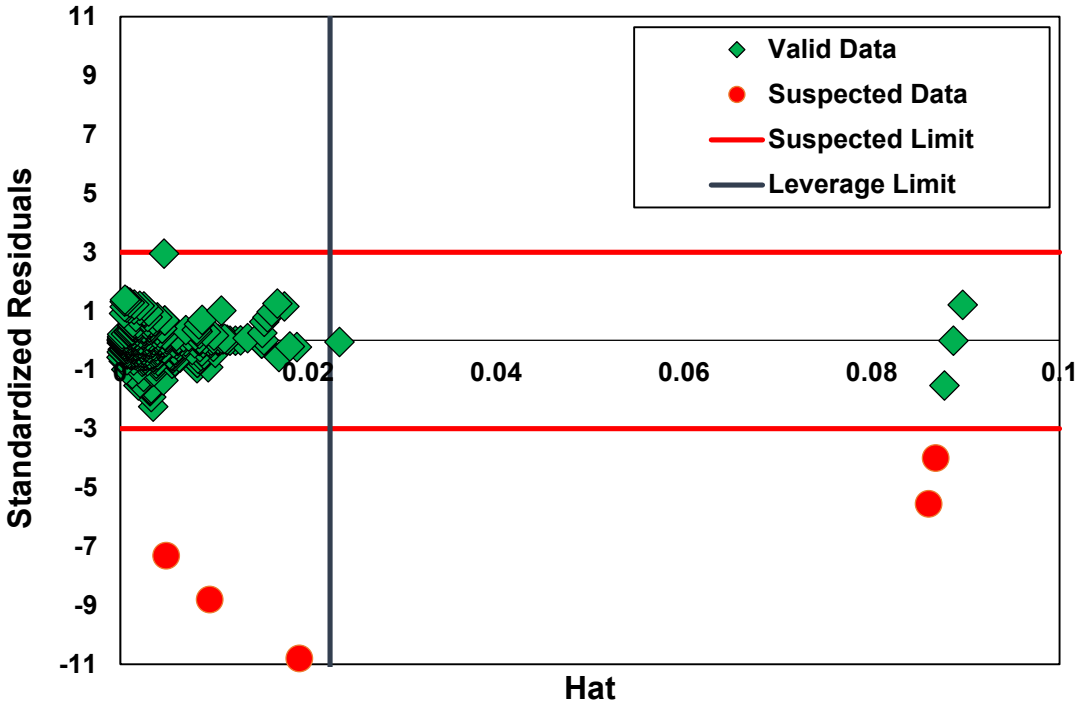

Fig. 5. 31 Detection of outlier data points existing in the saturated oil viscosity dataset during development of the model using the Leverage approach.

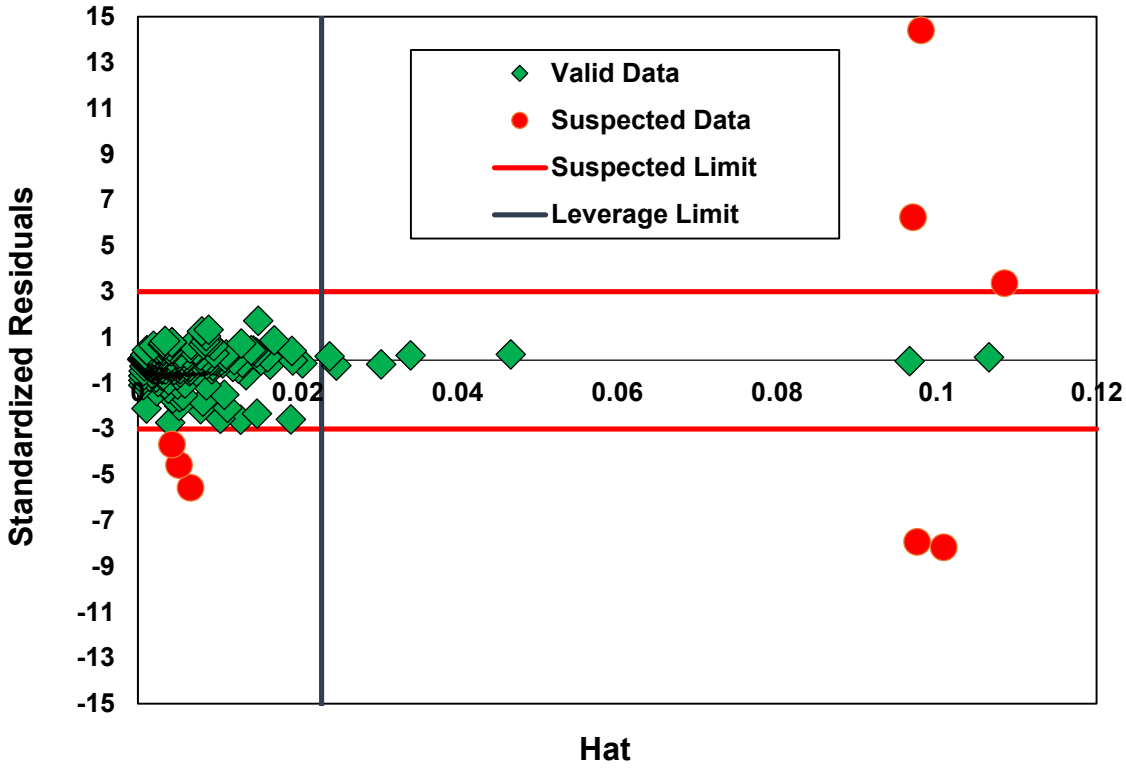

Fig. 5. 32 Detection of outlier data points existing in the under-saturated oil viscosity dataset during development of the model using the Leverage approach.

5.5. New Equation for Solution Gas–Oil Ratio

5.5.1. Development of the New Model

The key aim of the present study is to propose a comprehensive, accurate and reliable model for the determination of solution GOR using data collected from various crudes worldwide. To this end, the LINGO methodology (LINGO-Softwate, 2011) is used for pursuing our objective. Basically, the technique is an interactive linear and discrete tuning tool. As a result, the methodology has been utilized in mathematics, science, and industry, and employed to solve computer problems mathematically (Carvalho et al., 2012; Chuang et al., 2012; Vidal et al., 2007). Furthermore, quadratic programming, as well as linear, nonlinear, and integer programming are provided by LINGO software (Arabloo et al., 2014). As a result, the LINGO methodology can solve root and algebraic equations/problems linearly and nonlinearly.

It should be mentioned that the LINGO software includes a number of common mathematical functions within the programming language useful for finding/solving programming problems (Miao, 2011). In this study, the LINGO methodology is used to develop a reliable model to determine the solution GOR as a function of reservoir fluid properties, including bubble point/saturation pressure, reservoir temperature, and gas gravity, as well as oil gravity (API) as follows:

$$R_s = f(\gamma_g, T_R, API, P_b) \tag{5.12}$$

In the development of the model, the databank collected was separated into two sets of data, viz. the training and test sets. Approximately 80% of entire the databank was used in the development of the model (training set), and the rest (20%) was assigned to the test set for checking the model and evaluating its accuracy, performance, and capability. To measure the accuracy of the model, the average absolute relative deviation (AARD) is selected as an objective function. Finally, a simple form of equation with four easy functions including $\times, +, -, /$ was obtained as follows:

$$R_s = A + B - 15.849 \tag{5.13}$$

$$A = 0.14624\,P_b - 0.14624\,API + \frac{802.44}{P_b} + \frac{(2.7277\,P_b - API\,T_R + 2715.5)^2}{(API - 995.53)^2} \tag{5.14}$$

$$B = (0.0064332\,P_b + 0.0064332\,API\,\gamma_g) \times (API\,\gamma_g - 14.811) \tag{5.15}$$

where P_b denotes the bubble point pressure (psi), API stands for oil API gravity, T_R denotes the reservoir temperature (°F), and γ_g is gas specific gravity.

5.5.2. Performance Evaluation

A graphical statistical error analysis was conducted to evaluate the performance of the method over a wide range of the reservoir fluid properties, and to compare the results obtained using the model against the most widely used empirically derived correlations. Hence, AARD, root mean square error, and average relative percent error (ARPE) were considered as statistical error parameters, and a parity diagram or scatter plot, as well as

a relative distribution error curve, are used as two illustrations to evaluate the performance of the method proposed to estimate the solution GORs. The results of the error analysis are summarized in **Table 5. 9**.

The results in the table indicate that the model has acceptable accuracy with respect to the large number of data points and the wide range of reservoir fluid properties employed in its development. The AARD value reported for the model is 19.83%. The value indicates that the method output values are in agreement with the corresponding experimental records of the solution GOR. Furthermore, the calculated APRE and RMSE values are 1.73% and 203.05, respectively.

Table 5. 9 Error analysis performed for the proposed model and comparable methods investigated in this study.

Method	AARD, %	APRE, %	RMSE
Glaso (Glaso, 1980)	79.25	32.37	468.33
Petrosky and Farshad (Petrosky Jr and Farshad, 1993)	62.70	-48.16	217.38
Kartoatmodjo and Schmidt (Kartoatmodjo and Schmidt, 1994c)	57.80	-48.15	395.93
Standing (Standing, 1947b)	47.84	-38.61	312.88
Farshad et al. (Frashad et al., 1996)	43.07	-28.87	267.08
Vazquez and Beggs (Vazquez and Beggs, 1980)	42.29	-31.99	389.08
Al-Marhoun (Al-Marhoun, 1988)	42.01	-21.28	348.17
Dindoruk and Christman (Dindoruk and Christman, 2004)	36.89	-13.78	254.27
Macary and El-Batanony (Macary and El-Batanoney, 1993)	36.54	1.40	238.87
Al-Shammasi (Al-Shammasi, 1999)	32.95	-16.72	242.81
Baniasadi et al. (Baniasadi et al., 2015)	23.15	2.29	197.39
This study	19.83	1.73	203.05

To assess the performance of the proposed model, a scatter diagram, as well as a relative distribution error plot of values estimated, using Eq. (5.13), were plotted. **Fig. 5. 33** shows on a parity diagram a comparison of values estimated by the model versus

experimental values of solution GOR. It is clear from the figure that the estimated GOR values approximate the experimental values, resulting in data clustered around the parity line. This shows the capability of the new model in predicting more than 1000 data values for solution GOR.

Another graphical comparison is shown in **Fig. 5. 34**, which illustrates the calculated average relative percent error between the model and the experimental data for solution GOR. As can be seen in the figure, that the relative errors for the estimated data are clustered around the zero line. This demonstrates that there is acceptable agreement between the model predictions and experimental data for solution GOR. **Fig. 5. 35** shows the plot of the predicted values against experimental data for the solution GOR with respect to the sorted data index.

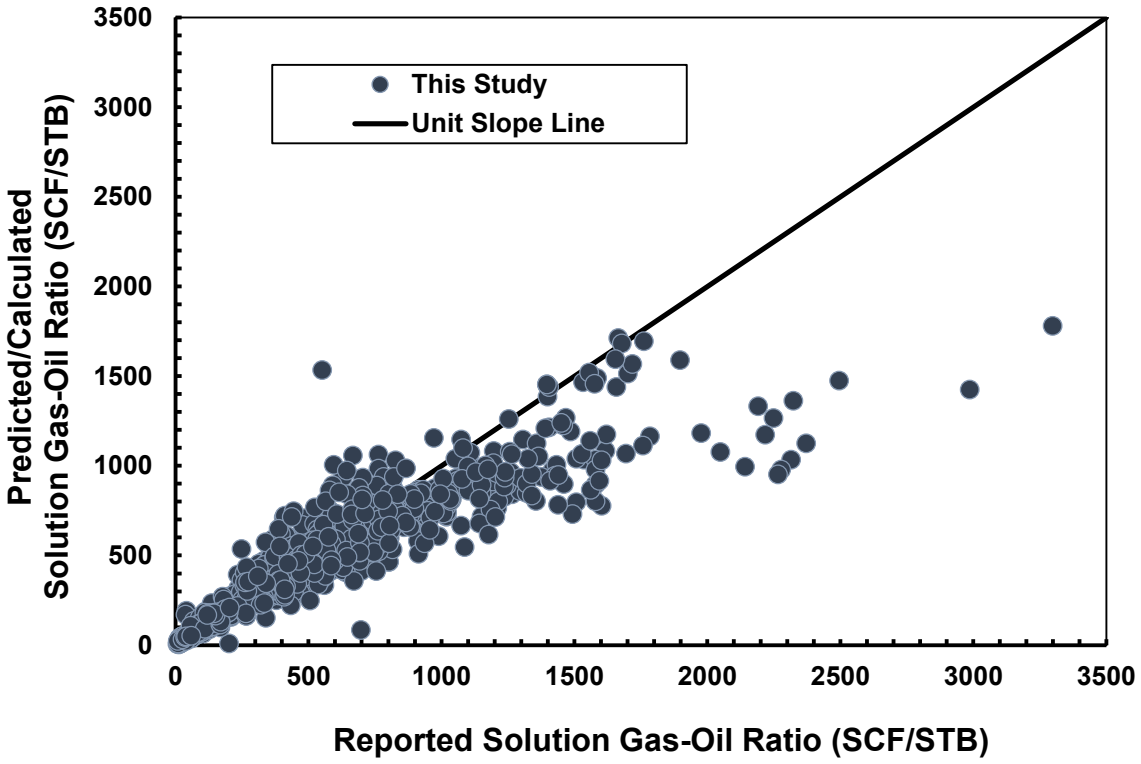

Fig. 5. 33 Scatter diagram of the predicted solution gas-oil ratio values versus the experimental records.

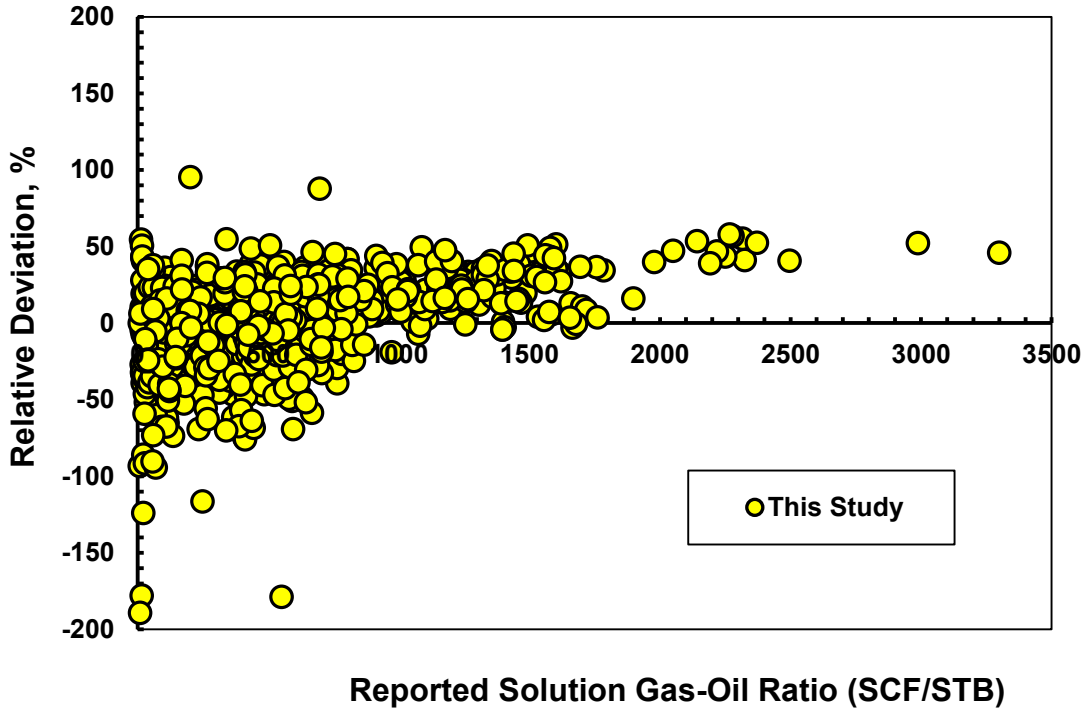

Fig. 5. 34 Relative error distribution plot of the predicted solution gas-oil ratio values versus the experimental records.

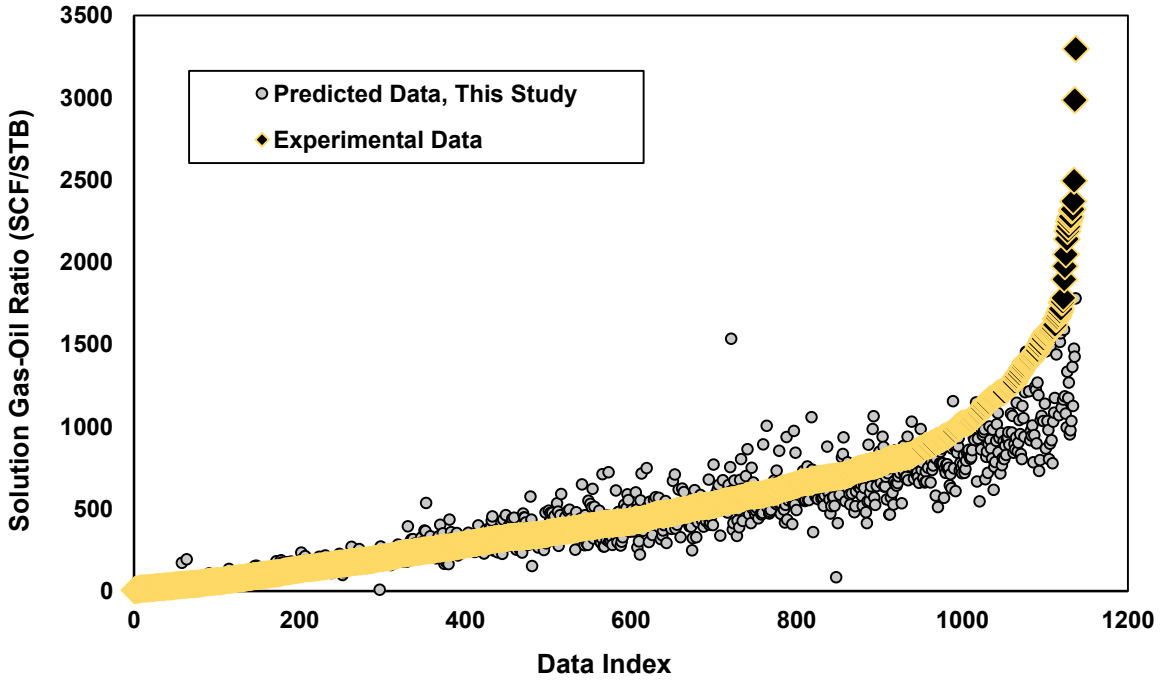

Fig. 5. 35 A fitting curve, as a sorted data index, for the predicted solution gas-oil ratio values versus the experimental records.

In order to evaluate the performance of the method in predicting solution GOR data, a comparative analysis was undertaken between the model developed in this study and widely-used empirically derived methods including the Farshad et al. method (Frashad et al., 1996), Macary and El-Batanony method (Macary and El-Batanoney, 1993), Petrosky and Farshad method (Petrosky Jr and Farshad, 1993), Vazquez and Beggs method (Vazquez and Beggs, 1980), Al-Marhoun method (Al-Marhoun, 1988), Kartoatmodjo and Schmidt method (Kartoatmodjo and Schmidt, 1994c), Al-Shammasi method (Al-Shammasi, 1999), Standing method (Standing, 1947b), Glaso (Glaso, 1980), Baniasadi et al. method (Baniasadi et al., 2015), and the Dindoruk and Christman method (Dindoruk and Christman, 2004). **Table 5. 9** reports on the statistical results obtained from the comparisons.

The table shows that the new model performs better for the calculation of solution GOR. **Fig. 5. 36** illustrates the calculated AARD for the model and the comparative methods. From **Table 5. 9** and **Fig. 5. 36**, it can be concluded that the methods of Baniasadi et al. (Baniasadi et al., 2015), Al-Shammasi (Al-Shammasi, 1999), Macary and El-Batanony (Macary and El-Batanoney, 1993), Dindoruk and Christman (Dindoruk and Christman, 2004), and Al-Marhoun (Al-Marhoun, 1988) are, after the method proposed in this study, the most accurate for the calculation of solution GOR with AARD values of 23.15, 32.95, 36.54, 36.89, and 42.01%, respectively.

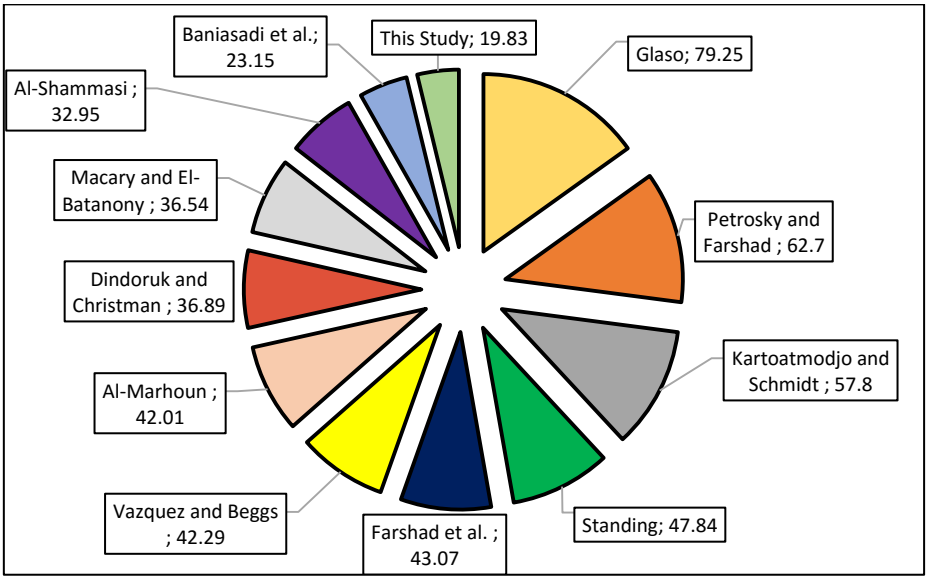

Fig. 5. 36 Graphical comparison of the developed model against the comparative methods studied in terms of the statistical error parameter of AARD.

Table 5. 10 lists some random data points selected from the databank, and **Table 5. 11** summarizes the estimated values for the data points presented in **Table 5. 10** using both the method developed and the above-mentioned empirical methods. **Table 5. 11** also confirms superior performance of the new model over the empirical methods to which it was compared.

Table 5. 10 Records of some data points existing in the databank compiled in this study.

Data Index	P_b	γ_g	API	T_R	R_{si}
1	2082.77	0.756	7.5	153.5	208.7
2	2076.97	0.815	10.5	152.6	260.0
3	554.99	0.68	12	74.9	52.4
4	599.98	0.74	14.8	82.8	68.0
5	825.28	1.411	19.4	172.4	177.8
6	3199.97	0.75	20.9	110.5	556.2
7	285	0.74	23	114.8	32.5
8	430.04	1.04	25	99.4	95.5
9	1499.98	0.64	27	107.7	239.9
10	909.97	0.67	30	88.25	171.8
11	3057	0.778	32	175	679.0
12	400.01	0.8	34	71.6	76.6
13	2775.01	0.823	35.7	140.5	689.4
14	1340	0.8	36.3	87.8	313.4
15	1415	1.2468	37.2	248	486.0
16	5760.99	0.924	40.1	302	1760.6
17	1153.05	0.85	40.4	105.5	299.6
18	2221	0.693	45.3	238	547.0
19	1386.97	0.763	46.5	116.03	367.6
20	1962.07	0.78	52.5	138.25	636.7
21	1170.47	0.649	56.8	140.7	300.9

Table 5. 11 A point-to-point comparison between the results obtained and comparative methods, as applied to the experimental records reported in Table 5. 10.

Data Index	This Study	ARD%	Baniasadi	ARD%	D& Christman	ARD%	Al-Marhoun	ARD%	Macary	ARD%	Al-Shamasi	ARD%
1	219.0	4.9	102.8	50.8	291.9	39.9	194.7	6.7	288.9	38.4	199.0	4.6
2	250.2	3.8	150.2	42.2	323.3	24.4	260.2	0.1	310.8	19.5	236.7	9.0
3	52.4	0.0	39.5	28.5	78.5	42.3	35.6	35.5	64.2	42.3	43.6	3.4
4	66.1	2.9	57.7	15.2	92.7	36.3	50.8	25.4	75.5	36.0	59.4	6.9
5	174.6	1.8	36.5	6.5	91.1	133.6	40.2	3.0	40.0	2.5	28.5	27.0
6	555.0	0.2	38.9	22.5	53.7	69.0	20.2	36.4	44.3	70.9	27.4	5.9
7	30.3	6.6	170.0	17.6	167.8	18.7	142.0	31.2	161.5	4.1	170.4	1.2
8	80.2	16.1	669.0	12.1	631.0	17.1	856.4	12.5	1107.4	78.3	900.9	45.0
9	240.4	0.2	241.8	0.8	184.0	23.3	163.4	31.9	233.7	19.4	264.7	35.2
10	152.3	11.3	167.3	2.6	123.2	28.3	106.1	38.2	132.3	5.6	163.9	16.9
11	658.8	3.0	776.1	10.2	729.1	3.5	996.8	41.5	1094.5	90.4	1064.8	85.2
12	75.8	1.1	324.2	5.2	257.2	24.7	364.2	6.5	265.4	4.9	366.0	31.2
13	678.0	1.6	749.8	23.6	653.9	7.8	665.6	9.7	909.5	83.7	967.1	95.3
14	311.7	0.5	331.2	5.7	250.8	20.0	344.2	9.8	252.9	1.1	362.2	41.6
15	490.7	1.0	366.7	3.8	404.7	6.2	591.2	55.2	200.3	47.4	356.3	6.5
16	1694.3	3.8	473.4	0.4	376.5	20.2	595.1	26.2	369.2	4.1	546.3	41.9
17	299.5	0.0	616.6	22.1	492.0	37.8	510.4	35.5	535.7	32.3	692.3	12.5
18	547.4	0.1	633.3	0.1	461.5	27.2	440.4	30.5	493.2	22.2	702.5	10.8
19	371.2	1.0	426.7	16.1	265.6	27.8	376.6	2.5	290.3	3.2	487.6	62.5
20	601.4	5.5	690.7	8.5	478.9	24.8	699.1	9.8	526.9	1.4	889.1	71.1
21	323.9	7.6	400.1	33.0	177.6	41.0	230.1	23.5	240.0	20.2	497.8	65.4

5.5.3. Influence of the Reservoir Fluid Properties on Solution GOR

As pointed out earlier, reservoirs containing light oils have more dissolved gases than reservoir with heavy oils. Therefore, it would be interesting to determine the accuracy of the model developed for various ranges of oil API gravity. To this end, the capability of the model presented in this study for estimating the solution GOR was scrutinized across the spectrum of the light to heavy oils. The solution GORs estimated by the proposed model were partitioned into four classes of oil API gravities, viz. 6-15, 15-25, 25-35, and 35-56.8°. The results of analysis in terms of the calculated AARD values is shown in **Fig. 5. 37**. As can be seen in the figure, the model errors for the estimation of solution GORs of light oils is less than that for heavy oils. It can therefore be concluded that the model developed in this study is more accurate for the measurement of crudes with higher values of oil API gravity.

To further investigate the influence of reservoir fluid properties, including saturation pressure, reservoir temperature, gas specific gravity, and oil API gravity, a sensitivity analysis was performed using the relevancy factor approach. **Fig. 5. 38** shows the results of sensitivity analysis. The figure indicates that bubble point pressure and gas specific gravity have the largest and smallest influences, respectively, on the solution GOR values predicted by the model.

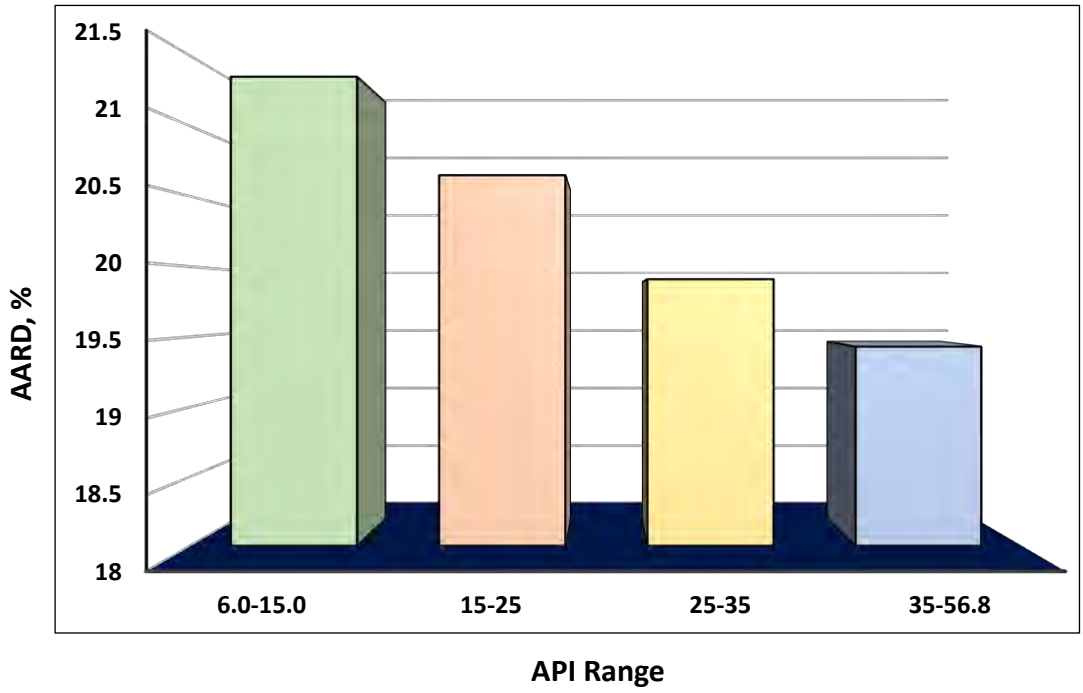

Fig. 5. 37 Accuracy of the model developed in this study in different API ranges.

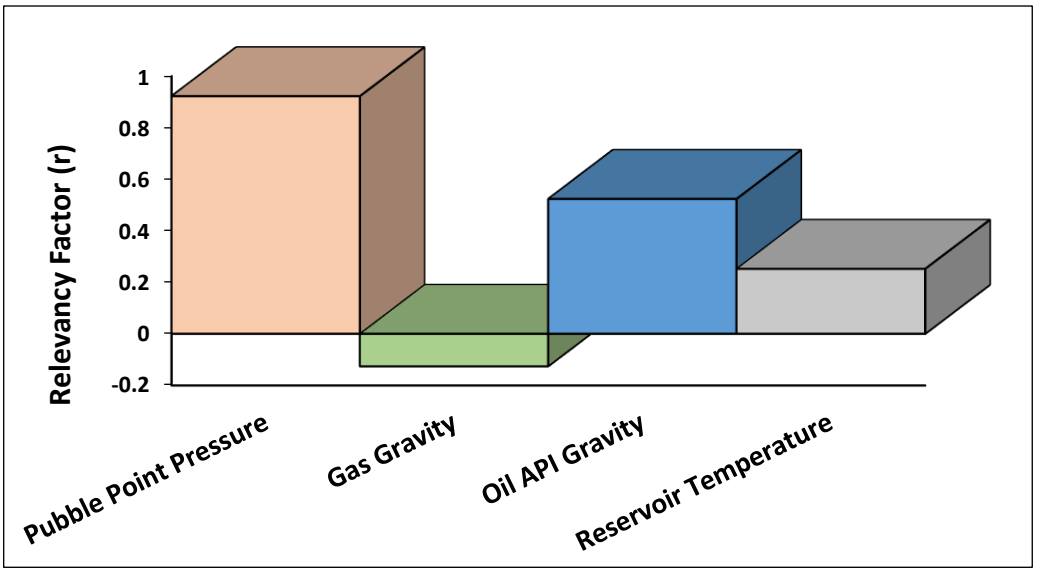

Fig. 5. 38 Degree of importance for each input parameter for the prediction of solution gas-oil ratio.

5.5.4. Detection of Outlier Solution GOR Data Points

The detection of outlier data points that exist in a databank used to develop a predictive model is important in order to determine the applicability domain of the model. To this end, the Leverage methodology (Gharagheizi et al., 2012a; Mohammadi et al., 2012c; Rousseeuw and Leroy, 2005) is utilized in this study to identify outlier data points in the solution GOR databank that was compiled. Detailed information on the Leverage methodology in terms of mathematical equations, as well as a step-by-step procedure is reported elsewhere (Gharagheizi et al., 2012a; Mohammadi et al., 2012c).

The Williams diagram is sketched to show the applicability domain of the proposed method. The existence of a majority of solution GOR data in the domain $0 \leq H \leq 0.1428$ and $-3 \leq$ Standardized Residuals ≤ 3 demonstrates that the method is statistically valid. The data points which are located in the domain range $-3 \leq$ Standardized Residuals ≤ 3 are recognized as valid solution GOR data, and data which are outside the range are considered as outliers. The results show that only 26 data points in the solution GOR databank (among more than 1000 data points) were identified as outlier data points (**Fig. 5. 39**).

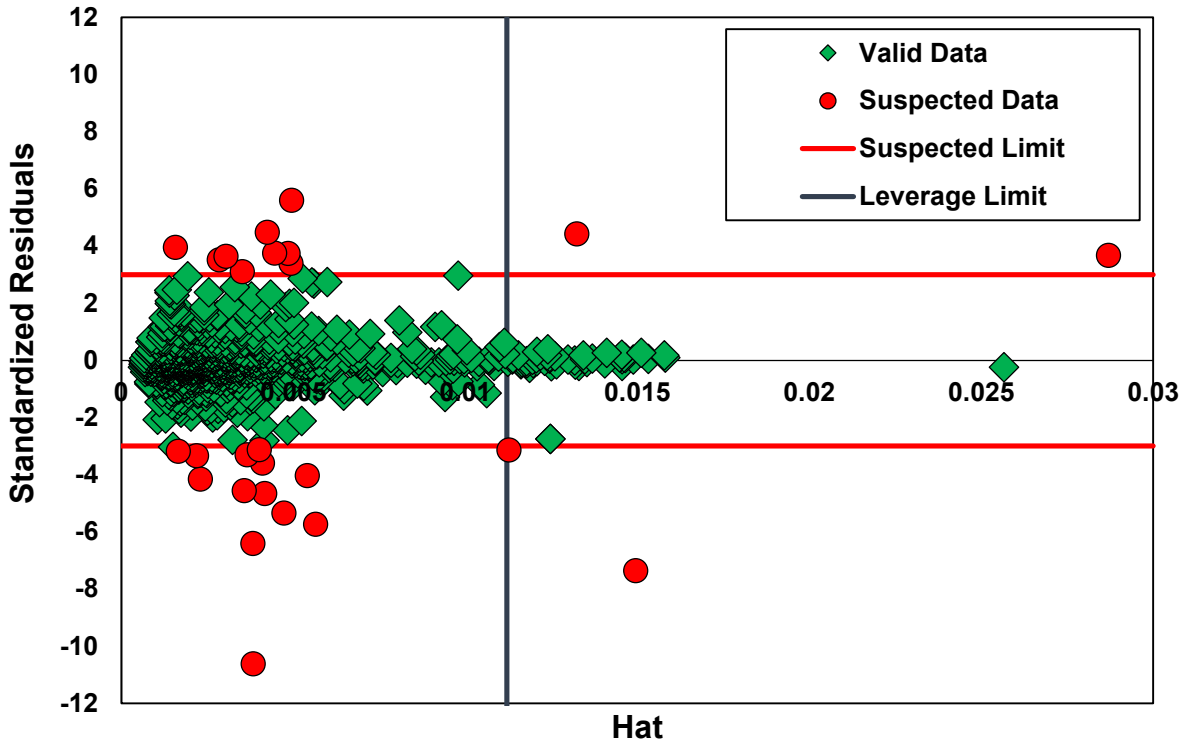

Fig. 5. 39 Plot of the Leverage analysis for the recognition of outlier data points.

5.6. Intelligent Model Development for Asphatene Precipitation

In this study, three reliable models have been proposed to predict the amount of asphaltene precipitated from a crude oil extracted from Iranian reservoirs. To this end, the asphaltene precipitation model is based on the solvent to oil dilution ratio, temperature, and solvent molecular weight. As mentioned earlier, the LSSVM modeling approach has two adjustable parameters including γ and σ^2 which should be optimized through an external optimization methodology.

In this study, a coupled simulated annealing (Atiqullah and Rao, 1993; Fabian, 1997; Vasan and Raju, 2009) was employed for obtaining the optimum values of the LSSVM parameters. As a result, the values adjusted by the CSA approach for the LSSVM developed to predict asphaltene precipitation are 1.0465 and 927850.8749 for σ^2 and γ, respectively.

To develop a DT model, we used the available MATLAB codes related to the regression DT approach. In the regression DT approach, the Y variable takes ordered values and a regression model is fitted to each node to give the predicted values of Y (Loh, 2011).

To propose the GEP-based model, we applied one gene with 30 chromosomes, and an average absolute relative deviation was utilized as an accuracy function. The head size is equal to 7 and a function set including *, /, – and + is selected while applying the GEP methodology. To achieve a highly accurate and capable model, the stop condition of the algorithm was set on maximum generation with a best number of 72 thousands. The final model obtained by the GEP algorithm is simple-to-use with the lowest possible coefficients as follows:

$$W_t = \frac{-0.9048\, R_v\, (M\ -\ 188.76)}{6.1211\, R_v\ +\ T\ +\ 33.302} \tag{5.16}$$

where W_t denotes the amount of asphaltene precipitated (wt. %), M stands for the solvent molecular weight, T expresses the temperature (°C), and finally, R_v is the solvent to oil dilution ratio (mL/g).

Furthermore, the AARD error parameter was applied to compare the results obtained by these models with those of the artificial neural network model proposed by Ashoori et al. (Ashoori et al., 2010) as well as the scaling equation, and asphaltene scaling equations presented by Hu et. al (Hu and Guo, 2001) and Rassamdana et al. (Rassamdana et al., 1996). **Fig.5. 40** illustrates the comparison of AARD calculated for all models developed in this study (*i.e.* the LSSVM, DT and GEP-based models) as well as the ANN modeling approach, and three asphaltene scaling equations proposed by Ashoori et al. (Ashoori et al., 2010), Hu et. al (Hu and Guo, 2001), and Rassamdana et al. (Rassamdana et al., 1996). The ANN model includes three layers of input, output, and hidden layers with 10 hidden neuron numbers.

In order to compare all models investigated in this study, two panels, namely, smart techniques and symbolic equations, should be considered. In the smart techniques panel, the LSSVM model shows the highest accuracy in comparison with the DT and ANN models. The AARD reported for the LSSVM model is 3 %, while AARDs calculated for the ANN and DT models are 5 and 11%, respectively. In the other panel, the equation

developed based on the GEP approach is more accurate than the scaling equations proposed by Ashoori et al. (Ashoori et al., 2010), Hu et. al (Hu and Guo, 2001), and Rassamdana et al. (Rassamdana et al., 1996). The AARD calculated for methods mentioned above are 8.5, 10.9, 17.3, and 17.9, respectively.

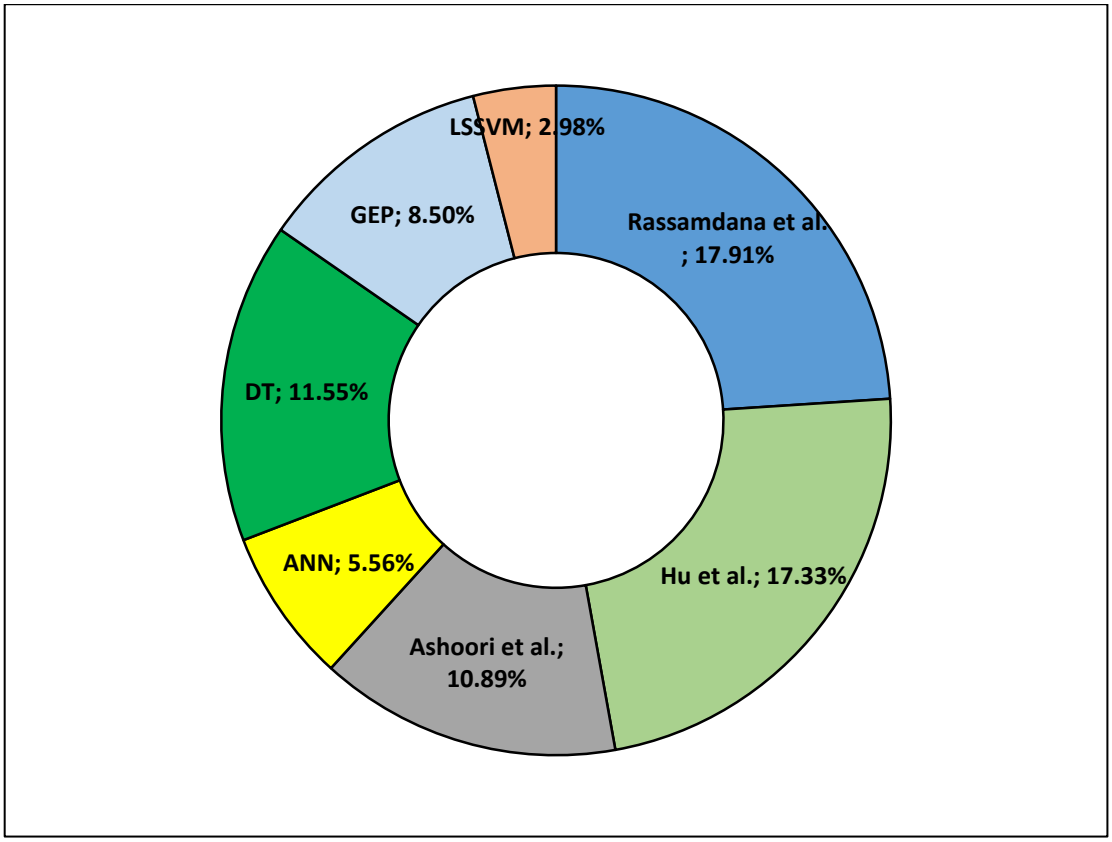

Fig. 5. 40 AARD calculated by means of the developed model and the comparative methods.

To show the capability performance of the models investigated in this study, a graphical analysis, using a parity diagram or crossplot and relative error distribution plot has been performed. To this end, the four most accurate methods investigated (*e.i.* the LSSVM, ANN, DT, and GEP-based models) in predicting the asphaltene precipitation have been considered. **Fig. 5. 41** provides the crossplots for four methods mentioned above. It is clear from the figure that the distribution of data points predicted by the LSSVM model, around the unit slope line, is lower than from the other methods. In other words, the R-squared error obtained by the LSSVM model is higher than from the ANN, DT, and GEP-

based models. **Fig. 5. 42** illustrates the relative error distribution plots for the LSSVM, ANN, DT, and GEP-based models. The figure clearly shows that the distribution of relative error around Y=0 (zero error) is lower than that from the other methods.

To show the smoothness performance of the models mentioned above, a trend analysis is undertaken of the asphaltene precipitation versus solvent to oil dilution ratio data at various temperatures for *n*-pentane, *n*-hexane, and *n*-heptane solvents. **Figs. 5. 43-45** indicate the trend plot of asphaltene precipitation changes versus the dilution ratio at a temperature of 30 °C for *n*-pentane, *n*-hexane, and *n*-heptane solvents, respectively. Furthermore, **Figs. 5. 46-48** illustrates the changes of asphaltene precipitation versus solvent to oil dilution ratio at temperature of 50 °C for *n*-pentane, *n*-hexane, and *n*-heptane solvents, respectively. Finally, **Figs. 5. 49-51** show the changes of asphaltene precipitation at 70 °C.

It is clear from the figures that the data points related to the LSSVM model are matched more closely with the experimental values than are the other models'. This shows that the LSSVM model is more capable for the prediction of asphalte precipitation as a function of temperature, solvent to oil dilution ration and solvent molecular weight. Additionally, the model proposed by the LSSVM approach has only two adjustable parameter, while the other methods require more parameters. A high number of adjustable parameters can increase the error of a model. The results obtained in this study confirm that the model is more appropriate for the prediction of targets in the petroleum industry, and also for the simulation of heavy organics precipitation such as asphaltene.

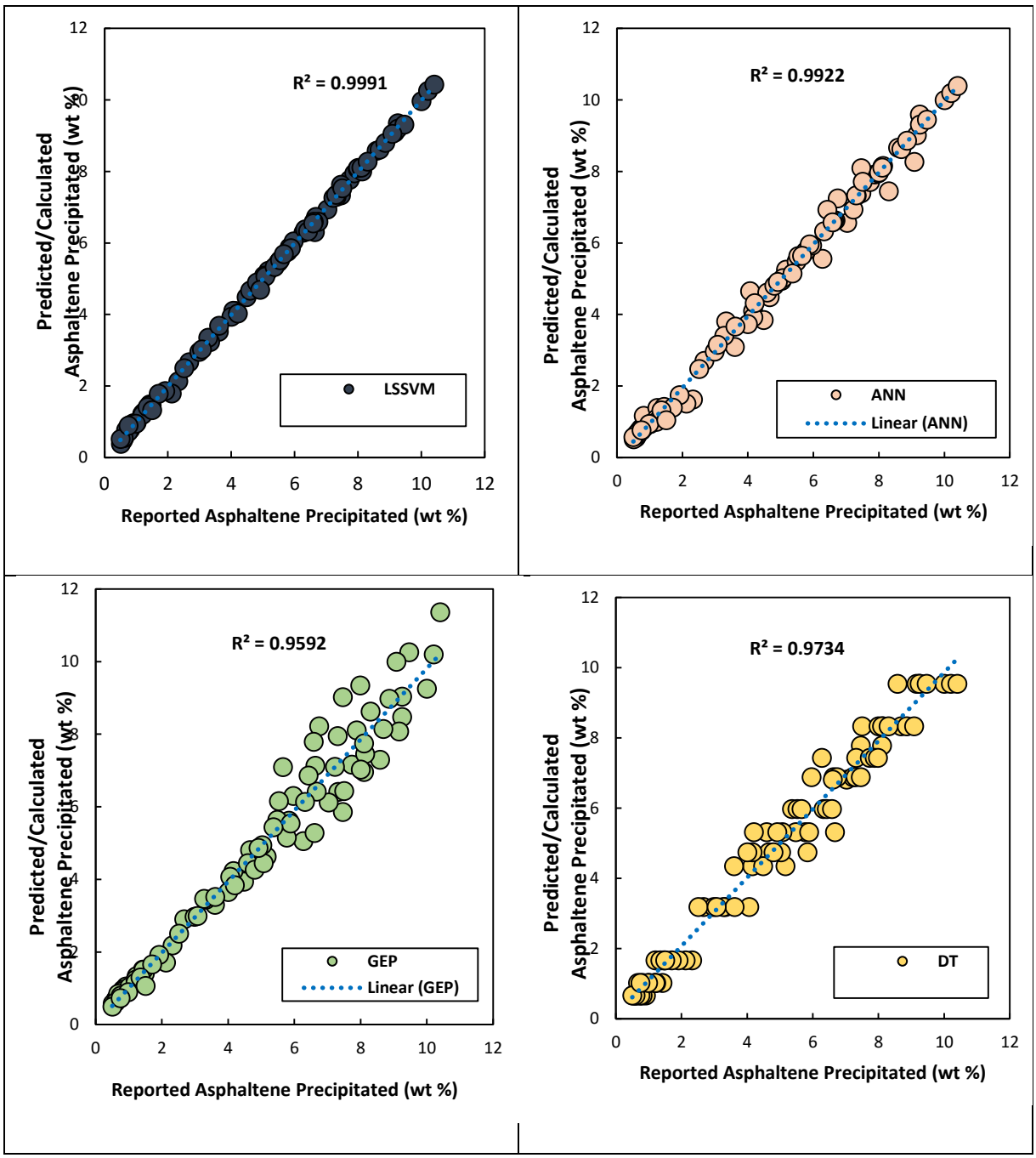

Fig. 5. 41 Crossplots for the different methods investigated in this study with respect to R-squared error.

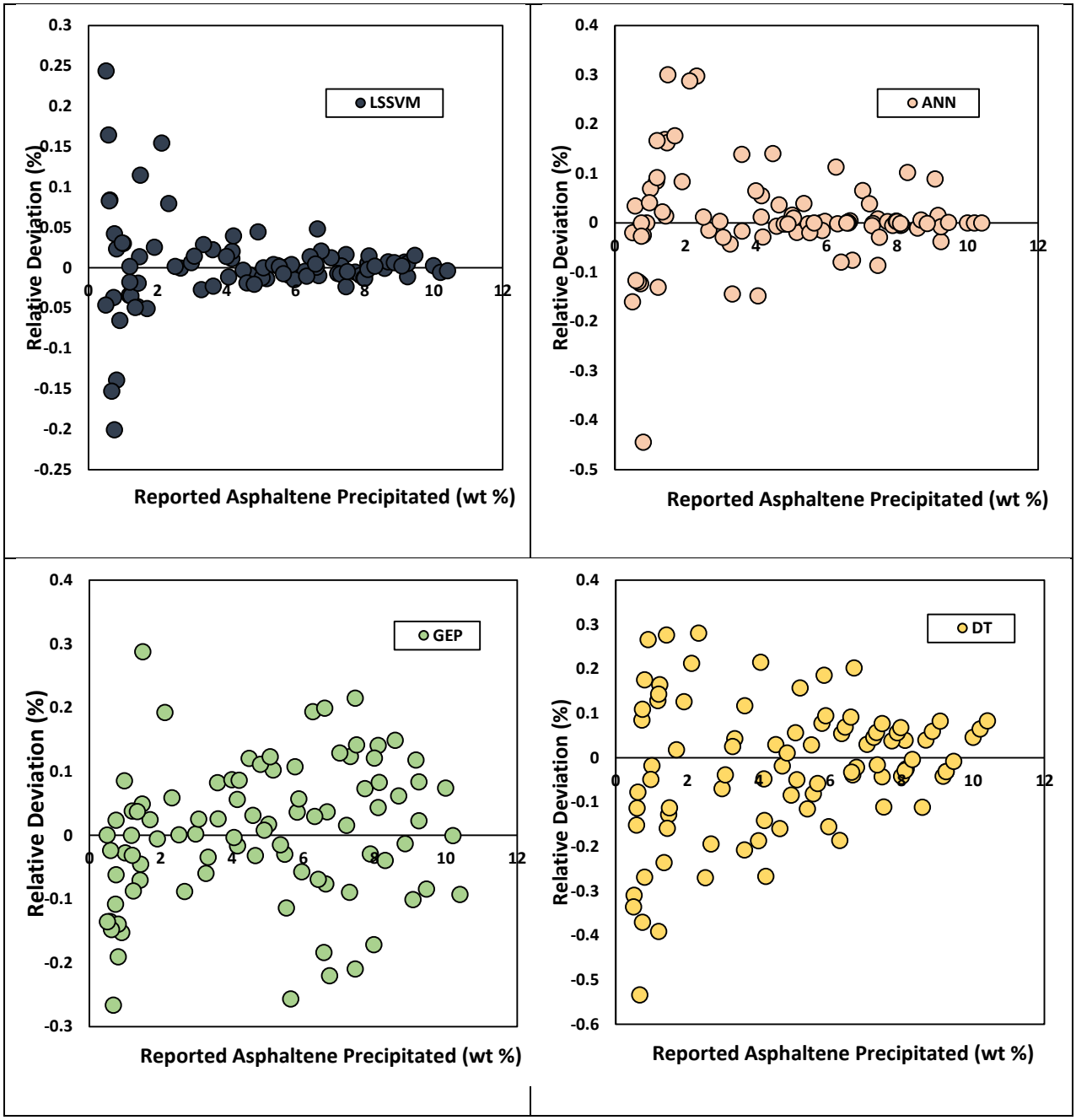

Fig. 5. 42 Relative error distribution analysis of the different methods investigated in this study.

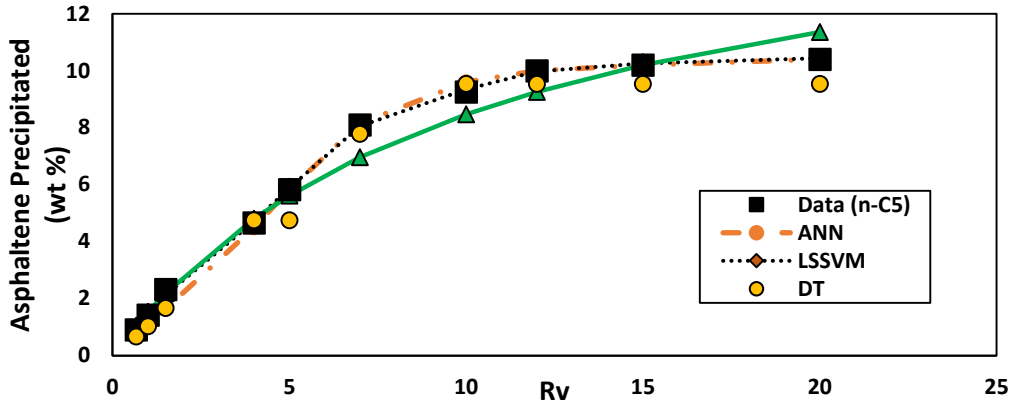

Fig. 5. 43 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 30 °C for n-C_5 solvent.

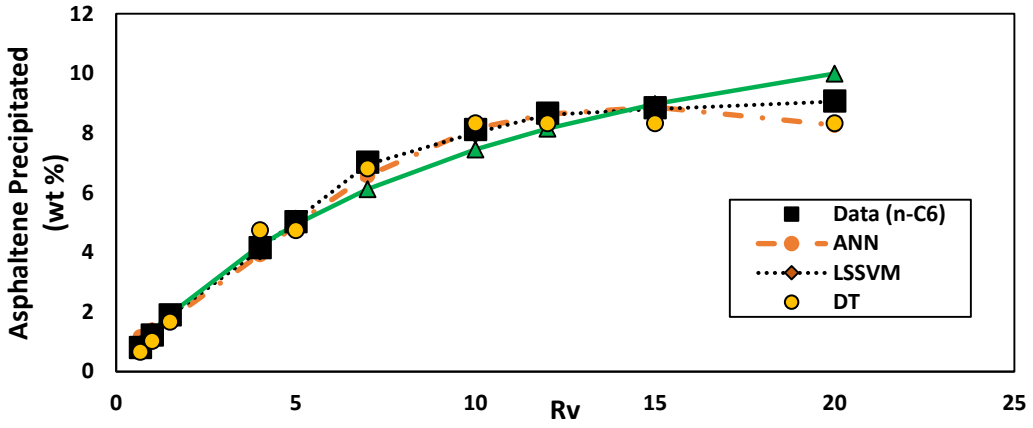

Fig. 5. 44 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 30 °C for n-C_6 solvent.

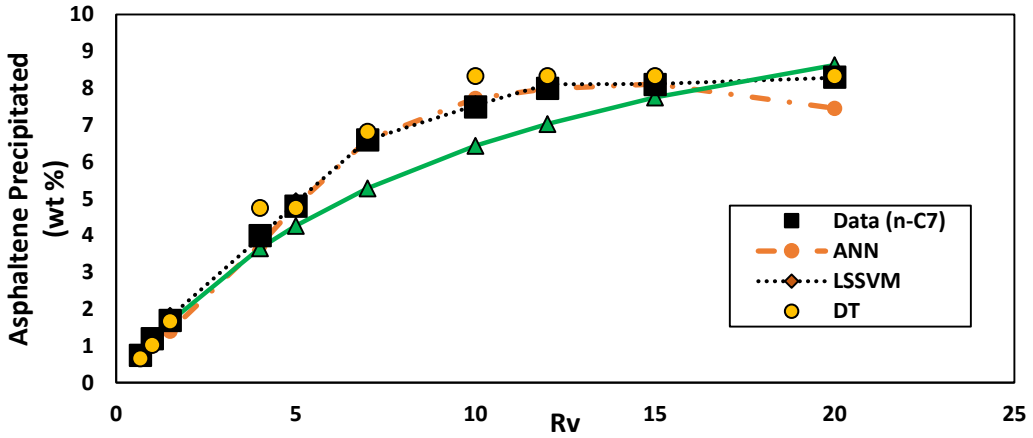

Fig. 5. 45 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 30 °C for n-C_7 solvent.

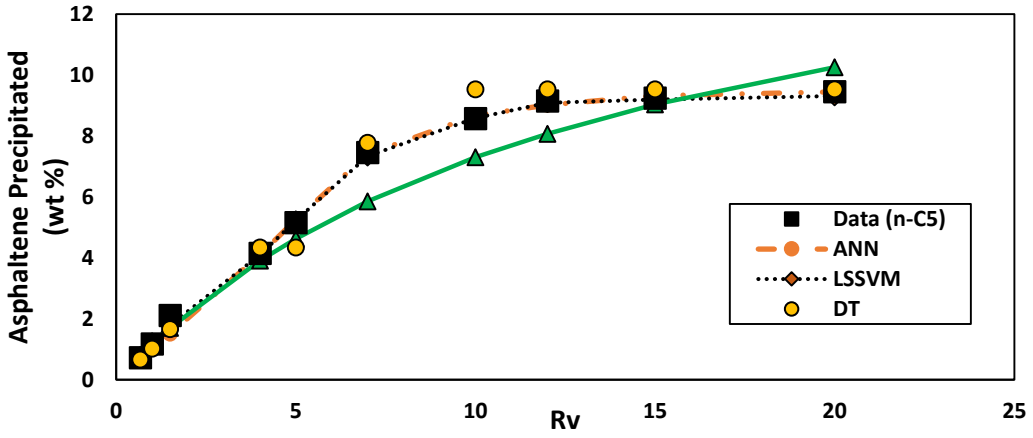

Fig. 5. 46 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 50 °C for n-C_5 solvent.

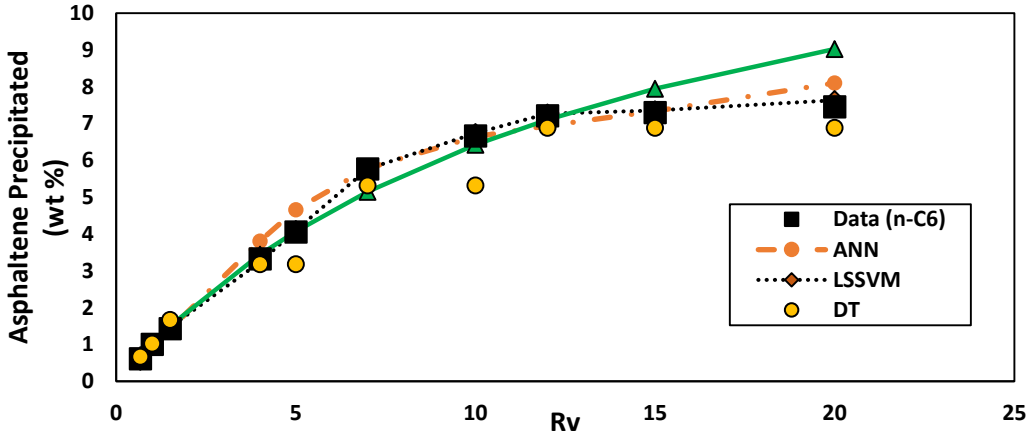

Fig. 5. 47 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 50 °C for n-C_6 solvent.

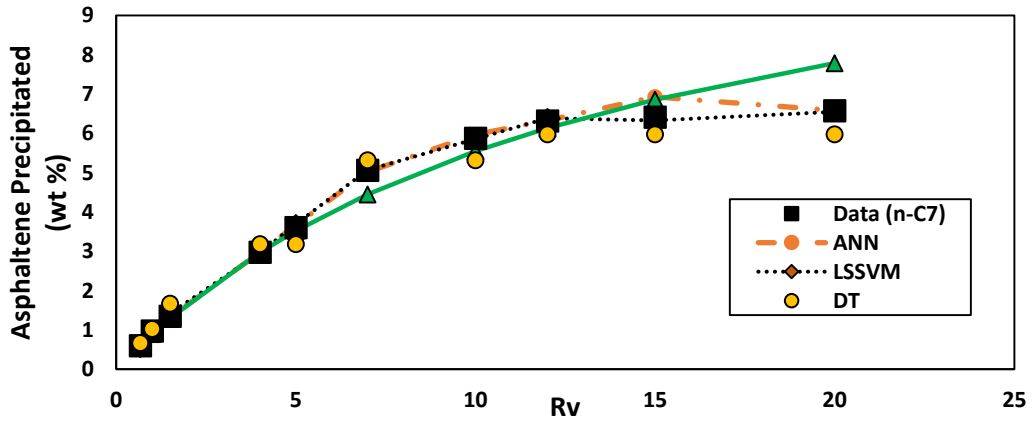

Fig. 5. 48 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 50 °C for n-C_7 solvent.

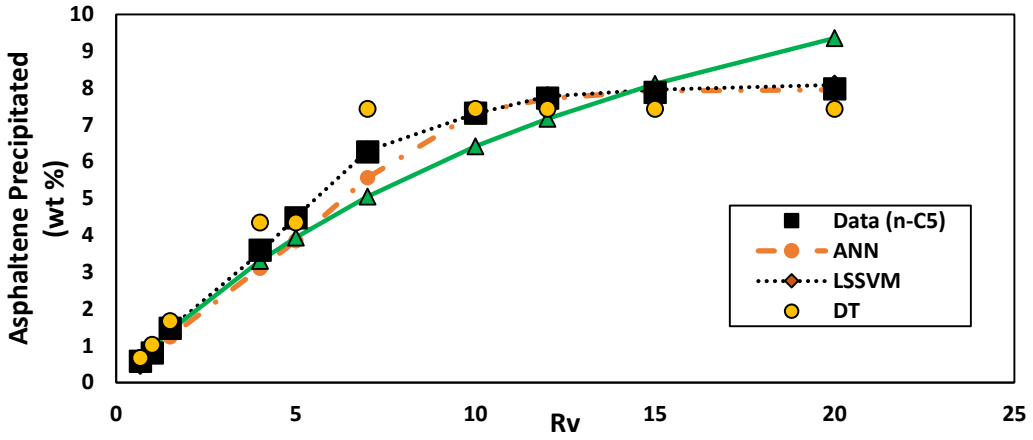

Fig. 5. 49 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 70 °C for n-C_5 solvent.

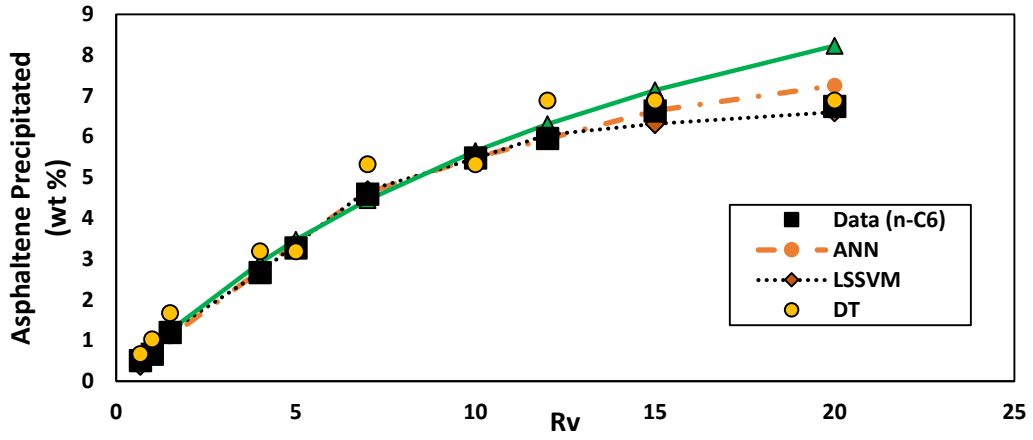

Fig. 5. 50 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 70 °C for n-C_6 solvent.

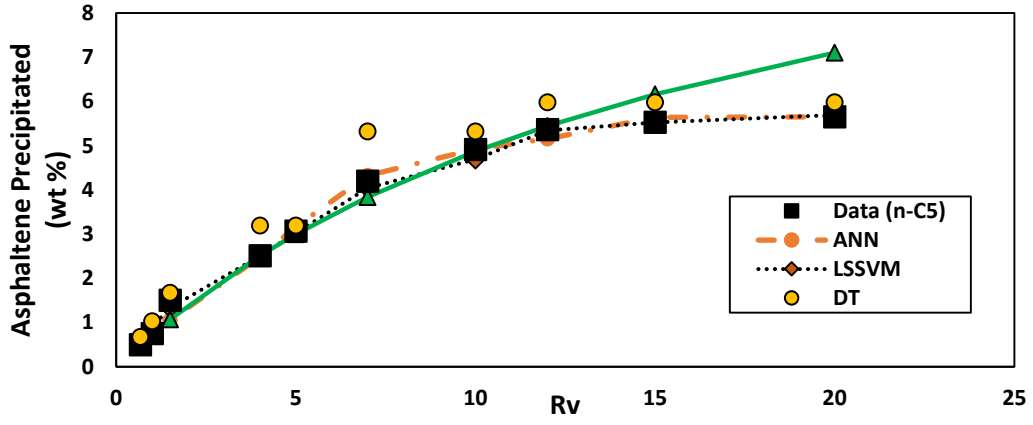

Fig. 5. 51 Trend plot of asphaltene precipitation changes versus dilution ratio at temperature of 70 °C for n-C_7 solvent.

Classification of crude oil by solubility is the most relevant to asphaltene association and solubility modeling. There are four major solubility fractions: saturates aromatics, resins, and asphaltenes (SARA). The input divided into two classes: saturate + asphaltene and resin + aromatic, and consequently output divided into three modes: severe, mild, and no/minor problems. To classify the asphaltene stability, the DT and ANFIS approaches have been employed in this study. Furthermore, the results obtained by the mentioned approaches are compared with the output of LSSVM algorithm (Chamkalani, 2015). Our conclusions elucidate that the proposed models can be implemented for asphaltene stability determination. The detailed results regarding the classification of asphaltene stability have been presented in **Appendix G**.

5.7. Model Development for Wax Disappearance Temperature

In order to develop an accurate and rapid predictive model for the determination of wax disappearance temperature, two input variables have been considered according to the existing thermodynamics methods, namely, molar mass and pressure.

A comprehensive database is first required in order to develop models for the evaluation of WDT. The database is randomly divided into two sets consisting of the "Training" set and the "Test" set. About eighty percent (203 data points) of the main data set was randomly separated for the "Training" set and the rest, twenty percent (51 data points), has been used to test the models developed by means of the ANN, LSSVM, and DT approaches. In this study, several statistical error parameters have been used through a comprehensive error analysis in order to visualize the accuracy and performance capability of the developed models for the determination of wax disappearance temperature. The statistical error parameters implemented in this study are squared correlation coefficient, average absolute percent relative error, average absolute percent relative error, standard deviation errors, and root mean square errors. In addition to the statistical error analysis, a graphical analysis was conducted consisting of a scatter plot (crossplot) and a relative error distribution.

For the ANN approach model, the tanh-axon was used as the transfer function, and Levenberg–Marquardt back propagation was employed for training the model on the basis of an ANN algorithm. As a result, the number of hidden neurons in the hidden layer

should be optimized using a trial and error method. The development of the most accurate ANN model was optimized by using 5 hidden neurons by means of AAPRE, R^2, and RMSE. In other words, the ANN model includes three layers of input, output, and hidden with 5 hidden neuron numbers. In the next stage, the LSSVM algorithm was coupled with an optimization strategy known as coupled simulated annealing (Atiqullah and Rao, 1993; Fabian, 1997; Vasan and Raju, 2009) for obtaining the optimum values of the LSSVM parameters (γ and σ^2). As a result, the values obtained by means of the CSA-LSSVM model for the prediction of wax disappearance temperature are 40.8325 and 2138.4288 for σ^2 and γ, respectively. Finally, the regression DT toolbox available in the MATLAB software was used to develop a predictive model for comparing the predicted values with the other methods.

Table 5. 12 lists the calculated statistical error factors for the models developed on the basis of ANN, LSSVM, and regression DT approaches for the prediction of wax disappearance temperature. Regarding the table, similar results are achieved from the use of the ANN and LSSVM models. The AAPRE and R^2 for ANN and LSSVM models are 0.6% and 0.95, respectively, while the regression DT model has provided better results. The statistical error analysis for the regression DT model shows an AAPRE=0.3 and R^2=0.97. The above results clearly indicate that the regression DT method is the most capable for the prediction of wax disappearance temperature with respect to the database collected in this study. To compare the performance of the models graphically, a scatter diagram as well as a relative error distribution plot were sketched.

Table 5. 12 Statistic error parameters calculated for the models developed in this study.

Method	AAPRE[a] %	APRE[b] %	SD[c]	RMSE[d]	R^2 [e]
ANN Model	0.6	0.017	0.004	2.2	0.95
LSSVM Model	0.6	-0.040	0.004	2.2	0.95
DT Model	0.3	-0.002	0.003	1.5	0.97

Table 5. 13 lists some predicted values of wax disappearance temperature data by the ANN, LSSVM and DT modelling approaches. **Fig. 5. 52** illustrates the experimental

data of WDT against the output values obtained from ANN, LSSVM, and regression DT models. As is clear from the **Fig. 5. 52**, all data corresponding to ANN, LSSVM, and regression DT models are almost placed on the unit slope line (Y=X), revealing that there is acceptable agreement between the model predictions and the experimental data of wax disappearance temperature. Additionally, **Fig. 5. 53** reveals the relative error distribution plots for the ANN, LSSVM, and regression DT models for the prediction of wax disappearance temperature. Regarding the **Fig. 5. 53**, the output values of all models developed in this study have a low scatter around the zero error line, and a small range of error has been observed through the prediction of wax disappearance temperature by using ANN, LSSVM, and regression DT approaches. The results clearly show that intelligent approaches such as ANN, LSSVM, and DT can be applied as an alternative to thermodynamic methods for the evaluation of phase behavior of wax deposition.

Finally, we selected the results obtained by the regression DT model (because they provided the most accurate prediction the WDT) to find the effects of molar mass and pressure on the predicted WDT values. To conduct such a sensitivity analysis, the relevancy factor (r) approach (Chen et al., 2014) is applied for assessing the influence degree of each input variable (*i.e.* molar mass and pressure) on the wax disappearance temperature. The result of the sensitivity analysis performed is indicated in **Fig. 5. 54**. It is clear from the figure that molar mass and pressure have positive effects on the WDT values predicted by the regression DT model. The results indicate that the molar mass has higher impact on the WDT predicted by the regression DT model compared to the pressure variable. The reported relevancy factor for molar mass and pressure are 0.70 and 0.33, respectively.

Table 5. 13 Summary of some correlated wax disappearance temperature values provided by the models developed in this study in comparison with the experimental data.

P(MPa)	M(g/mol)	WDT (K)	ANN	AAPRE	LSSVM	AAPRE2	DT	AAPRE3
0.1	199.4	277.4	276.2	0.4	276.1	0.5	277.8	0.1
0.1	199.75	277.1	276.4	0.3	276.3	0.3	276.6	0.2
0.1	200.1	276.2	276.5	0.1	276.4	0.1	276.6	0.2
0.1	205	278.7	278.6	0.0	278.7	0.0	277.9	0.3
0.1	208.5	280.8	280.2	0.2	280.3	0.2	279.4	0.5
0.1	212	283.2	281.8	0.5	281.9	0.5	282.2	0.4
20	200.1	280.6	281.2	0.2	281.1	0.2	281.1	0.2
20	200.8	280.9	281.5	0.2	281.4	0.2	281.1	0.1
20	201.5	281.2	281.8	0.2	281.7	0.2	281.1	0.0
20	205	283	283.5	0.2	283.4	0.1	282.2	0.3
20	208.5	285.2	285.1	0.0	285.0	0.1	284.7	0.2
40	200.8	285.1	285.7	0.2	286.0	0.3	285.3	0.1
40	201.5	285.3	286.1	0.3	286.3	0.3	285.3	0.0
40	205	287.3	287.8	0.2	288.0	0.2	286.0	0.5
40	208.5	289.5	289.5	0.0	289.6	0.0	288.8	0.2
40	212	291.9	291.2	0.2	291.3	0.2	293.3	0.5
60	200.1	288.8	289.5	0.2	290.0	0.4	289.3	0.2
60	200.8	289.1	289.8	0.2	290.3	0.4	289.3	0.1
60	201.5	289.4	290.1	0.3	290.7	0.4	289.3	0.0
60	205	291.3	291.8	0.2	292.4	0.4	290.0	0.5
60	208.5	293.5	293.6	0.0	294.1	0.2	292.8	0.3
60	212	296.2	295.3	0.3	295.7	0.2	296.9	0.2
80	200.1	292.7	293.5	0.3	294.0	0.5	293.7	0.3
80	200.8	292.9	293.8	0.3	294.4	0.5	293.7	0.3
80	201.5	293.2	294.1	0.3	294.8	0.5	293.7	0.2
80	205	295.1	295.8	0.2	296.5	0.5	293.0	0.7
80	208.5	297.4	297.5	0.0	298.2	0.3	296.6	0.3
100	199.75	298.9	297.2	0.6	297.6	0.4	297.8	0.4
100	200.1	296.3	297.4	0.4	297.8	0.5	297.8	0.5
100	200.8	296.7	297.7	0.3	298.1	0.5	297.8	0.4
100	201.5	296.9	298.0	0.4	298.5	0.5	297.8	0.3
100	205	298.8	299.7	0.3	300.3	0.5	296.6	0.7

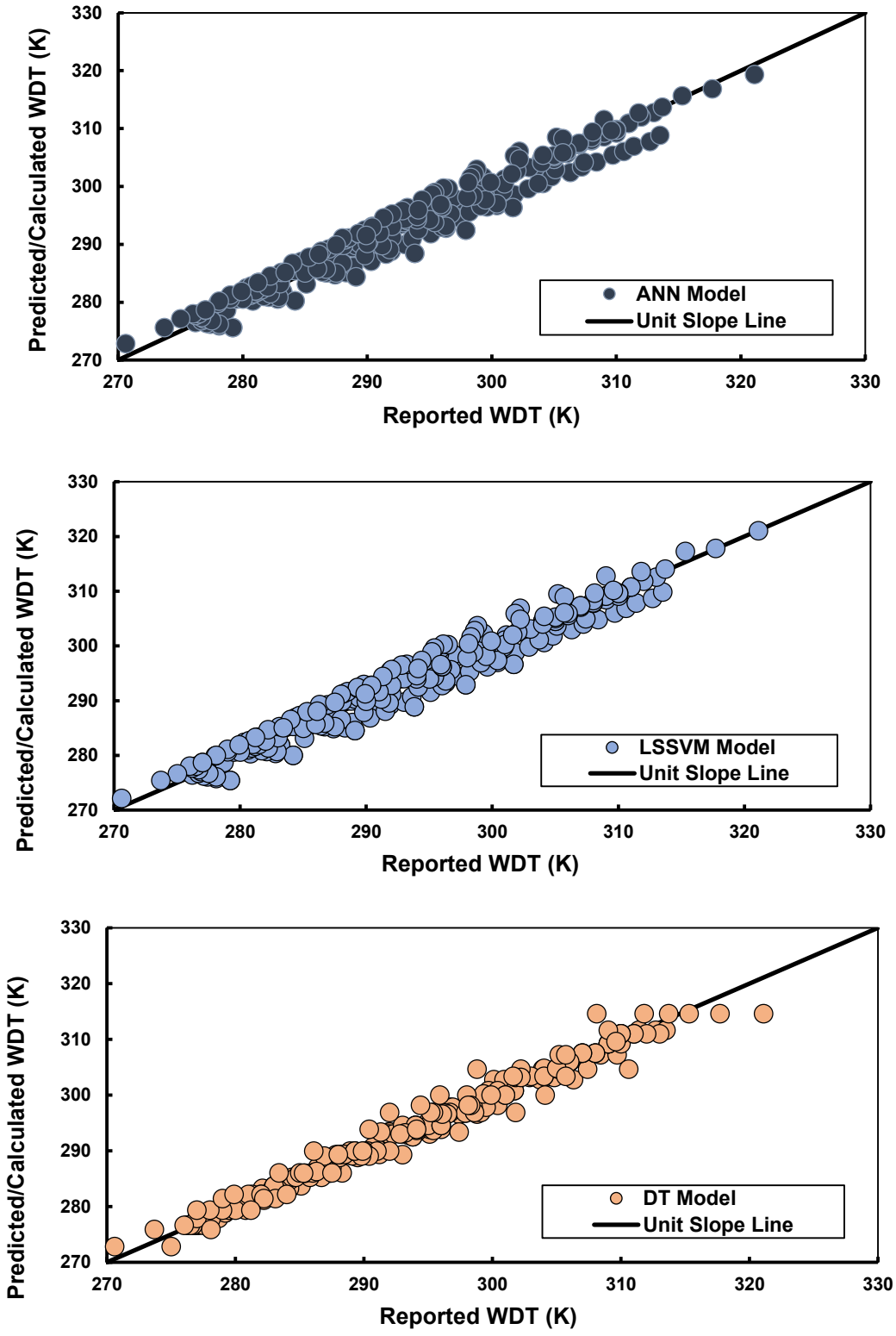

Fig. 5. 52 Crossplots for the models developed in this study.

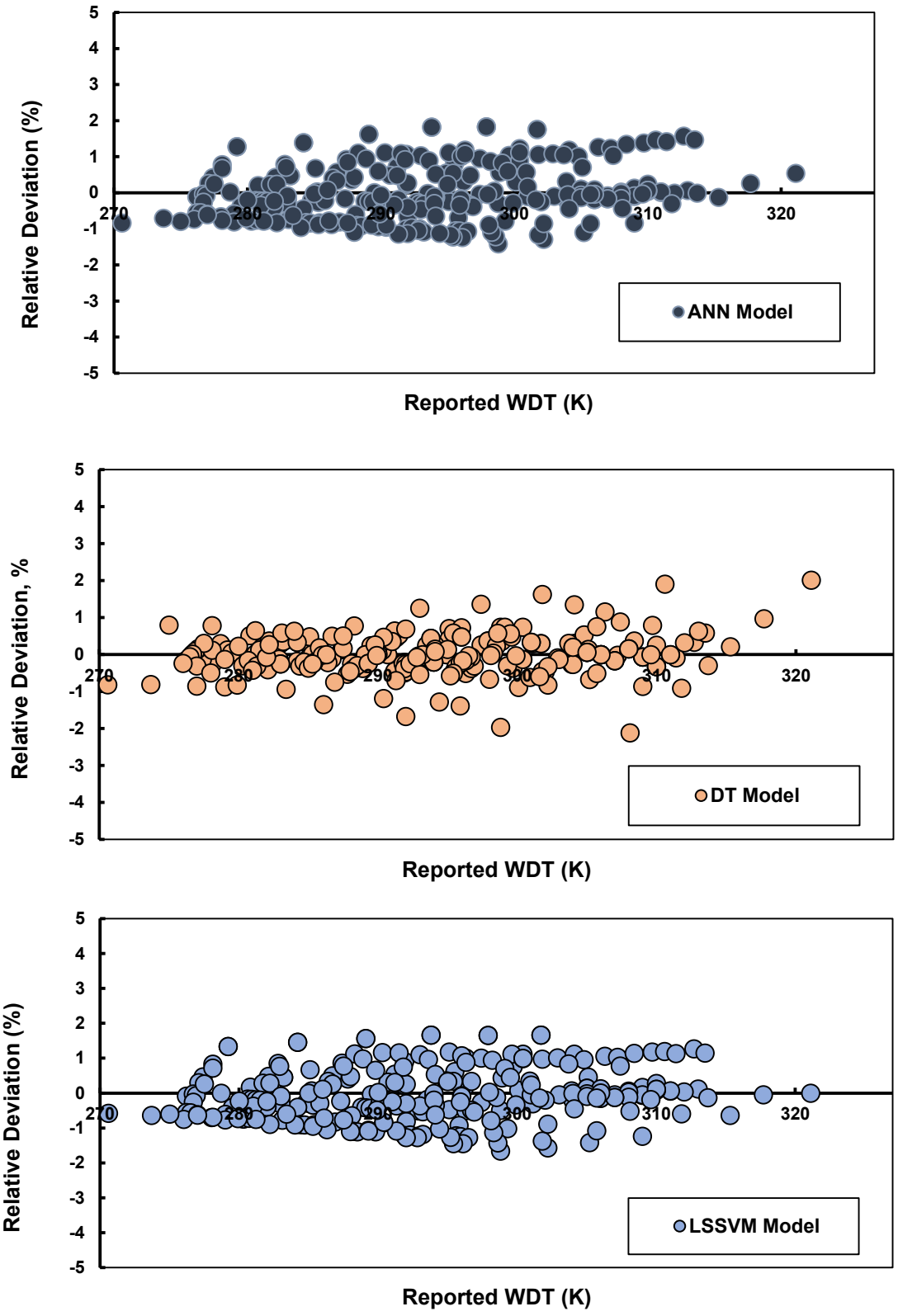

Fig. 5. 53 Relative deviation distribution plots for the models developed in this study.

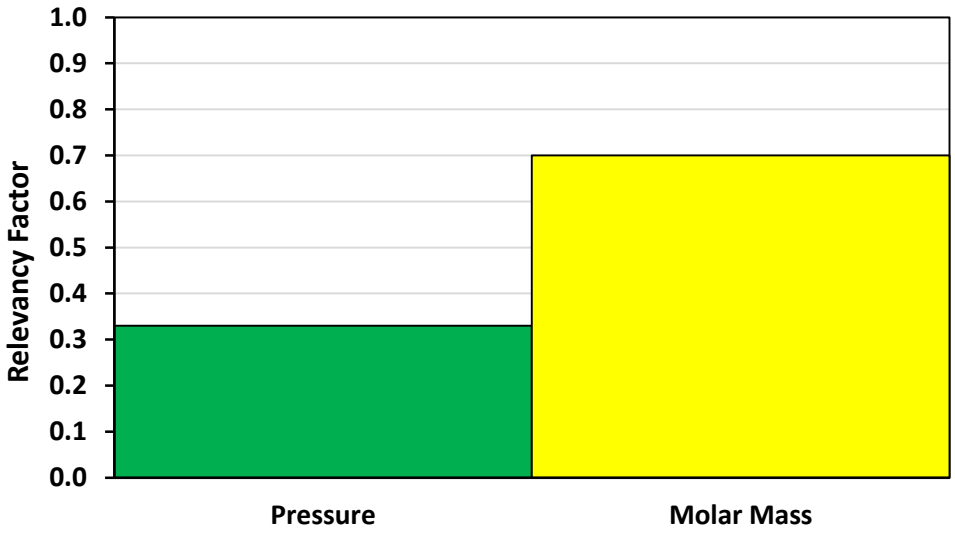

Fig. 5. 54 The result of sensitivity analysis conducted to find the impact of molar mass and pressure variables on the WDT predicted by the regression DT model.

5.8. Intelligent Model Development for Hydrocarbon–plus (C_{7+}) Properties

In order to develop reliable models to characterize the distributed properties of heptane-plus components (ie. molecular weight, specific gravity, and boiling point temperature) four modelling strategies are used, all with the same input variables, *i.e.* bulk molecular weight, bulk specific gravity, and cumulative weight fraction.

In order to develop and test the models, the database collected is randomly divided into two sets consisting of the "Training" set and the "Test" set. About eighty percent of the whole data set was randomly separated for the "Training" set and the twenty percent has been used to test the models developed by the ANN, LSSVM, GEP, and DT approaches.

In this study, two important statistical error parameters have been used to conduct a comprehensive error analysis in order to demonstrate the accuracy and performance capability of the models. The statistical error parameters implemented in this study are the squared correlation coefficient and the average absolute percent relative error. In addition to the statistical error analysis, a graphical analysis is provided in which a scatter plot (crossplot) is performed.

To ensure a reliable ANN model, the tanh-axon was considered as a transfer function, and the Levenberg–Marquardt back propagation was employed for training the model developed on the basis of the ANN algorithm. As a result, the number of hidden neurons in hidden layer should be optimized using a trial and error method. An ANN model was developed using 10 hidden neurons relating to AARD and R^2. In other words, the ANN model includes three layers of input, output, and hidden layers with 10 hidden neuron numbers. In the next stage, the LSSVM algorithm was coupled with an optimization strategy known as coupled simulated annealing (Atiqullah and Rao, 1993; Fabian, 1997; Vasan and Raju, 2009) in order to obtain the optimum values of the LSSVM parameters (γ and σ^2).

The values produced by means of the CSA-LSSVM model for the prediction of the distributed properties of heptane-plus components are σ^2=0.521696 and γ=6020.4 for the boiling point model; σ^2=5.342234 and γ=126813.9 for the specific gravity model; andσ^2= 414.181743 and γ= 1.2 for the molecular weight model. Finally, the regression DT toolbox available in the MATLAB software was used to develop a predictive model for comparing the predicted values with the other methods.

Three GEP-based models for the distributed properties (*ie.* molecular weight, specific gravity, and boiling point temperature), were proposed by applying one gene with 30 chromosome for each model, and an average absolute relative deviation was utilized as the accuracy function. Furthermore, a function set including power, cube root, *, /, – and + is selected during the GEP methodology. The final models obtained by the GEP algorithm developed in this study are simple-to-use with the lowest possible coefficients as follows:

a) Molecular weight.

$$MW = \left({CX_w}^2 + {CX_w}^{\frac{1}{3}}\right)\left(MW_b - \frac{1}{{SG_b}^{12.1871}} + 11.37\right) \tag{5.17}$$

b) Specific gravity.

$$SG = \left(\frac{1}{\left(\frac{17.064\,SG_b}{{CX_w}^2}\right)^{\left(\frac{8.575}{MW_b + SG_b}\right)}}\right) \tag{5.18}$$

c) Boiling point temperature.

$$T_b = MW_b + 3.9123\, CX_w\, MW_b + 396.29 + \frac{CX_w}{SG_b} \quad (5.19)$$

where MW, SG, and T_b shows the distributed properties (*ie.* molecular weight (g/mole), specific gravity, and boiling point temperature (°R), respectively). MW_b expresses the bulk molecular weight (g/mole), SG_b stands for the bulk specific gravity, and CX_w is the cumulative weight fraction.

Table 5. 14 summarizes the calculated statistical error factors for the models developed on the basis of ANN, LSSVM, GEP and regression DT approaches for the prediction of distributed properties (*ie.* molecular weight, specific gravity, and boiling point temperature). It can been seen in the table that the DT modelling is superior to the ANN, LSSVM, and GEP methods for the prediction of all three properties studied. Additionally, the LSSVM model provides a more accurate prediction of boiling point temperature in comparison to the GEP and ANN models.

The results indicate that the specific gravity values characterized by the GEP and LSSVM models are in satisfactory agreement with the actual data, while the ANN model could not predict the specific gravity properly. Furthermore, **Table 5. 14** shows that the ANN, LSSVM and GEP models developed in this study approximate the same performance of heptane-plus data in predicting molecular weight. **Fig. 5. 55** illustrates the AARD % calculated for the prediction of C_{7+} properties (*ie.* T_b, SG, and MW). It clearly demonstrates that the DT model outperforms other methods.

Table 5. 14 Statistical error analysis for the models developed in this study.

Parameter	ANN	DT	LSSVM	GEP
AARD %				
Boiling temperature	3.809	1.64	3.2	4.46
Specific gravity	11.434	0.677	1.6	2.4
Molecular weight	11.85	5.1	10.387	12.7
R-squared				
Boiling temperature	0.8803	0.9779	0.9139	0.8193
Specific gravity	0.2292	0.9724	0.8362	0.7148
Molecular weight	0.8338	0.9567	0.8649	0.8049

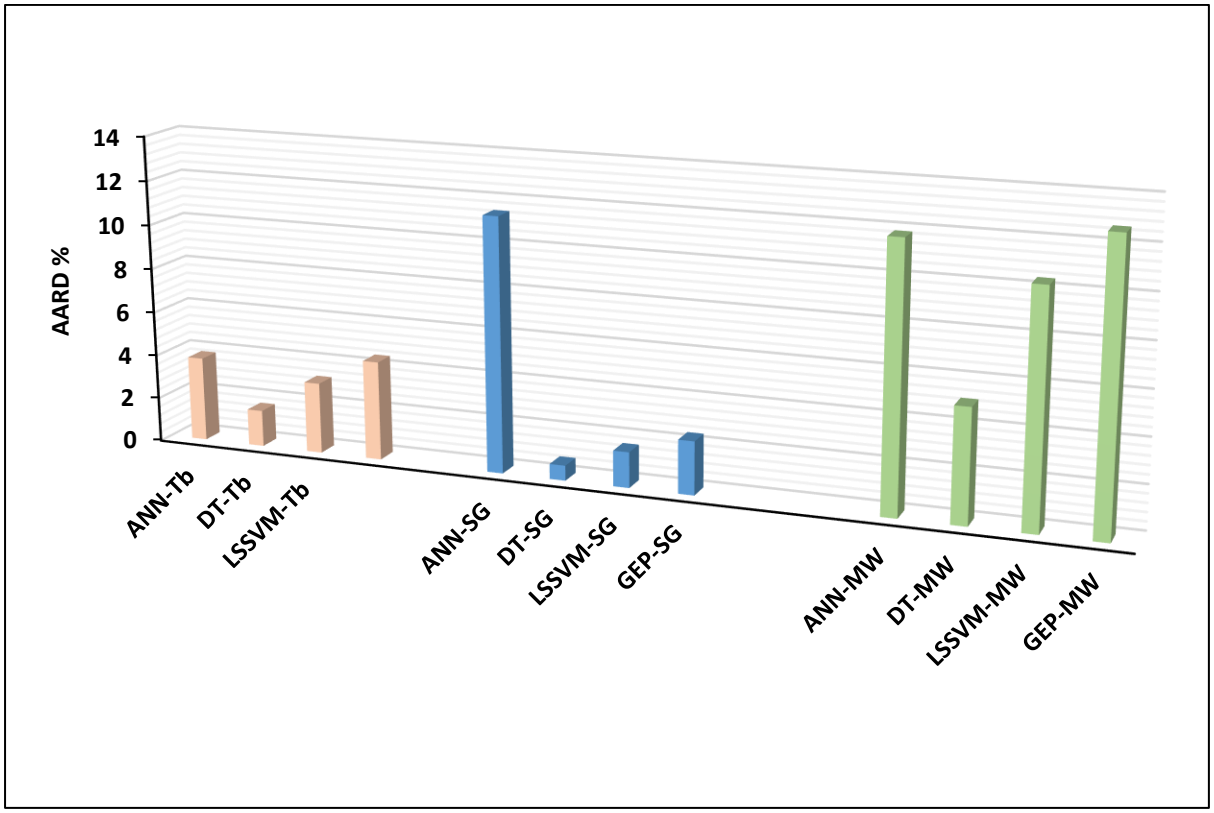

Fig. 5. 55 AARD % calculated graphically for the prediction of C_{7+} properties (*i.e.* T_b, SG, and MW) using the models developed in this study.

For further comparison of the performance of these models, scatter diagrams of the data predicted versus actual data of C_{7+} properties were sketched. **Figs. 5. 56-58** illustrate the output values obtained from ANN, LSSVM, GEP and regression DT models against the experimental data of C_{7+} properties *ie.* T_b, SG, and MW, respectively. As clear from the **Figs. 5. 56-58**, the data corresponding to ANN, LSSVM, GEP and regression DT models cluster around the unit slope line (Y=X), revealing there is acceptable agreement between the model predictions and the experimental data on C_{7+} properties. Additionally, **Fig. 5. 59** reveals the trend plot of boiling point temperature versus the cumulative weight fraction of the different models developed (*ie.* ANN, DT, LSSVM, and GEP) against the actual data. It is clear from the figure, the DT and LSSVM have better smoothness in predicting the C_{7+} properties. The results obtained in the current study demonstrate that all methods developed in this study are applicable for the characterization of C_{7+} properties (*ie.* T_b, SG, and MW) of crude oils (black oil samples), and gas-condensates. Moreover, simple symbolic equations presented in this study are a reliable alternative for

the existing distribution methods for the characterization of the heavier and complex components of crude oils

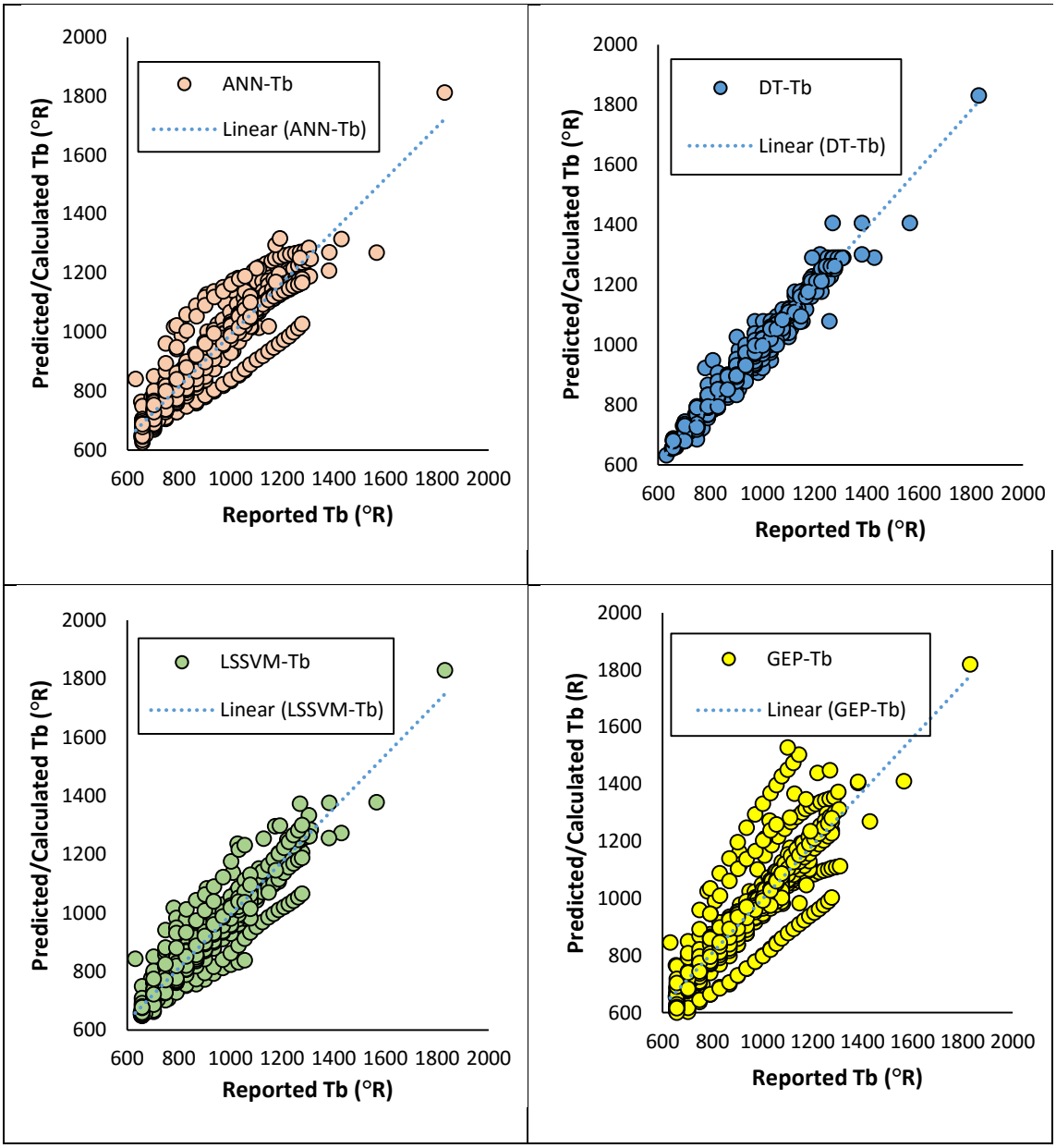

Fig. 5. 56 Graphical comparison (crossplot) between the results obtained by the models developed (*i.e.* ANN, DT, LSSVM, and GEP) and the actual data of T_b.

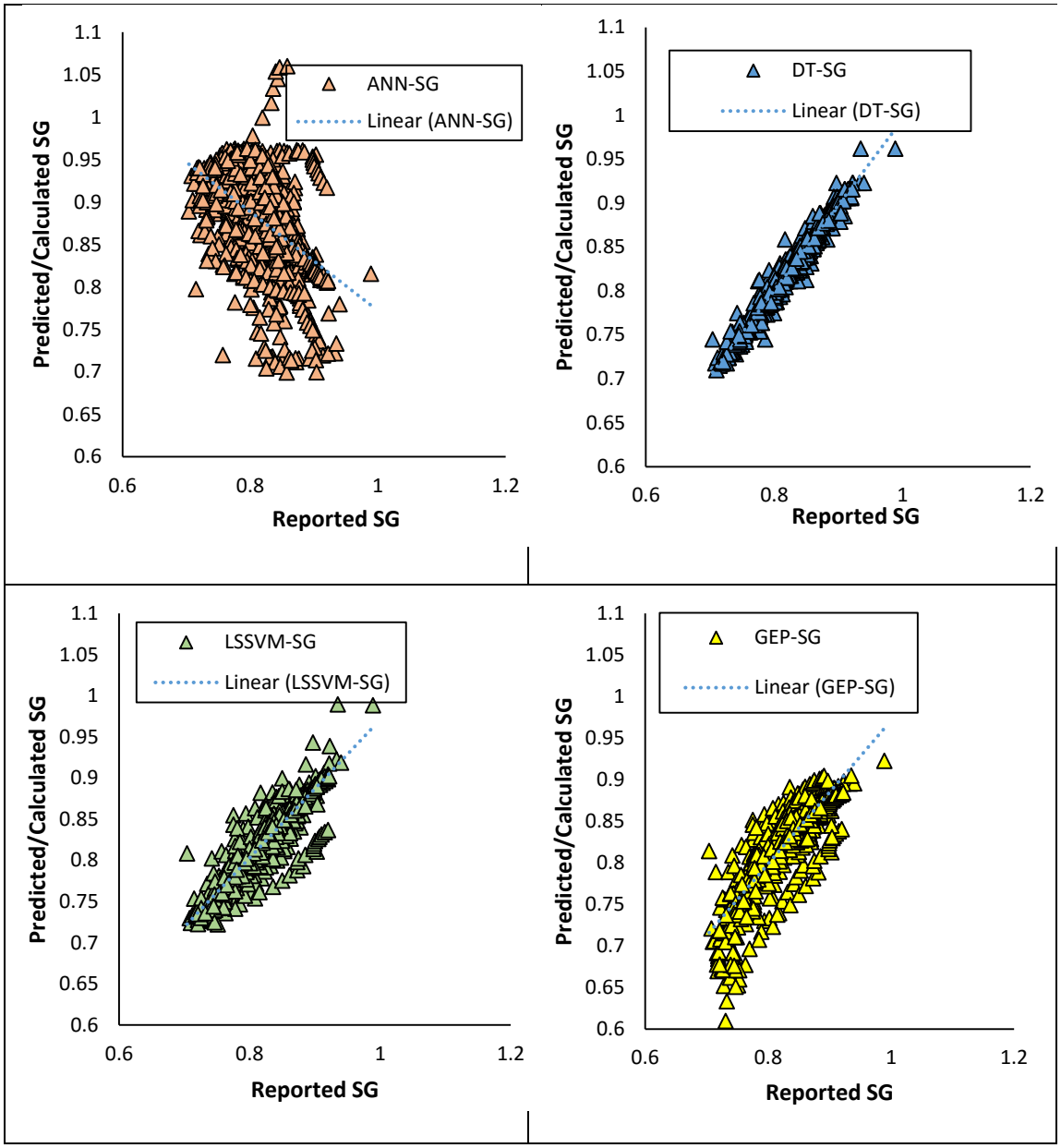

Fig. 5. 57 Graphical comparison (crossplot) between the results obtained by the models developed (*i.e.* ANN, DT, LSSVM, and GEP) and the actual data of SG.

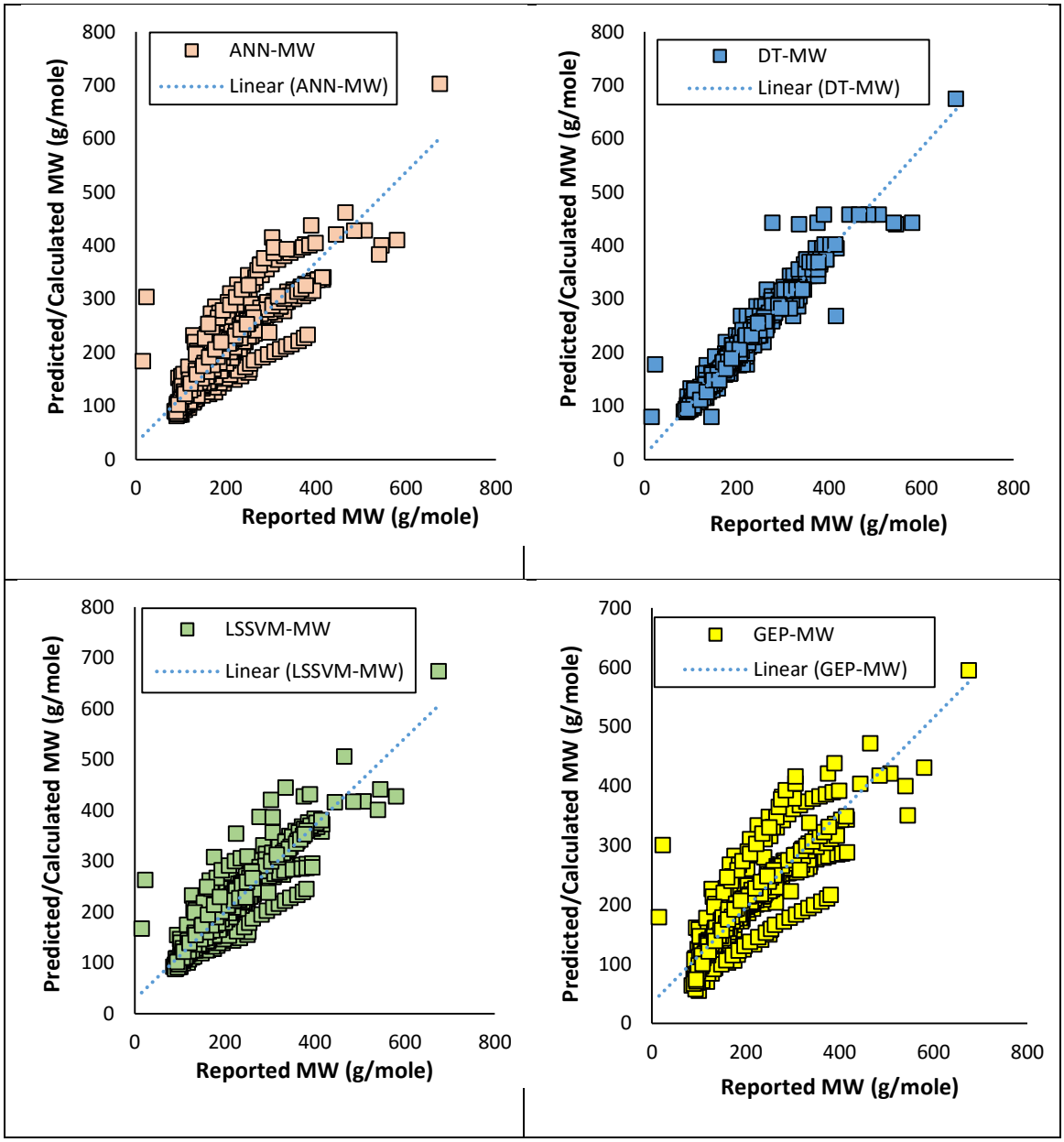

Fig. 5. 58 Graphical comparison (crossplot) between the results obtained by the models developed (*ie.* ANN, DT, LSSVM, and GEP) and the actual data of MW.

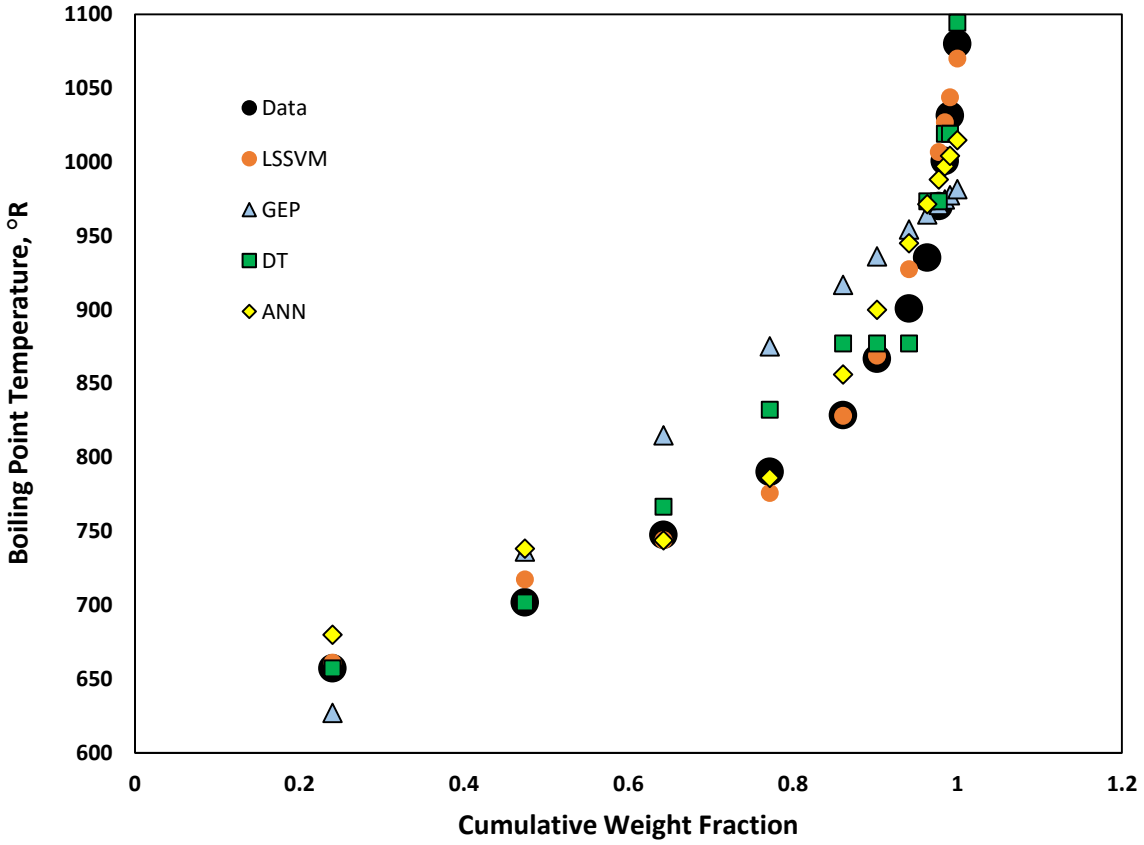

Fig. 5. 59 Trend plot of boiling point temperature versus cumulative weight fraction of the different models developed (*ie.* ANN, DT, LSSVM, and GEP) against the actual data.

5.9. Model Development for Vaporization Enthalpy

5.9.1. Precision Assessment

An assessment of accuracy of the models for testing vaporization enthalpy is required because there is a nonlinear relationship between the vaporization enthalpy and the parameters which affect it, viz. S, M and T_b (Parhizgar et al., 2013). The effect of these variables on the vaporization enthalpy has been shown in **Fig. 5. 60**. Additionally, the calculated R-squared for each input parameter versus the vaporization enthalpy are shown in **Fig. 5. 60**. This figure reveals that the boiling point temperature, as well as the molecular weight, have the highest correlation coefficient, whilst specific gravity has the lowest correlation coefficient. As has been mentioned earlier, the predictive methods developed for prediction/correlation of the vaporization enthalpy are divided into two

classes, viz. those which relate the vaporization enthalpy to the critical properties at the normal boiling point temperature, and those which are a function of the S, M, and T_b. However, the second class, which requires S, M and T_b are more applicable in the petroleum industry (Fang et al., 2003; Vetere, 1995).

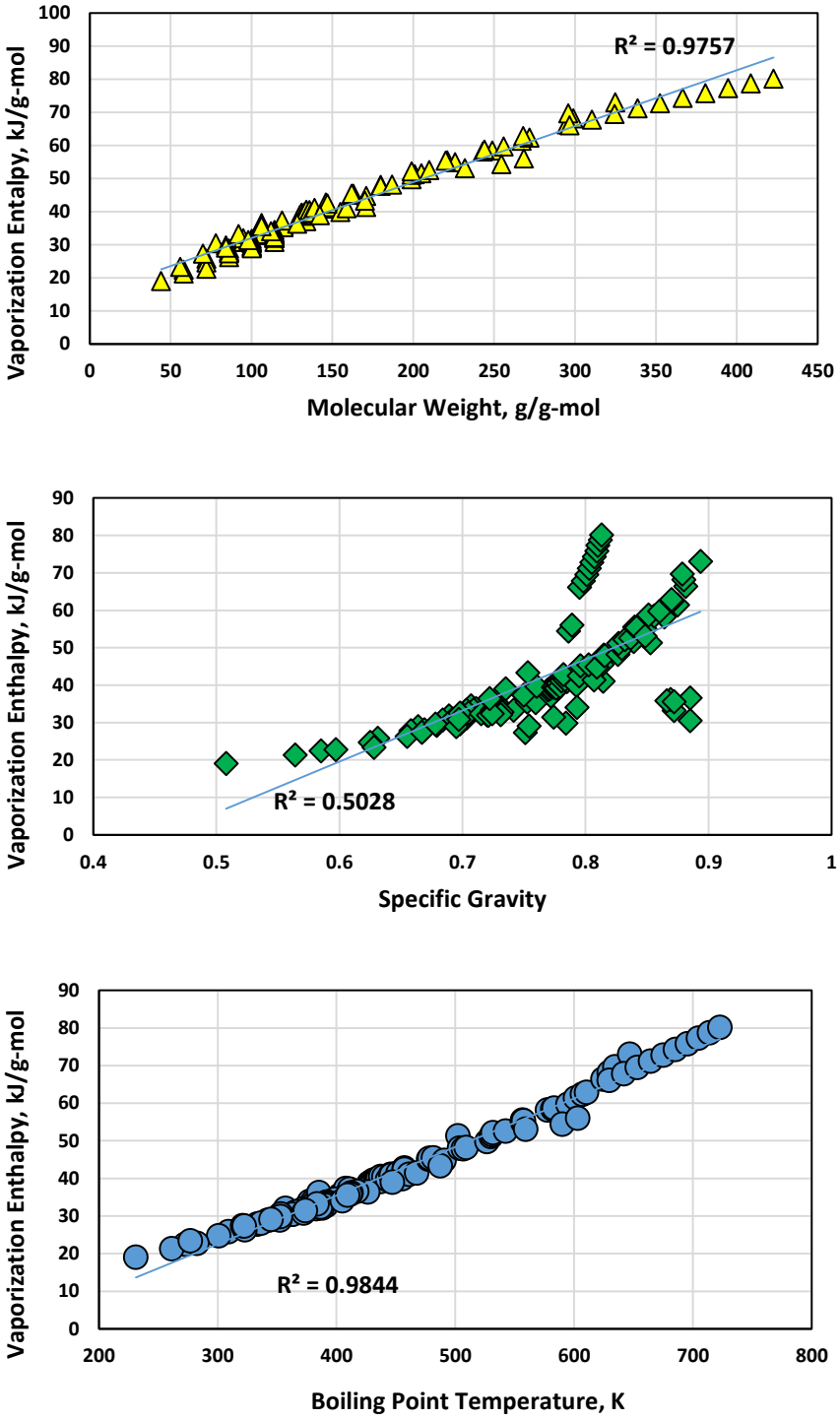

Fig. 5. 60 Vaporization enthalpies of petroleum fractions and pure hydrocarbons as a function of T_b, M, and S.

In view of the above issues, four LSSVM models were developed with various input parameters to achieve the highest accuracy in the determination of the vaporization enthalpies of pure hydrocarbon components and petroleum fractions. To this end, a statistical error analysis, in which standard deviation, average absolute relative deviation (AARD or E_a), average relative deviation (E_r), root mean square error, and R-squared error (R^2) are employed. A graphic error analysis is also provided, in which a crossplot is sketched. More details about statistical and graphical error analyses can be found elsewhere (Hemmati-Sarapardeh et al., 2014a; Kamari et al., 2013b; Montgomery, 2008; Shafiei et al., 2013; Zendehboudi et al., 2011; Zendehboudi et al., 2009).

Table 5. 15 lists the statistical error parameters for the four LSSVMs proposed with the various input parameters. Furthermore, this table summarizes the optimized LSSVM parameters for these four models developed with various input parameters. As has been pointed out earlier, an efficient tuning technique is required to optimize the parameters of the model during the calculations for predicting the vaporization enthalpies of pure hydrocarbon components, as well as for petroleum fractions. The optimum values of the two parameters of the LSSVM model developed in this study, σ^2 and γ, were optimized and determined by implementing the coupled simulated annealing (CSA) (Atiqullah and Rao, 1993; Fabian, 1997; Vasan and Raju, 2009) optimization strategy. As seen from **Table 5. 15**, the LSSVM model developed for three input parameters which include S, M, and T_b have the highest accuracy compared to the other models. The total AARD and R^2 for this model is 1.15514% and 0.9982, respectively. For the training and testing phases, the AARD and R^2 values are 1.08062% and 1.45939%, and 0.9983 and 0.9973, respectively. **Fig. 5. 61** shows the structure of the final LSSVM model developed with its corresponding inputs and output.

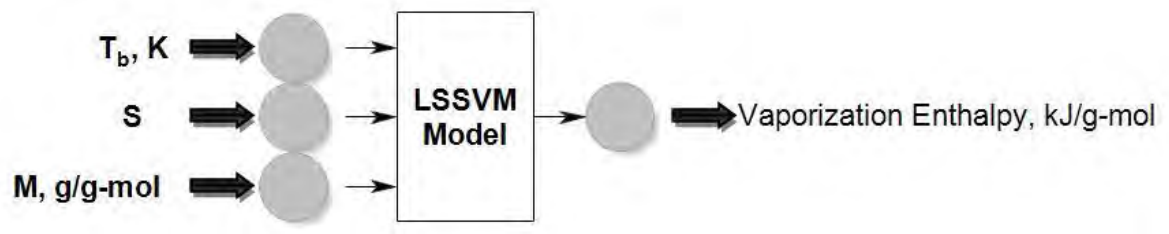

Fig. 5. 61 Structure of the LSSVM model developed in this study with three input variables of T_b, M, and S.

Table 5. 15 The error parameters calculated for the model in order to determine the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

Model No.	Inputs Parameters	γ	σ^2	E_a[a] %	E_r[b] %	SD[c]	RMSE[d]	R^2 [e]
1	S and M	214.330	24.938	2.35	-0.268	0.012	1.23	0.9927
2	T_b and M	137.759	10.452	1.96	-0.209	0.010	1.23	0.9927
3	T_b and S	4699824.055	2275.917	1.85	-0.267	0.009	1.08	0.9944
4	T_b, S and M	114.161	5.111	1.15	0.063	0.005	0.62	0.9982

Fig. 5. 62 shows a point by point comparison between values obtained with the LSSVM model and that from literature for the vaporization enthalpy. As can be seen in this figure, there is satisfactory agreement between the points. A crossplot for the estimated vaporization enthalpy, versus reported values from literature, for both the training and test sub-sets of data is shown in **Fig. 5. 63**. A tight grouping of points around the 45˚ line for the training and test datasets illustrates the capability of the new LSSVM model for forecasting the vaporization enthalpies of pure hydrocarbon components and petroleum fractions. Furthermore, **Fig. 5. 64** shows the absolute relative deviation distributions of the suggested LSSVM model for forecasting the vaporization enthalpies of petroleum fractions and pure hydrocarbon components. The figure clearly illustrates that the LSSVM model developed in this study has low scatter/distribution around the zero error. From **Figs. 5. 62-64** it can be concluded that there is excellent agreement between the forecasted values by the LSSVM approach and literature data on the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

The capability of the LSSVM model for forecasting the vaporization enthalpies was compared with results from some widely-utilized empirical correlations, including Fang et al. (Fang et al., 2003), Mohammadi and Richon (Mohammadi and Richon, 2007), as well as Parhizgar et al. (Parhizgar et al., 2013) and an intelligent method, namely ANN modeling (Mohammadi and Richon, 2007). The bar plots in **Fig. 5. 65** show the average

absolute percent relative deviations of the vaporization enthalpy for the suggested LSSVM model, ANN modeling and the empirically correlations. In **Fig. 5. 65**, the results display that the LSSVM model has a higher accuracy compared to other methods investigated. It is worth mentioning that the CSA-LSSVM model has been developed on the basis of only "two adjustable parameters" while other methods investigated like ANN methodology require more adjustable parameters for forecasting the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

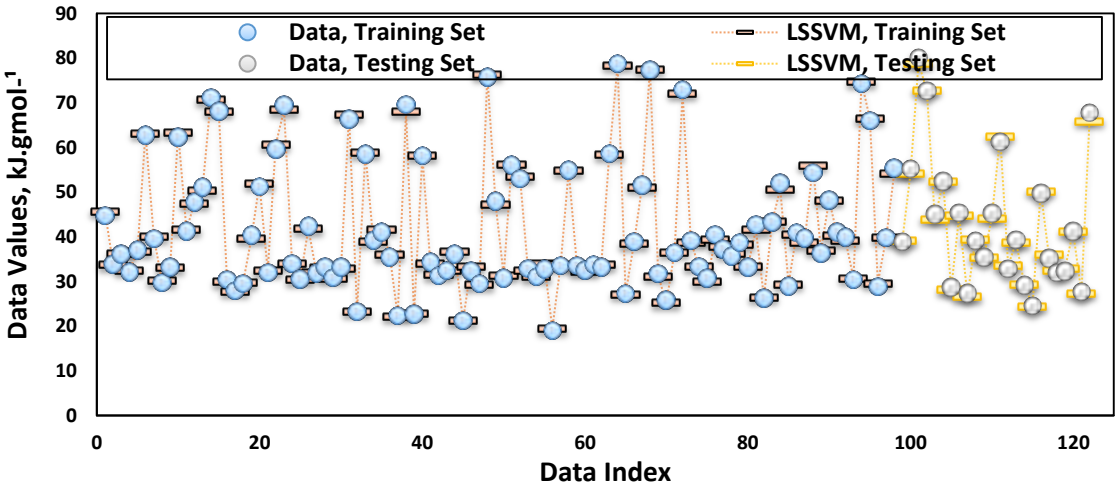

Fig. 5. 62 Point by point comparison between results of the LSSVM model proposed and the actual data for the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

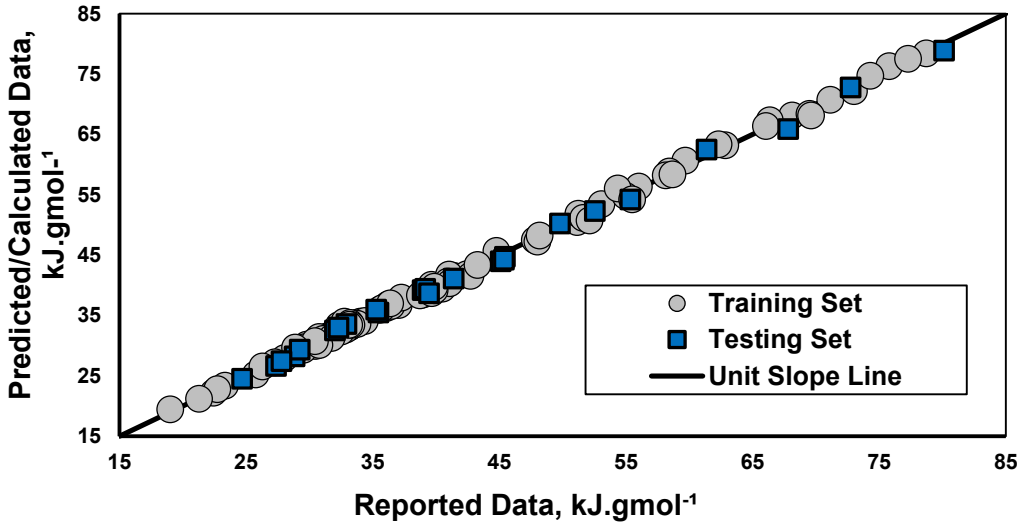

Fig. 5. 63 Scatter diagram comparing the results of the developed LSSVM model and the actual values of vaporization enthalpies of petroleum fractions and pure hydrocarbons.

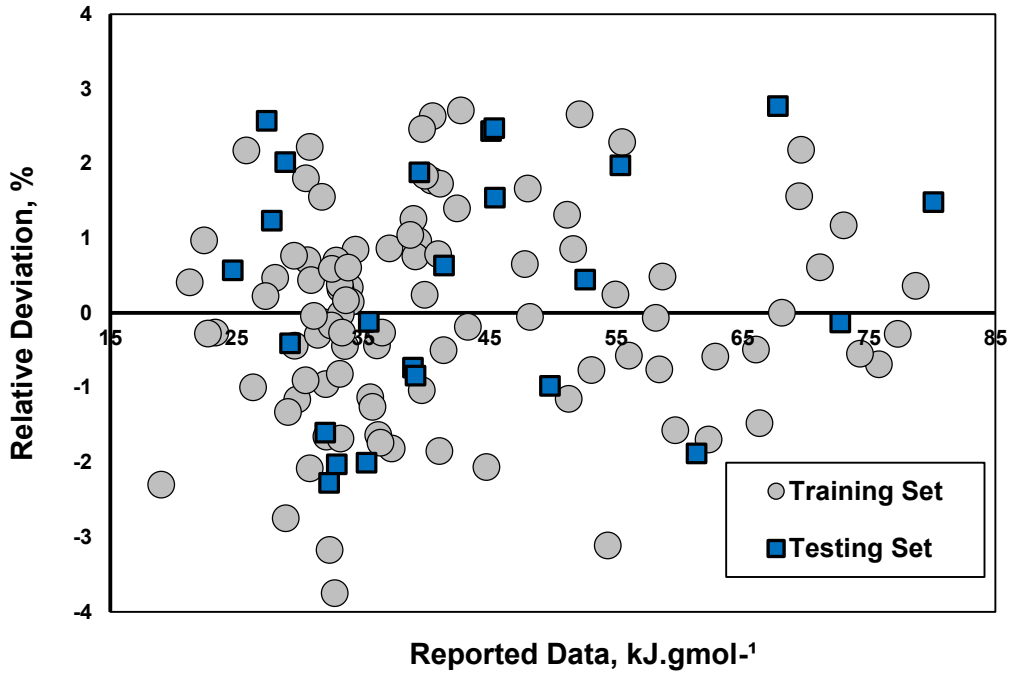

Fig. 5. 64 Distribution of relative deviation of the estimated vaporization enthalpies of petroleum fractions and pure hydrocarbons using the LSSVM model.

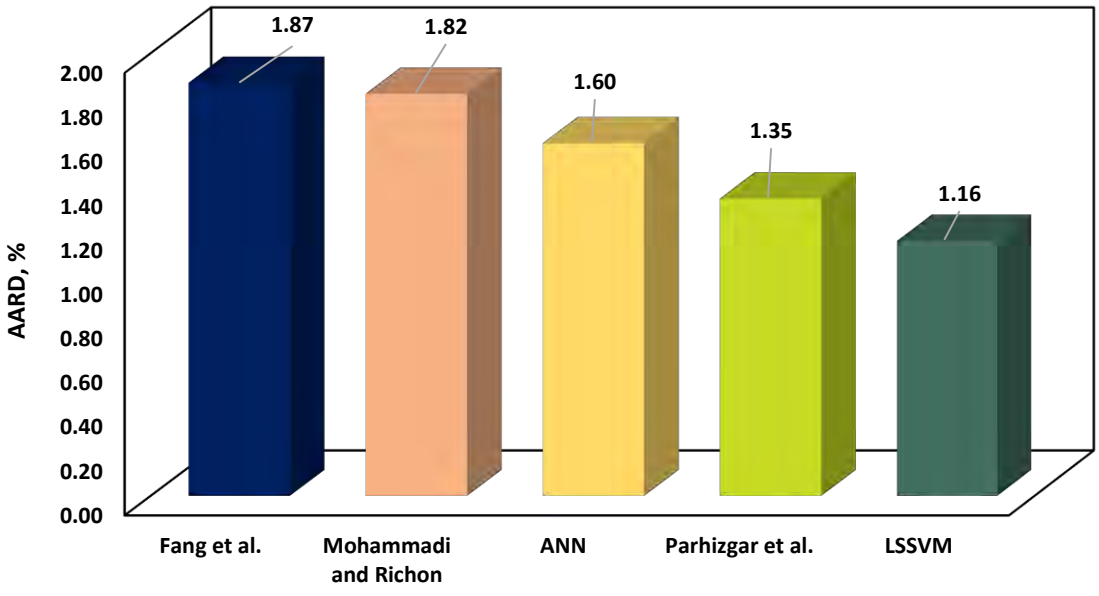

Fig. 5. 65 Calculated average absolute percent relative deviation for the empirically derived methods, ANN methodology, as well as the LSSVM model proposed.

From the results obtained in this study, it can be ascertained that the LSSVM strategy displays superior capability for modeling vaporization enthalpies of petroleum fractions than traditional neural networks and thus makes it a more attractive choice for the prediction of vaporization enthalpy of petroleum fractions. The ANN methodology does not perform well using small datasets due to its high number of adjustable parameters. Furthermore, the ANN-based models have an over-fitting problem and may lead to errors in the outer range of data. This indicates that they are probably not appropriate for extrapolation. Furthermore, the many more adjustable parameters are required to develop an ANN-based model than for the LSSVM methodology. The LSSVM methodology has only two adjustable parameters. Normally, a higher number of adjustable parameters causes an over-fitting problem, and decreases the capability and reliability of the modeling approach in solving nonlinear problems.

The advantage of the LSSVM approach is that it does not need to use a large number of data points in order to optimize and achieve the best (optimal) condition for prediction. Other local regression methods such as neural networks have poor performance for prediction in the presence of a small dataset size, as pointed out earlier. Additionally, prior determination of the network topology is not required in the LSSVM approach and can be automatically determined in the training process. Moreover, the number of hidden nodes and hidden layers does not need to be determined in the CSA-LSSVM model. Furthermore, this model has fewer adjustable parameters (typically two) compared to ANN methods. Additionally, the empirical correlations normally require an initial value, longer computations, and considerable effort to determine a proper relationship for fitting the literature/experimental values.

However, despite its attractive benefits, the LSSVM approach has some potential disadvantages: 1) every data point of an existing database contributes to the model developed and the relative importance of a data point is given by its support value; 2) the second problem is that it is well known that the utilization of a sum squared error cost function without regularization might lead to predictions which are less robust (Suykens et al., 2002a). The LSSVM model developed in this study can be integrated with commercial software applicable available in the industry in order to improve its accuracy and reliability, while decreasing the uncertainties accompanying these applications.

5.9.2. Outlier Diagnosis

It is important to be able to identify the outlier points existing in a databank during the development of a model for forecasting the vaporization enthalpies of pure hydrocarbon components and petroleum fractions because the results of a regression analysis may be biased by such data points. Recognizing the outlier data point(s) is a critical step in the development of a mathematical model and in the assessment of the applicability of a new model (Esfahani et al., 2015; Gharagheizi et al., 2012a; Mohammadi et al., 2012c). Determining the outliers is required in order to recognize individual datum (or groups of data) that may differ from the bulk of the data existing in a database (Gharagheizi et al., 2012a; Gramatica, 2007; Mohammadi et al., 2012c; Rousseeuw and Leroy, 2005). The Leverage method (Goodall, 1993; Gramatica, 2007; Shateri et al., 2014) was implemented in this study. The graphical determination of the suspect data or outliers is carried out by means of sketching the Williams plot based on the calculated H values (Gharagheizi et al., 2012a; Mohammadi et al., 2012c). A detailed explanation of the procedure and formulas for the Leverage strategy have been reported elsewhere (Gharagheizi et al., 2012a; Mohammadi et al., 2012c).

The Williams plot is sketched in **Fig. 5. 66** for the predictions of the vaporization enthalpies of petroleum fractions and pure hydrocarbon components using the LSSVM model suggested in the current study. The results indicate that the majority of data points are in the ranges of $0 \leq H \leq 0.09836$ and $-3 \leq R \leq 3$ which confirms the efficiency and credibility of the LSSVM model. As is clear from **Fig. 5. 66**, that there is only one data point which is outside of the applicability domain of the model that may be an outlier, from the available vaporization enthalpy database.

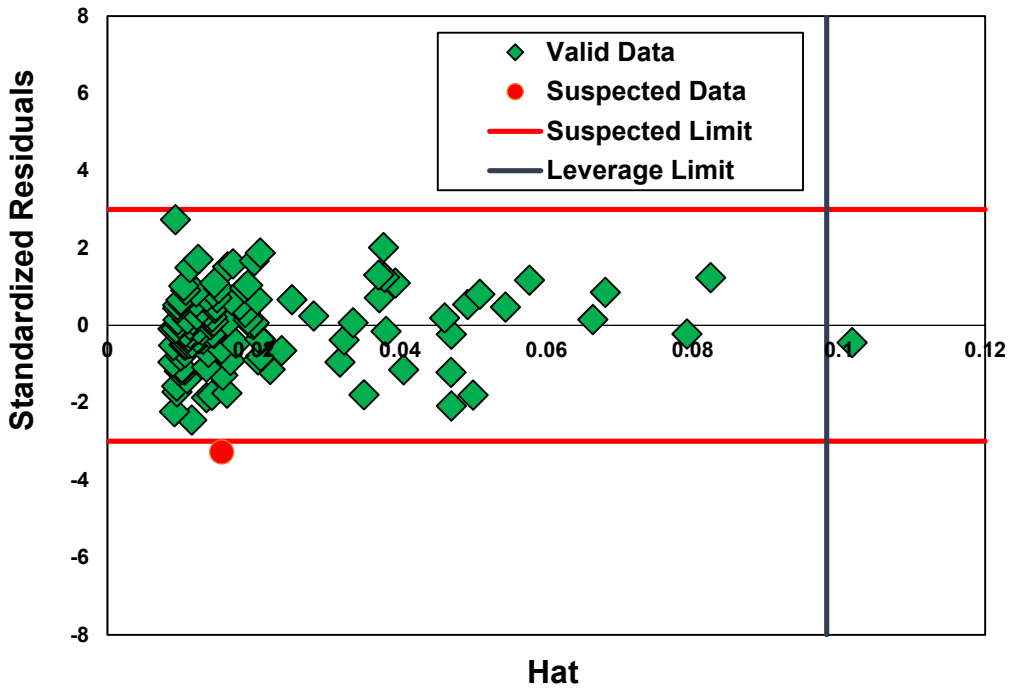

Fig. 5. 66 The Leverage approach illustrating the suspected data existing in the collected database of the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

5.10. Model Development for Gasoline Properties

In this study, an efficient LSSVM model was developed for the estimation of gasoline properties (SG, MON, RON and RVP). In the selection of a proper set of regularization parameters, Kernel parameters such as σ^2 for RBF function and γ, are important in obtaining an improved LSSVM result which in turn improves the success of the LSSVM mathematical strategy. Hence, these two parameters are optimized by means of a reliable external technique. The optimum values of σ^2 and γ were determined and calculated using the coupled simulated annealing (CSA) (Atiqullah and Rao, 1993; Fabian, 1997; Vasan and Raju, 2009) optimization strategy. The optimum values of σ^2 and γ for the LSSVM model developed for prediction of the gasoline properties (SG, MON, RON and RVP) are listed in **Table 5. 16**.

The performance of the LSSVM models proposed for calculating the SG, MON, RON and RVP in terms of statistical model validation parameters is given in **Table 5. 16**. As can be seen in this table, the R^2s for the testing phase for SG, MON, RON and RVP are 0.994, 0.928, 0.919 and 0.828, respectively. The crossplots of the literature-reported SG, MON,

RON, and RVP data against the predicted values from the LSSVM model, for both training and testing stages, are shown in **Fig. 5. 67**. Furthermore, the error percentage distributions between the predicted and literature-reported values of SG, MON, RON, and RVP are shown in **Fig. 5. 68**. The graphical plots confirm that there is a good fit between the literature-reported data and the predictions of the LSSVM model for SG.

Recently, Albahri (Albahri, 2014) developed a predictive model on the basis of the neural network approach for determining gasoline properties using the same databank. Their results indicated acceptable accuracy with a R^2=0.99 and a maximum deviation of 1.71 for SG, R^2=0.99 and a maximum deviation of 0.13 for RVP, R^2=0.99 and a maximum deviation of 8.25 for RON, and R^2=0.99 and a maximum deviation of 11.3 for MON. The capability and applicability of any mathematical predictive model depends on the complexity of the model and the number of adjustable parameters, in addition to accuracy.

As a comparison, the model developed in the present study is satisfactory in terms of time to develop, speed of the computation, ease of development, applicability and capability. Additionally, it should be noted that the LSSVM model developed in this study for prediction of gasoline properties (SG, MON, RON and RVP) has only "two adjustable parameters" (σ^2 and γ), while other intelligent methods such as neural networks require many more adjustable parameters for their estimation targets. The model developed in this study has additional advantages over classical techniques, advantages such as convergence to the global optimum, does not require the optimization and evaluation of the number of neurons in hidden layers (in other words finding the topology of networks); has very low possibility for over and under-fitting problems; and does not need large databases (Cristianini and Shawe-Taylor, 2000; Gharagheizi, 2007; Suykens and Vandewalle, 1999).

Table 5. 16 The results obtained using the LSSVM model for SG, MON, RON, and RVP.

Performance	E_a%	E_r%	SD	RMSE	R^2
SG model (γ: 740.4370534; σ²: 114.4808574)					
Total	0.62	0.009	0.0044	0.007	0.990
Training	0.62	-0.011	0.004	0.007	0.989
Testing	0.61	0.090	0.001	0.005	0.994
MON model (γ: 20.5295675; σ²: 3.8287799)					
Total	2.51	-0.332	0.018	2.780	0.933
Training	2.41	-0.274	0.016	2.730	0.936
Testing	2.90	-0.56	0.008	2.966	0.928
RON model (γ: 30.9073519; σ²: 3.7480896)					
Total	2.39	-0.088	0.018	2.929	0.955
Training	2.05	-0.238	0.014	2.515	0.967
Testing	3.74	0.504	0.011	4.182	0.919
RVP model (γ: 9.0376950; σ²: 6.1108949)					
Total	15.29	-5.411	0.176	0.092	0.92
Training	13.79	-4.731	0.137	0.089	0.932
Testing	21.23	-8.12	0.110	0.101	0.828

In order to develop an accurate predictive model, one needs a collection of the most reliable data points, from which one has detected and excluded suspect or unreliable data. Moreover, there is also a need to assess the available literature-reported data for the gasoline properties (SG, MON, RON and RVP) because large uncertainties will influence the estimation capability of the model. Hence, the Leverage value statistics technique (Goodall, 1993; Gramatica, 2007) is used in this study to detect outliers data points. Graphical detection of the outliers or suspect data is determined using the Williams plot on the basis of the calculated H value (Mohammadi et al., 2012a; Mohammadi et al., 2012c). Details on the Williams plot and the related mathematical

equations and calculation procedure are available in the literature (Mohammadi et al., 2012a; Mohammadi et al., 2012c).

The Williams plot for the SG model, the MON model, the RON and the RVP model are shown in **Fig. 5. 69**. As can be seen from the existence of the majority of data points in the ranges $0 \leq H \leq 0.1685$ and $-3 \leq R \leq 3$ for the SG model, $0 \leq H \leq 0.2022$ and $-3 \leq R \leq 3$ for the MON model, $0 \leq H \leq 0.2022$ and $-3 \leq R \leq 3$ for the RON model and $0 \leq H \leq 0.0668$ and $-3 \leq R \leq 3$ for the RVP model, the mathematical method proposed in this study is statistically satisfactory and useable. As a result, high good domain leverage points are placed in the area of $0.1685 < H$ for the SG model, $0.2022 < H$ for the MON and RON models, and $0.0668 < H$ for the RVP model. **Fig. 5. 69** shows that only three data points for the SG model, four data points for the MON and RON models, and six data points for the RVP model are located in the bad Leverage domain or region. Here it should be mentioned that some data points in **Fig. 5. 69** are located in the high leverage zone which does not affect the overall performance of the LSSVM model.

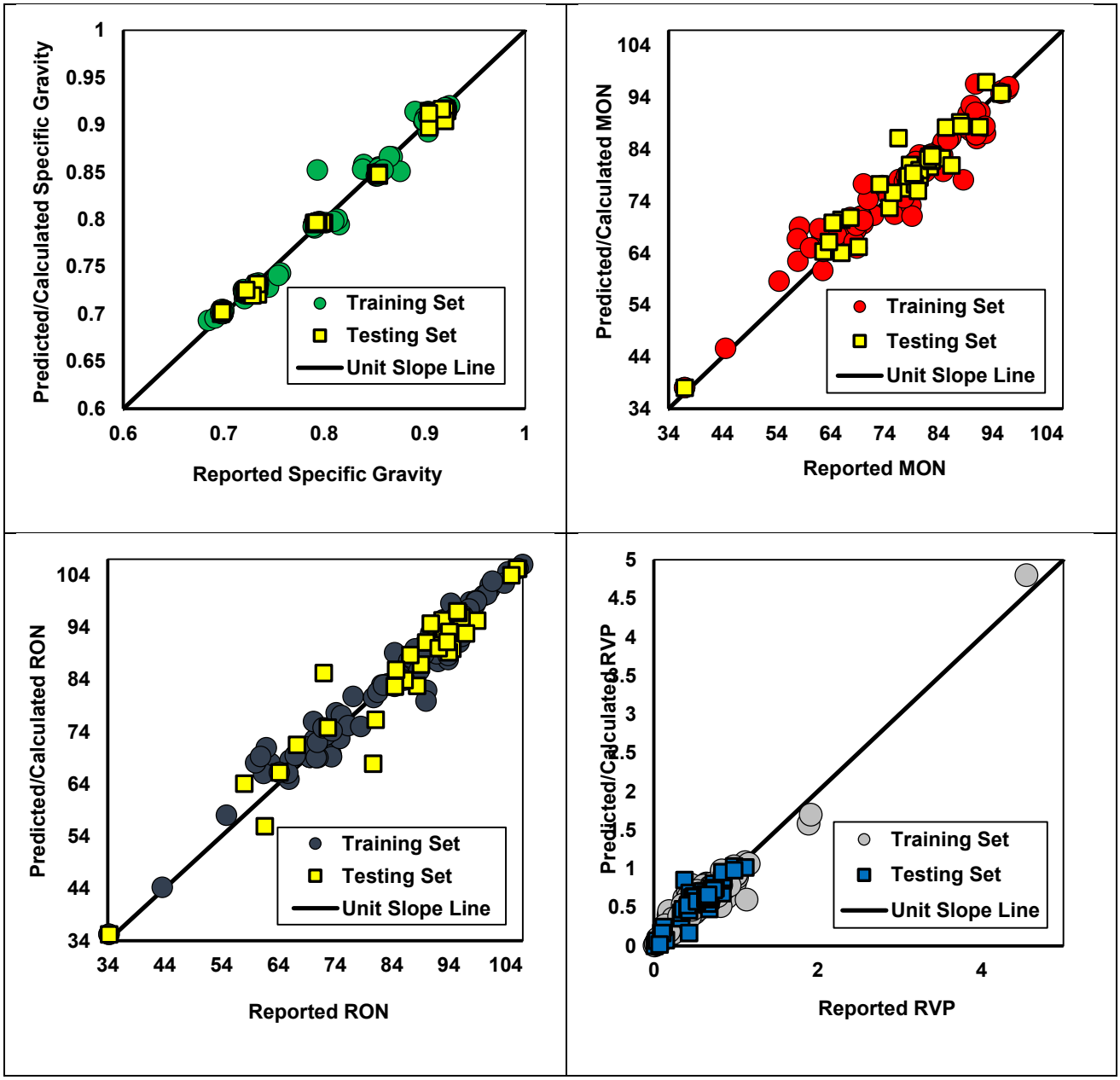

Fig. 5. 67 Comparison of the estimated and literature values of SG, MON, RON, and RVP.

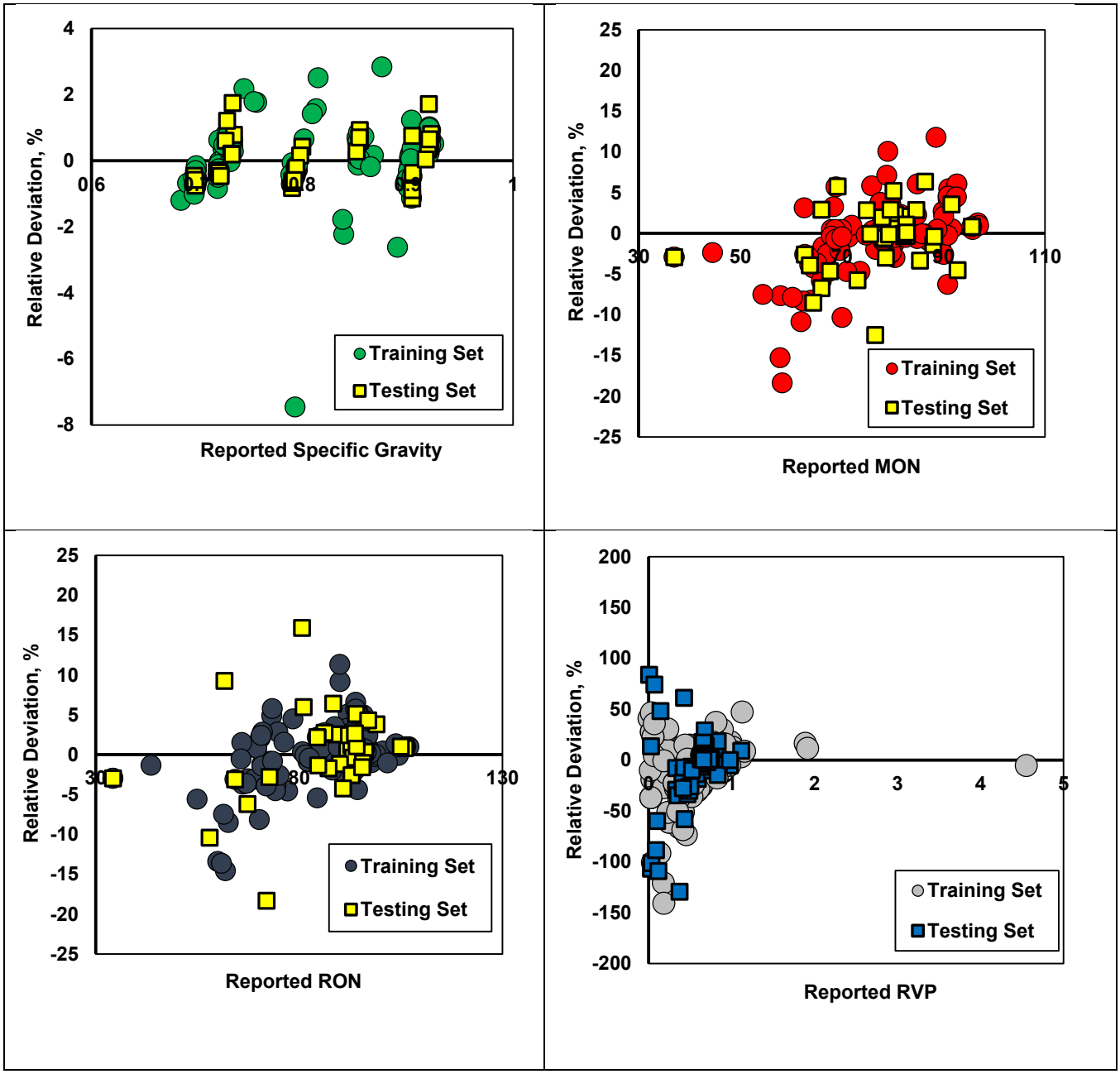

Fig. 5. 68 Comparison of estimated and literature values of SG, MON, RON, and RVP with respect to the residual error percentage.

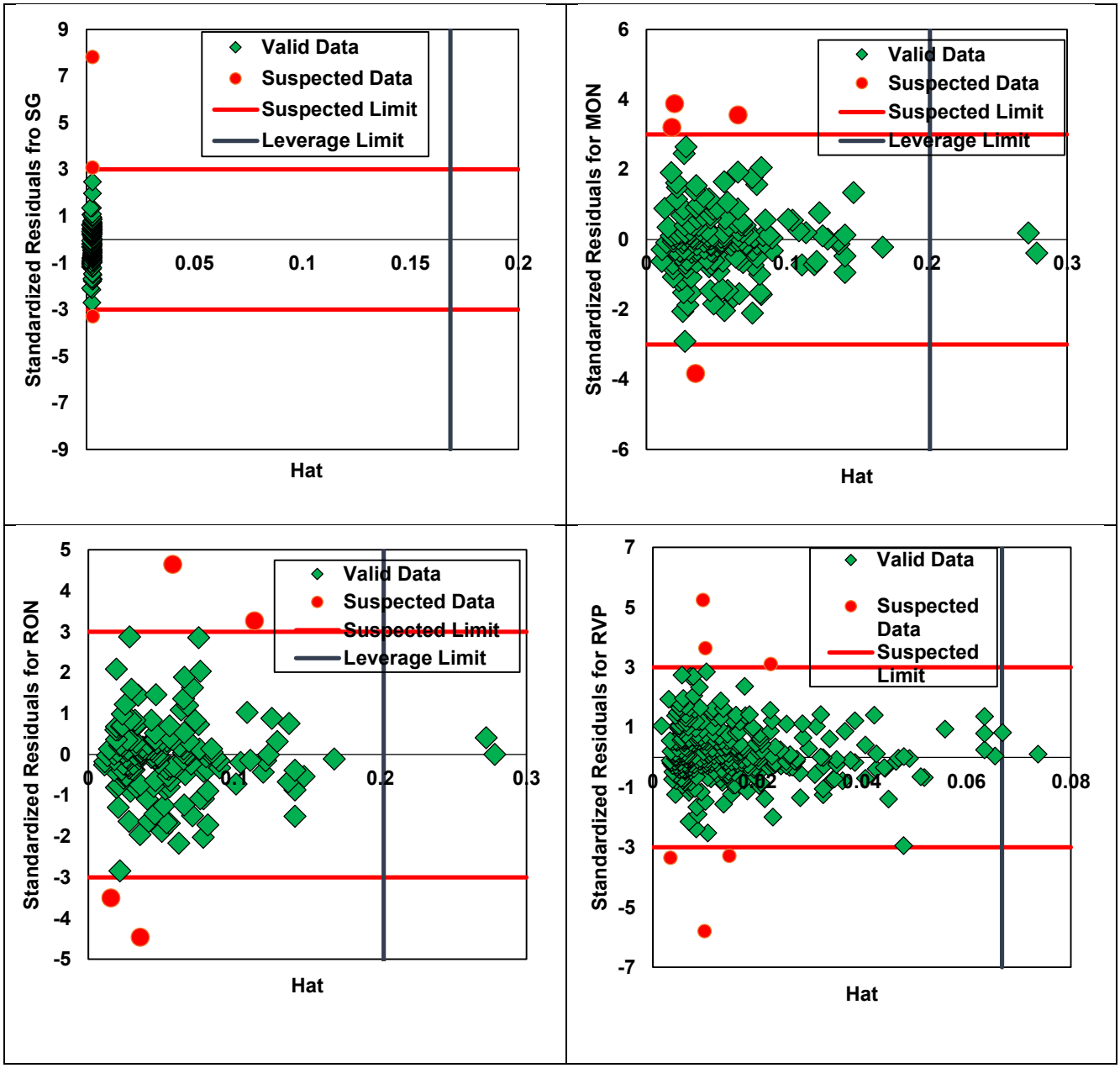

Fig. 5. 69 Identification of doubtful data for SG, MON, RON, and RVP in LSSVM modeling using Hat values.

5.11. New Equation for Gas Compressibility Factor

The computational steps as described above were followed to achieve an efficient, reliable, and capable GEP model for the prediction of *z*-factor. Moreover, as previously mentioned, for assessing the capability and performance of the GEP model, a statistical error analysis, in which AAPRE, APRE, SD, RMSE and R^2, as well as a graphical error analysis, in which a parity diagram and APRE error distribution plot are sketched, have been implemented. It is thereby shown that the GEP (Ferreira, 2006) approach calculations express the required parameters, which yield the most precise model from the introduced parameters (P_{pr} and T_{pr}). Hence, one can consider several independent parameters for a particular problem and find the ones which have the most positive impacts on the desired output results. The ultimate form of *z*-factor equation obtained can be expressed as follows:

$$Z = 0.2625136 + \frac{3.1263651}{T_{pr}} + \frac{-3.8916368}{T_{pr}^{\,2}} + \frac{1.0551763}{T_{pr}^{\,3}} + 0.5638878\ln(P_{pr}) - 0.3372525\ln(P_{pr})^2$$
$$+0.061688\ln(P_{pr})^3 + \frac{-1.3976452\ln(P_{pr})}{T_{pr}} + \frac{0.5217521\ln(P_{pr})}{T_{pr}^{\,2}} + \frac{0.447935\ln(P_{pr})^2}{T_{pr}} \quad (5.20)$$

where T_{pr} is pseudo-reduced temperature and, P_{pr} denotes pseudo-reduced pressure.

To obtain the equation above, the number of significant digits for the coefficients has been calculated by conducting sensitivity analysis of the predicted results in relation to the actual values. The statistical error parameters of the results obtained show that the average absolute percent relative errors and R^2 of the three sub-data set (total) results are about 3.44 and 0.898, respectively. This indicator demonstrates acceptable accuracy of the method developed for calculation of the *z*-factor of the gasses studied.

A detailed statistical error analysis of the proposed model for *z*-factor in this work is listed in **Table 5. 17**. The results listed for the training, validation and testing phases in **Table 5. 17** reveal that the new model has a reliable performance. A crossplot of the training, validation, and test datasets for *z*-factor, obtained by Eq. (5.20) is illustrated graphically in **Fig. 5. 70**. The results indicate that the new model provides a more precise estimation of the *z*-factor. Additionally, it is obvious that almost all data

points obtained by the newly developed GEP model lie on the unit slope line and this reveals its prediction capability**. Fig. 5. 71** represents the error distribution of the model for the determination of the *z*-factor of natural gasses. The figure confirms that the proposed model has a small error range and a low scatter around the zero error line. This indicates the potential of Eq. (5.20) for estimation of *z*-factor with a small expected error.

Table 5. 17 Statistical error parameters to determine the *z*-factor of the developed model (including training, validation and prediction sets).

Statistical Parameter	
training set	
R^2	0.897
Average Absolute Percent Relative Error	3.47
Standard deviation error	0.04
Root mean square error	0.04
N	784
validation set	
R^2	0.883
Average Absolute Percent Relative Error	3.47
Standard deviation error	0.04
Root mean square error	0.04
N	97
test set	
R^2	0.921
Average Absolute Percent Relative Error	3.46
Standard deviation error	0.04
Root mean square error	0.04
N	97
total	
R^2	0.898
Average Absolute Percent Relative Error	3.44
Standard deviation error	0.04
Root mean square error	0.04
N	978

The performance of the model for determination of the *z*-factor for the experimental data studied has been compared with that of some of the most widely-utilized empirically derived models and equations of state available in the literature,

including six empirically derived models, viz. Dranchuk-Abu-Kassem (Dranchuk and Kassem, 1975), Dranchuk-Purvis-Robinson (Dranchuk et al., 1973), Hall-Yarborough (Hall and Yarborough, 1973), Beggs-Brill (Beggs and Brill, 1973), Shell Oil Company (2003), Gopal (Gopal, 1977), Heidaryan et al. (Heidaryan et al., 2010a), Azizi et al. (Azizi et al., 2010), and Sanjari-Lay (Sanjari and Lay, 2012b) and three EoS-based models viz. van der Waals (van der Waals, 2004), Peng-Robinson (Peng and Robinson, 1976), Lawal-Lake-Sil berberg (Lawal, 1999), Soave-Redlich-Kwong (Soave, 1972), Patel-Teja (Patel and Teja, 1982). Additionally, the feed-forward multi-layer artificial neural network (ANN) approach has been employed to conduct a further comparison of the model developed with other kinds of artificial intelligence techniques. To this end, two reliable optimization methods viz. particle swarm optimization (PSO) (Kennedy, 2010) and genetic algorithm (GA) (Holland, 1975) have been used to tune the ANN adjustable parameters, including weight and bias. More information regarding these methods can be found elsewhere (Chamkalani et al., 2013b; Kamyab et al., 2010a; Shokir et al., 2012).

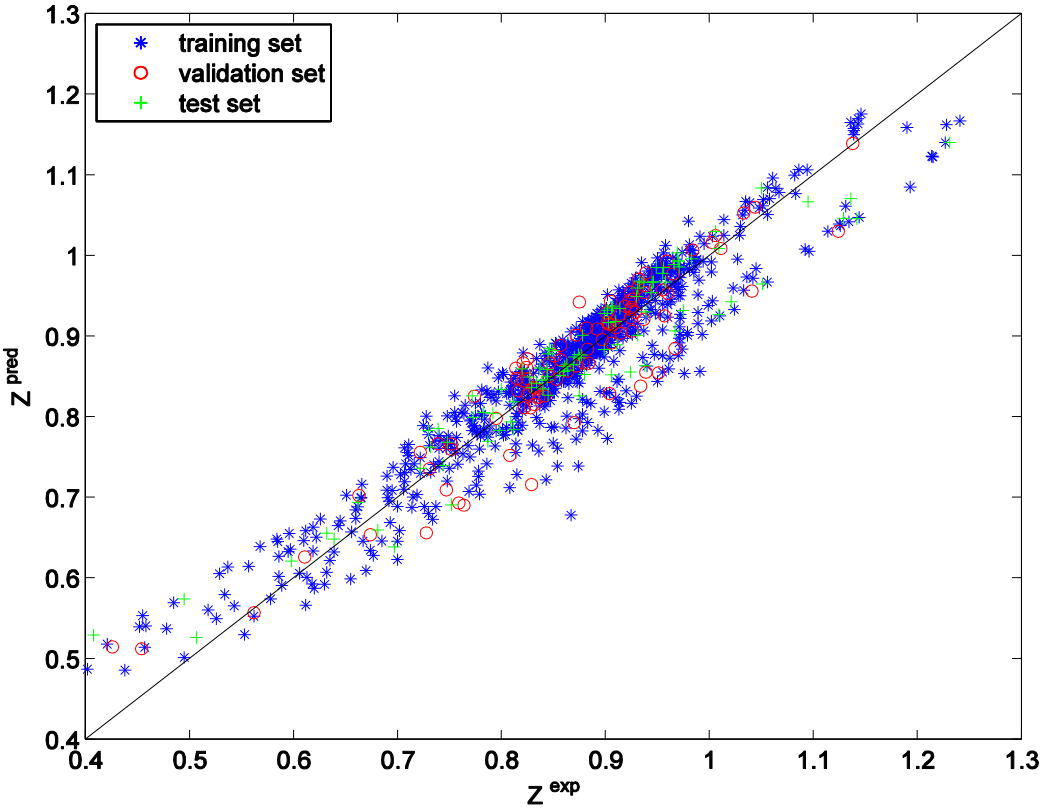

Fig. 5. 70 Comparison between the results of the model developed and the database values of *z*-factor.

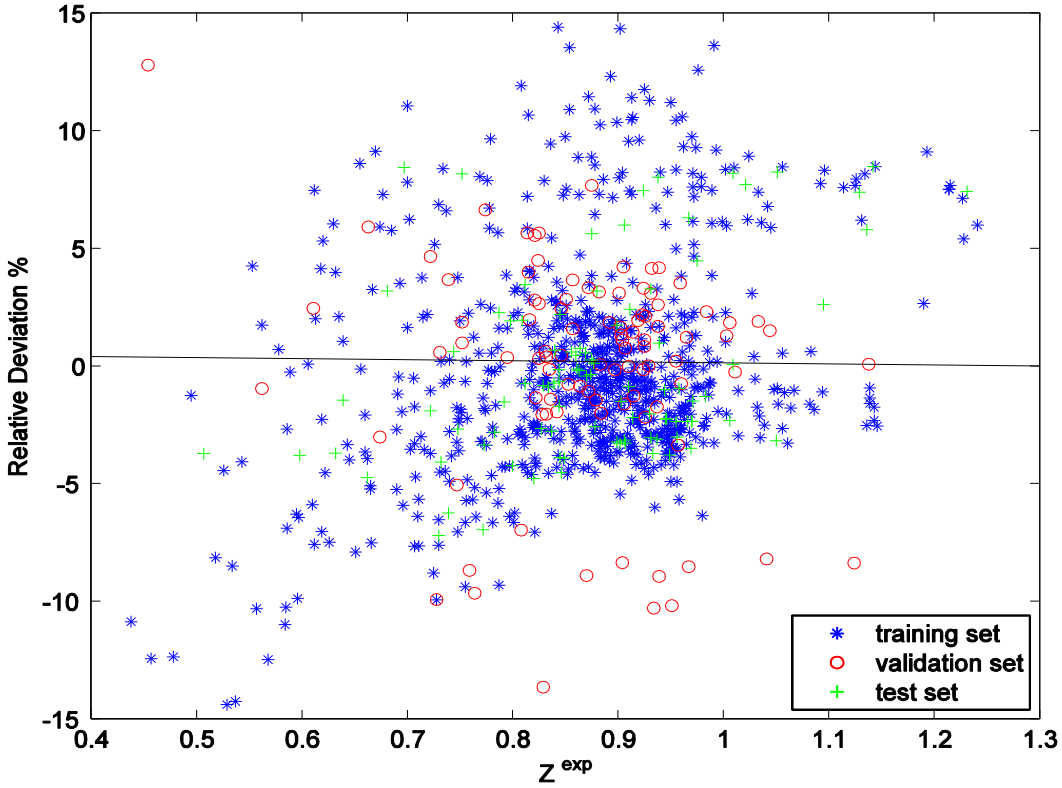

Fig. 5. 71 Relative deviations of the represented z-factor values by Eq. (5.20) from the database values.

Table 5. 18 reports on the correspondence of results of a comparative study. It is clear from a reading of **Table 5. 18** that the model in this study is simple and leads to a reasonable deviation of the determined *z*-factor values in comparison to the literature correlations, EoS-based models, and ANN based models. The bar plots in **Fig. 5. 72** represent the average absolute percent relative errors of the *z*-factor for the newly proposed method, EoS-based models, the ANN based model, and the empirical correlations. In **Fig. 5. 72**, the results show that Eq. (5.20) has a reliable level of accuracy. The proposed model is easy-to-use and does not need any soft-computing programs for calculation.

The results clearly demonstrate that the GEP algorithm is more powerful than ANN methodology in terms of accuracy, capability, and future usability. While the GEP algorithm model developed in the current study estimates the gas compressibility factors

of natural gasses with an AARD=3.44%, the AARD obtained for ANN optimized with the PSO and GA methods are 6.13, and 7.46 %, respectively. Additionally, over-fitting is a major problem faced in the ANN approach, in particular when a small dataset is used, because of the high number of adjustable parameters viz. weights and bias. It is worth noting that this is the first time that the GEP mathematical approach has been implemented for estimation of the *z*-factor of gasses. The results obtained indicate that the mathematical strategy implemented is very promising for evaluation of other petroleum fluid.

Table 5. 18 Comparative statistical error analysis of the empirical correlations, EoSs, and an artificial intelligent technique, and the newly developed model.

Method	**E_r %**	**E_a %**	**RMSE**	**R^2**
van der-Waals (van der Waals, 2004) EoS	0.31	6.42	0.0696	0.771
Peng-Robinson (Peng and Robinson, 1976) EoS	-5.34	6.10	0.0599	0.891
Lawal-Lake-Silberberg (Lawal, 1999) EoS	-2.66	4.43	0.0453	0.894
Patel-Teja (Patel and Teja, 1982) EoS	-1.18	4.15	0.0447	0.880
Soave-Redlich-Kwong (Soave, 1972) EoS	-3.14	4.82	0.0493	0.893
Dranchuk-Abu-Kassem (Dranchuk and Kassem, 1975) Corr.	4.21	8.18	0.0992	0.574
Dranchuk-Purvis-Robinson (Dranchuk et al., 1973) Corr.	4.66	4.77	0.0555	0.906
Hall-Yarborough (Hall and Yarborough, 1973) Corr.	1.46	3.59	0.0429	0.892
Beggs-Brill (Beggs and Brill, 1973) Corr.	4.95	5.07	0.0574	0.904
Shell Oil Company (2003) Corr.	5.34	5.40	0.0596	0.908
Gopal (Gopal, 1977) Corr.	6.12	6.26	0.0910	0.737
Azizi et al. (Azizi et al., 2010) Corr.	4.26	6.25	0.0792	0.772
Heidaryan et al. (Heidaryan et al., 2010a) Corr.	3.61	5.80	0.0762	0.778
Sanjari-Lay (Sanjari and Lay, 2012b) Corr.	0.66	5.67	0.0697	0.811
ANN-PSO	-1.13	6.13	0.0647	0.736
ANN GA	-1.85	7.46	0.0766	0.624
Eq. (5.20)	-0.25	3.44	0.04	0.898

To show the applicability domain of all methods investigated in this study graphically, the absolute percent relative error contour of the gas compressibility factor has been sketched together with the collected database for the input variables of Ppr and Tpr. **Fig. 5. 73** illustrates the absolute percent relative error contour of the gas compressibility factor predicted by the model developed in the current study. It is evident from the figure that the model is able to predict gas compressibility factors in the dataset range. However, the model developed in the current study could not estimate the gas compressibility factor with high accuracy in the Tpr range of 1-1.2, and Ppr range of 1-3.

However, the absolute percent relative error contours of the gas compressibility factor estimated by the comparative methods mentioned above are illustrated in Figs. 5 and 6. These figures clearly indicate that the method presented in this study is superior to the comparative methods. Furthermore, **Figs. 5. 74** and **Fig. 5. 75** show that the comparative methods have high errors in the Tpr range of 1-1.2 and Ppr range of 1-3, similar to the developed method. This may be due to the experimental errors when conducting laboratory measurements for the gas compressibility factors.

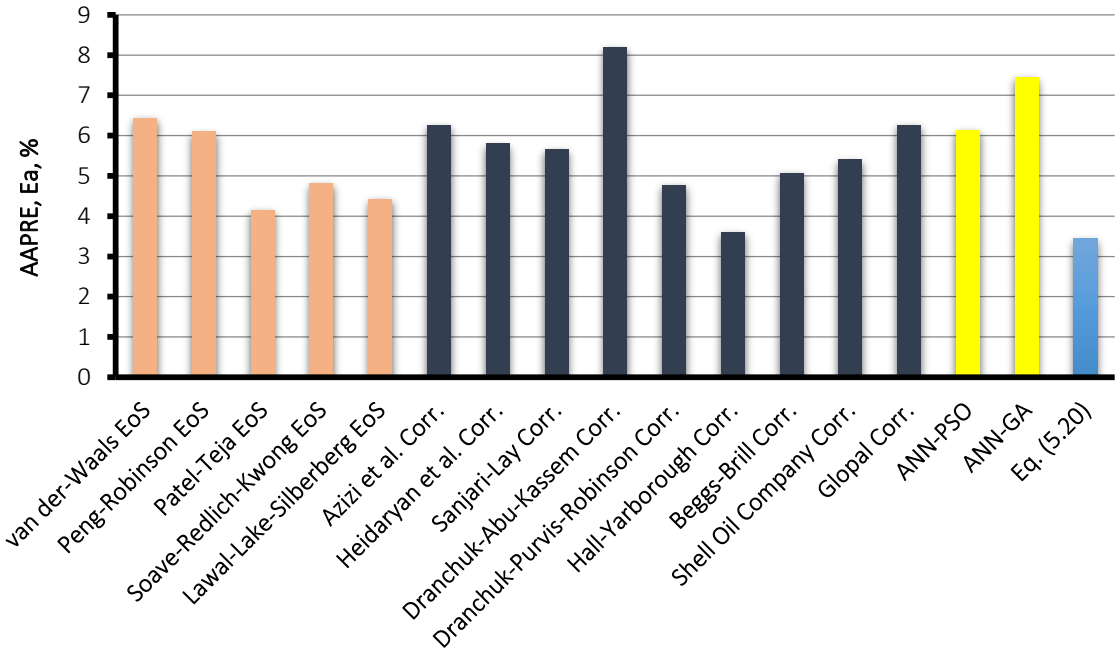

Fig. 5. 72 Calculated average absolute percent relative error for the empirical correlations, EoSs, and the artificial intelligent technique, as well as the proposed model (Eq. (5.20)).

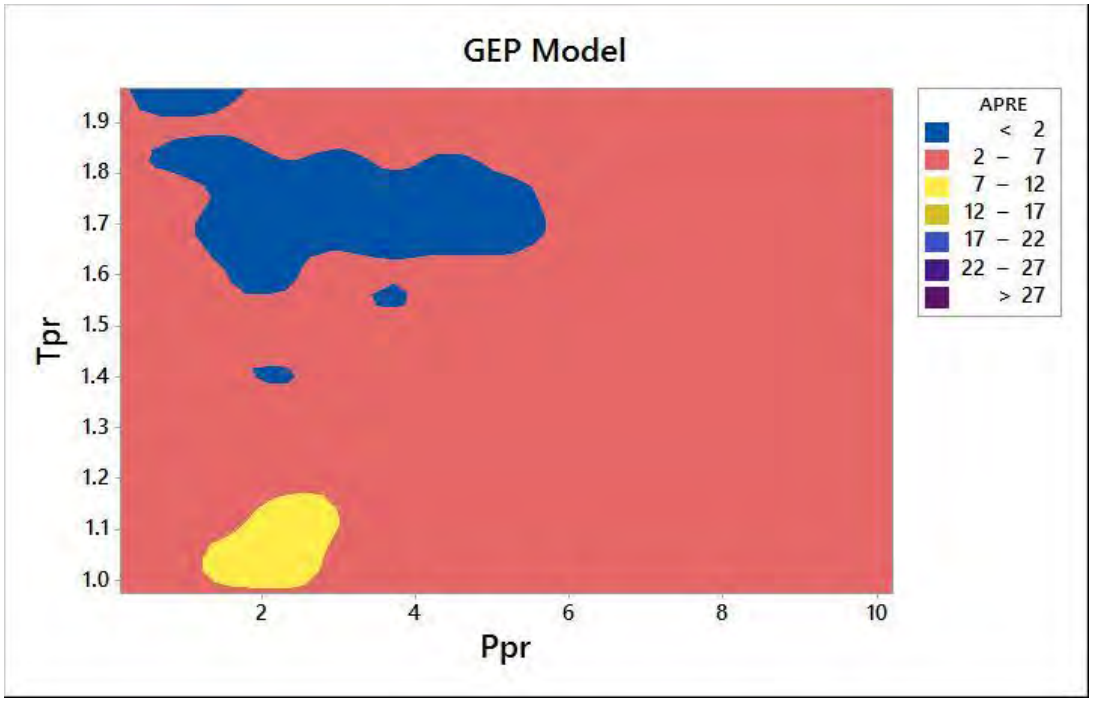

Fig. 5. 73 Absolute percent relative error contour of gas compressibility factor for the Eq. (5.20) in the ranges of P_{pr} and T_{pr}.

In the development of a predictive model or correlation, the leverage technique (detection of the outlier data points) plays a significant role to assess a group or groups of data which may differ from the bulk of the data present in a dataset (Mohammadi et al., 2012a; Mohammadi et al., 2012c; Rousseeuw and Leroy, 2005). The main objective of the leverage technique is the detection, in each experimental/literature databank, of data located outside of the applicability domain of the model. A detailed description of computational procedure and equations for the leverage technique can be found elsewhere (Mohammadi et al., 2012a; Mohammadi et al., 2012c; Rousseeuw and Leroy, 2005). Hence, to check whether the GEP model is statistically acceptable; the Williams plot has been provided.

The existence of the majority of data points in the ranges $0 \leq H \leq 0.0092$ and $-3 \leq$ Standardized Residuals ≤ 3 confirms that the GEP model developed for the calculation of *z*-factor is statistically accurate and reliable. As a consequence, good high leverage data points are located in the domain of $0.0092 < H$ for the method presented. Those good leverage points which are outside of the ranges $-3 \leq$ Standardized Residuals ≤ 3 may be regarded as outlier data points in terms of the applicability domain of the presented GEP

model. The results of the *z*-factor predictive method illustrate that a few of the data points are located in the aforementioned domain (**Fig. 5. 76**).

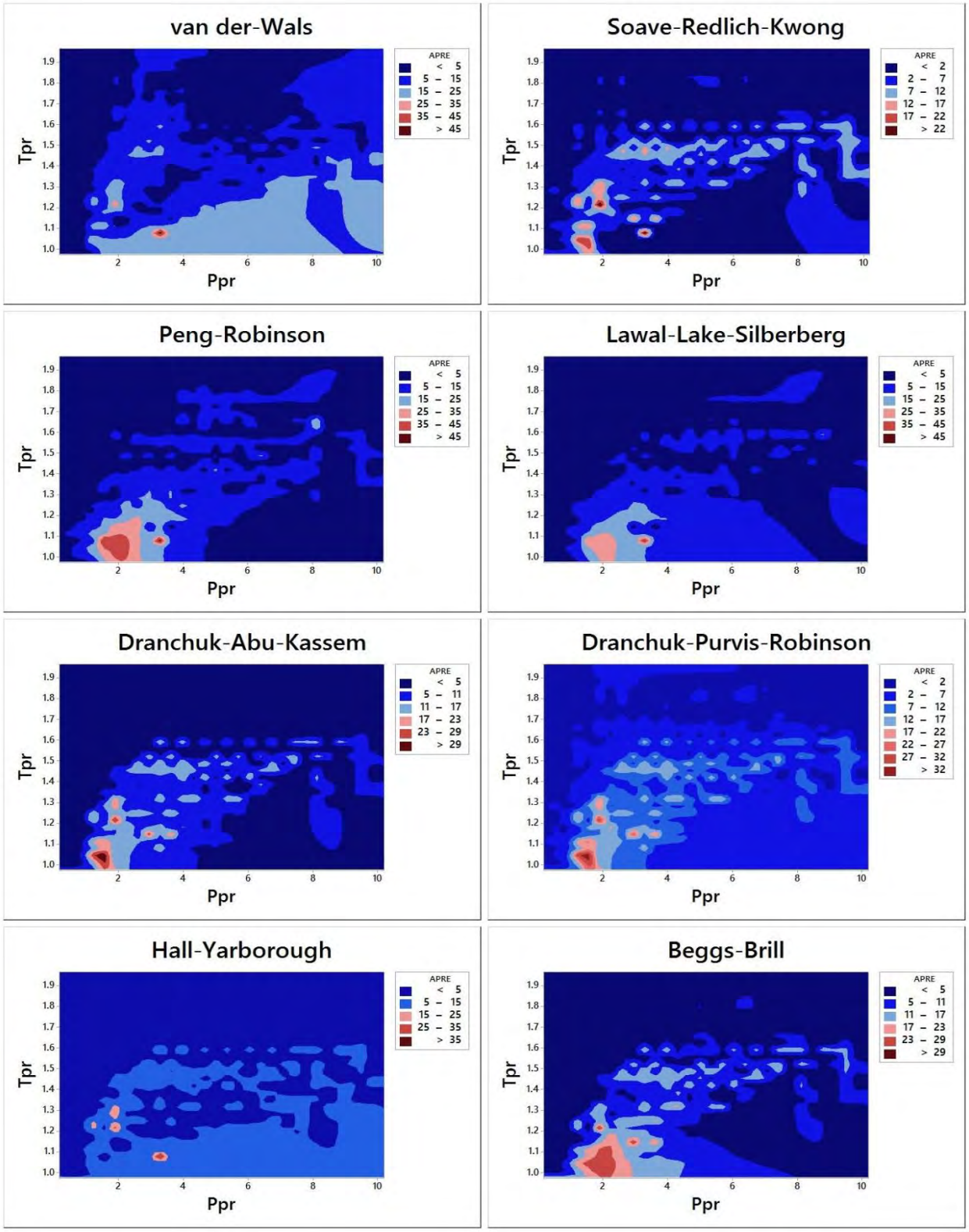

Fig. 5. 74 Absolute percent relative error contour of gas compressibility factor for the comparative methods (set I) in the ranges of P_{pr} and T_{pr}.

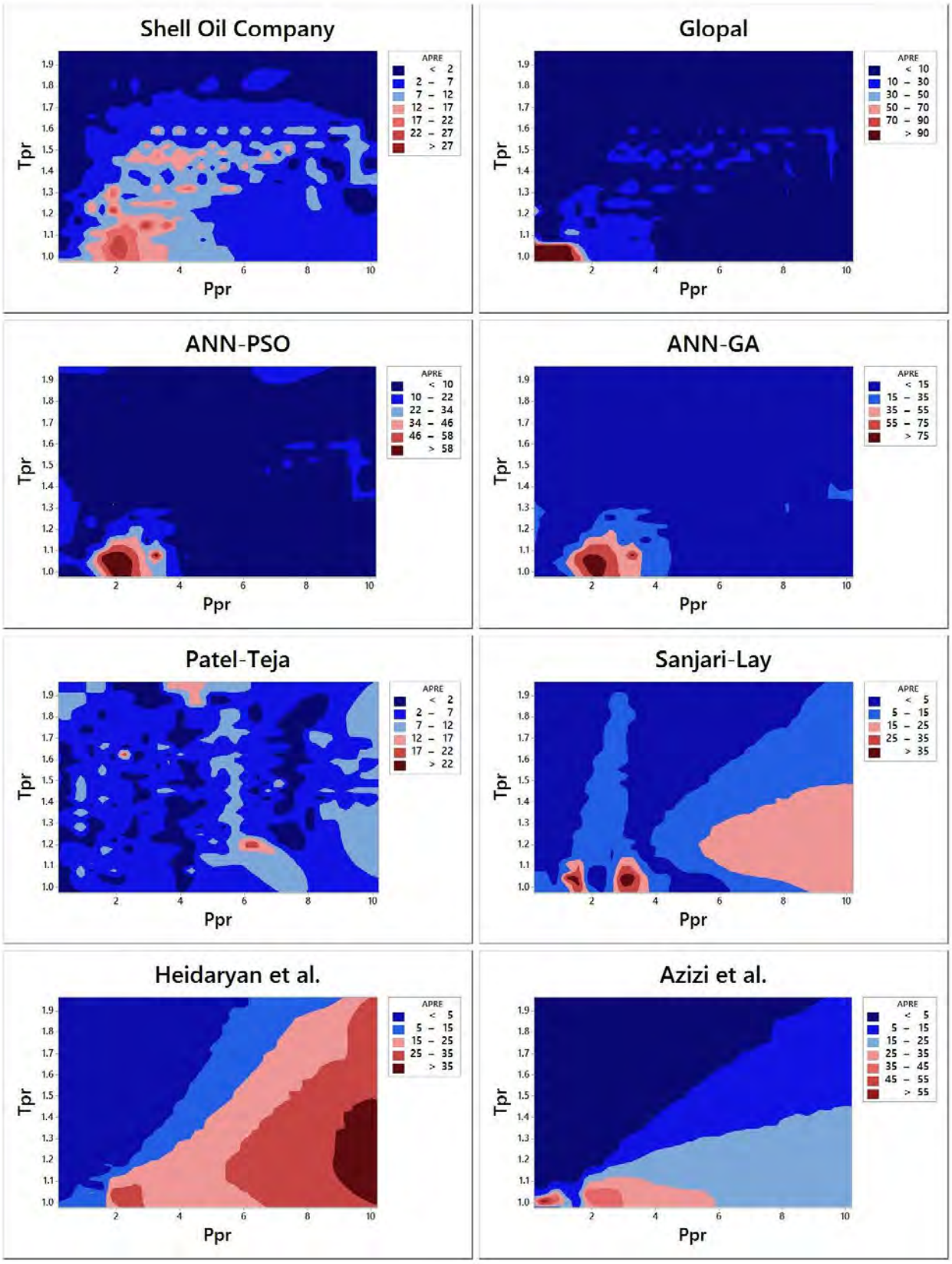

Fig. 5. 75 Absolute percent relative error contour of gas compressibility factor for the comparative methods (set II) in the ranges of P_{pr} and T_{pr}.

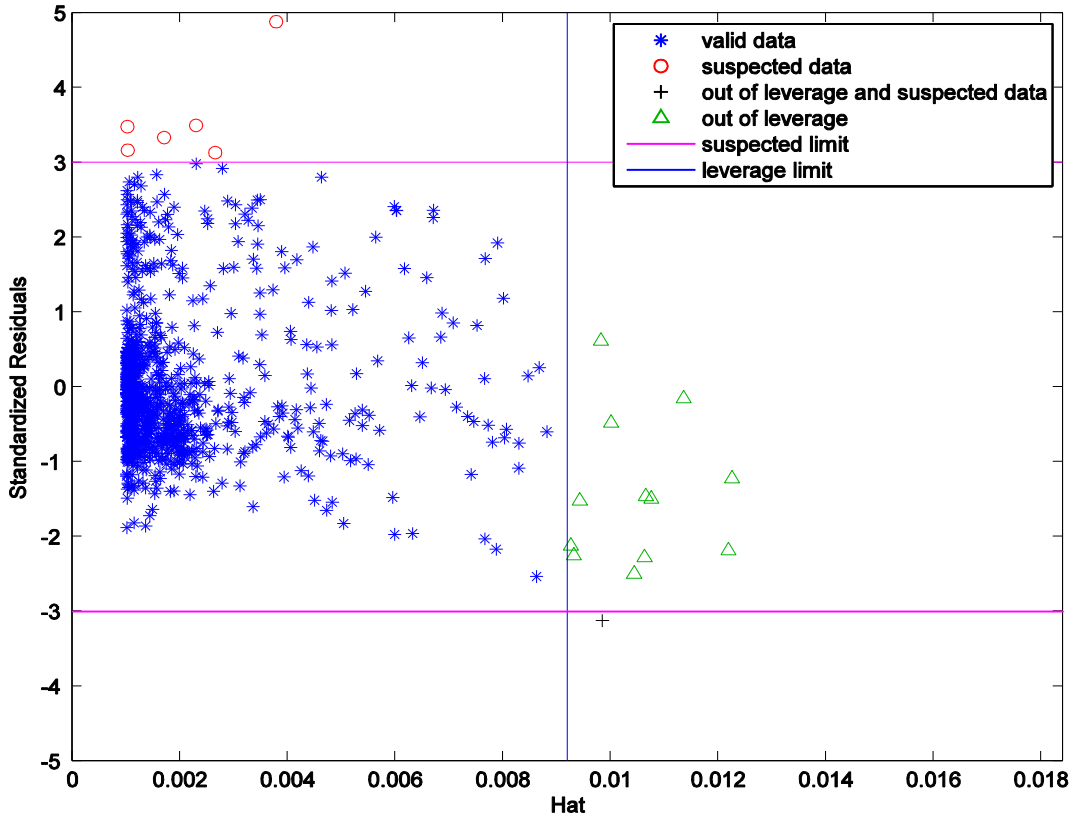

Fig. 5. 76 Detection of probable doubtful data of the *z*-factor and the applicability domain of the model developed.

5.12. Structural Based Models for the Estimation of Watson Characterization Factors of Hydrocarbon Components

Petroleum is one of the most complex mixtures and is defined as a substance, generally liquid, which occurs naturally in the earth and is composed mainly of mixtures of chemical compounds of carbon and hydrogen with or without other nonmetallic elements such as sulfur, oxygen, and nitrogen. Exact identification of the components available in unique petroleum cuts is highly desired in petroleum science and industry. For this identification, many characterization factors have been defined. One of the characterization factors which is widely used is the Watson characterization factor or Universal Oil Products Company (UOP) characterization factor.

5.12.1.Model with 10 adjustable parameters

In order to obtain an accurate and reliable correlation for the determination of Watson characterization of hydrocarbon components, a number of chemical substructures prepared were introduced into the sequential search mathematical algorithm. Furthermore, in order to obtain the optimal correlation in terms of both the number of chemical substructures and accuracy, a threshold value of 0.01 was considered for the reduction in the AARD as a stopping criterion. It means that when the improvement of the model AARD% was less than 0.01, the sequential search mathematical strategy was automatically stopped and reported the final model. The best model derived by using group contribution approach to predict the Watson characterization data for a ten-chemical structures correlation equation, with a total AARD= 14.60, is tabulated in **Table 5. 19**:

Table 5. 19 Equation with 10 parameters for the estimation of Watson characterization factor using GC approach. A detailed definition is presented in the Dragon software.

Equation	Definitions
Y = 4.941786494 * X1	X1 : E : nCp
+ 0.108759048 * X2	X2 : F : nCs
+ 0.330039526 * X6	X6 : J : nCrt
- 0.919667638 * X14	X14 : R : n=C=
- 4.628165746 * X25	X25 : AC : NsCH3
- 1.217880596 * X27	X27 : AE : NtCH
- 0.665938713 * X35	X35 : AW : F01[C-C]
+ 0.183872627 * X38	X38 : AZ : F04[C-C]
+ 0.158262938 * X39	X39 : BA : F05[C-C]
+ 0.254532833 * X41	X41 : BC : F07[C-C]

5.12.2.Model with 15 adjustable parameters

The best model derived by using group contribution approach to predict the Watson characterization data for a fifteen-chemical structures correlation equation, with a total AARD= 17.36, is tabulated in **Table 5. 20**:

Table 5. 20 Equation with 15 parameters for the estimation of Watson characterization factor using GC approach. A detailed definition is presented in the Dragon software.

Equation	Definitions
Y = 0.296361869 * X2	X2 : F : nCs
+ 0.299771183 * X3	X3 : G : nCt
- 0.195144625 * X5	X5 : I : nCrs
- 0.938337165 * X14	X14 : R : n=C=
+ 2.174648042 * X17	X17 : U : C-001
- 0.317516851 * X20	X20 : X : C-015
- 2.037246983 * X22	X22 : Z : C-021
- 2.283929986 * X25	X25 : AC : NsCH3
- 1.685603621 * X35	X35 : AW : F01[C-C]
+ 0.497567616 * X36	X36 : AX : F02[C-C]
+ 0.150798585 * X37	X37 : AY : F03[C-C]
+ 0.186566269 * X38	X38 : AZ : F04[C-C]
+ 0.179758444 * X39	X39 : BA : F05[C-C]
+ 0.252660707 * X40	X40 : BB : F06[C-C]
+ 0.161989378 * X42	X42 : BD : F08[C-C]

5.12.3.Model with 20 adjustable parameters

The best model derived by using group contribution approach to predict the Watson characterization data for a twenty-chemical structures correlation equation, with a total AARD= 16.90, is tabulated in **Table 5. 21**:

Table 5. 21 Equation with 20 parameters for the estimation of Watson characterization factor using GC approach. A detailed definition is presented in the Dragon software.

Equation	Definitions
Y = 2.649191783 * X1	X1 : E : nCp
- 2.766611746 * X4	X4 : H : nCq
- 0.096054873 * X5	X5 : I : nCrs
+ 0.248508342 * X6	X6 : J : nCrt
+ 0.458263980 * X7	X7 : K : nCrq
- 1.135284865 * X8	X8 : L : nCar
+ 0.924619192 * X9	X9 : M : nCbH
- 0.265558515 * X12	X12 : P : nR=Cs
- 1.319737539 * X13	X13 : Q : nR=Ct
- 2.406420367 * X25	X25 : AC : NsCH3
- 1.843847001 * X27	X27 : AE : NtCH
- 1.042638849 * X28	X28 : AF : NsssCH
- 1.262497738 * X29	X29 : AG : NddC
- 1.799478217 * X35	X35 : AW : F01[C-C]
+ 0.875342269 * X36	X36 : AX : F02[C-C]
+ 0.144627020 * X37	X37 : AY : F03[C-C]
+ 0.206540886 * X38	X38 : AZ : F04[C-C]
+ 0.198352266 * X39	X39 : BA : F05[C-C]
+ 0.242696246 * X40	X40 : BB : F06[C-C]
+ 0.172614950 * X42	X42 : BD : F08[C-C]

CHAPTER 6

6. Conclusions and Recommendations

This study set out to develop simple, efficient, accurate and reliable predictive models for the determination of petroleum reservoir fluid properties. To this end, the largest possible databanks were gathered from two main sources including the actual field data and previously published data available in the literature. The databases collected cover a wide ranges of reservoir rock and fluid properties coming from the various geographical regions of world.

To develop predictive models for the determination of petroleum reservoir fluid properties, robust artificial intelligence strategies viz., gene expression programming, artificial neural networks, least square support vector machine, adaptive neuro-fuzzy inference system, and decision tree computational schemes, were utilized. In order to tune the adjustable parameters associated with the algorithms mentioned above, different optimization techniques were employed simultaneously viz., couples simulated annealing, particle swarm optimization, and genetic algorithm. Many empirically derived correlations were developed and compared.

The results show that the models are superior in terms of accuracy and simplicity of use to those in the literature.

6.1. Conclusions

The main conclusions for each property studied are listed as follows:

6.1.1. Model for determination of surfactant retention in porous media during chemical flooding

The error analysis indicated an $R^2 = 0.9464$ for the predicted values of surfactant retention is reported for the new model. A high quality of results are obtained by using

the GEP mathematical strategy, which are quantified by the average absolute relative deviation error obtained. The average absolute relative deviation error for the newly developed correlation is approximately 9.44%. These results indicate that the newly developed model predicts the surfactant retention values with an acceptable accuracy. The accuracy of method presented in this study is suitable for utilization in the design of EOR processes in petroleum reservoir disciplines. It can therefore be widely applied in situations where experimentally measured records are not available.

6.1.2. Model for the calculation of dew point pressure in gas condensate reservoirs

The proposed model has an overall AARD of 7.88 % and an R-squared equal to 0.89. It is observed that the model has a superior performance to the other correlations with respect to calculated statistical error parameters. Furthermore, there is good agreement between the actual data and values calculated using the GEP-based model proposed in this study. The comparative analysis conducted in this study also confirms that the GEP-based model is able to calculate the desired parameter (*i.e.* the DPP in gas condensate reservoirs) with greater accuracy and consistency. The model also has a smaller number of adjustable, resulting in the model optimization and development of GEP-based models being faster, less laborious, and less costly.

6.1.3. Models for the calculation of the oil PVT properties

Our conclusions indicated that the GEP model that is developed for the calculation of oil formation volume factor has an AAPRE=3.62% and a R^2=0.93 and for bubble point pressure the results are 15.3% and 0.88, respectively. These amounts of error confirm the level of accuracy of the new model for the prediction of PVT properties. The results obtained from the comparison analysis confirm that the models presented are quicker to calculate, and are more accurate, reliable and capable. The Leverage analysis performed in current study showed that there are 12 data points for the oil formation volume factor model and 13 data points for the bubble point pressure model that are outside of the applicability domain of the GEP models and are accounted as outliers whose values may be doubtful, compared with the corresponding actual data.

6.1.4. Models for the determination of dead, saturated and under-saturated reservoir oil viscosities

The results reveal that GEP-based models calculate the dead, saturated and under-saturated reservoir oil viscosities in a wide range of reservoir properties of Iranian crude oils. The models developed for dead, saturated and under-saturated reservoir oil viscosities are shown to be more accurate than the studied comparative methods in terms of all error factors investigated. The calculated AARD % for the developed dead, saturated and under-saturated reservoir oil viscosities models are 17.29, 13.55, and 1.47, respectively. From the results obtained in this study it can be concluded that the proposed models are reliable for the estimation of developed dead, saturated and under-saturated reservoir oil viscosities.

6.1.5. Model for calculating solution GOR data

The applicability domain of the proposed method was determined through the detection of outlier data points using the Leverage approach. It is found that only 26 data points (among more than 1000 data values) are identified as outlier data points. The results obtained indicate that the model used, with an AARD value of 19.83%, outperforms all comparable models studied. A sensitivity analysis conducted in this study indicates that bubble point pressure and gas gravity have the largest and smallest influences, respectively, on the predicted solution GOR data. Furthermore, the model has more applicability for the estimation of solution GORs for reservoirs containing light oils.

6.1.6. Models to calculate the asphaltene precipitated versus solvent to oil dilution ratio

A study was conducted on the changes in asphaltene precipitation versus solvent to oil dilution ratio in an Iranian asphaltenic crude oil at different temperatures for *n*-pentane, *n*-hexane, and *n*-heptane solvents. It was found that the previously published scaling equations are not fully accurate and satisfactory. To compare all methods investigated, two panels are considered. In the smart based panel, the LSSVM approach has the most accuracy and could predict the asphaltene precipation with an AARD of 3%. In the symbolic equations panel, the method proposed on the basis of GEP approach gave an acceptable AARD of 8.5 %.

The results indicate that the methods developed are applicable for the simulation of asphaltene precipitation in black oil software. Moreover, a simple symbolic method presented can be a reliable alternative to existing scaling equations and complex thermodynamic methods.

6.1.7. Modeling of phase behavior of wax deposition

Our conclusions demonstrate similar results for ANN and LSSVM models. The AAPRE and R^2 for ANN and LSSVM models are 0.6% and 0.95, respectively. As a result, the regression DT model provided more acceptable results compared to the ANN and LSSVM approaches. The statistical error analysis for regression DT model shows an AAPRE=0.3 and R^2=0.97. Finally, it has been shown in this study that application of intelligent approaches such as ANN, LSSVM, and DT could be preferable to complex thermodynamics methods for the evaluation of phase behavior of wax deposition.

6.1.8. Models for the characterization of the heptane-plus properties of crude oil, and gas-condensate

The results indicate that all of the methods developed in this study can be applied for the characterization of C_{7+} properties. The statistical error analysis revealed that decision tree modelling is superior to the ANN, LSSVM, and GEP methods for the prediction of all three properties studied. Additionally, simple symbolic equations proposed by gene expression programming is a capable alternative for the existing distribution methods to characterize the heavier and complex components of crude oils. The AARD obtained by using GEP algorithm were reported 4, 2, and 12 % for boiling point, specific gravity, and molecular weight, respectively.

6.1.9. Model to determine the vaporization enthalpies of pure hydrocarbon components and petroleum fractions

The results of the comparison confirmed the advantages of the CSA-LSSVM model over previously reported methods investigated. The Leverage approach was used to identify probable data outliers. It was found that only one data point is outside of applicability domain of the model suggested. The total AARD and R^2 for the LSSVM model proposed in the current study are 1.15514% and 0.9982, respectively. The model is

considered to be reliable for the estimation of the vaporization enthalpies of pure hydrocarbon components and petroleum fractions.

6.1.10.Model to predict gasoline properties

The results obtained indicated the success of the mathematical strategy pursued in this study for the prediction of gasoline properties, SG, MON, RON and RVP. The coefficient of determination, R^2, of the proposed model is 0.990, 0.933, 0.955 and 0.920 for SG, MON, RON and RVP, respectively. The results obtained indicate that the LSSVM model can be applied as a reliable tool for predicting the properties of gasoline and consequently, can be used to determine the quality of the gasoline.

6.1.11.Model for the calculation of z-factor values

A comparison, based on the statistical and graphical analyses, made with other models (empirical correlations, equations of state, and an artificial intelligent technique), showed the superiority of the newly developed method via indices such as R^2, Ea and RMSE of 0.898, 3.45, and 0.04, respectively. This statistically indicates a satisfactory predictive tool. The model proposed in this study also provides a considerable improvement over previous proposed correlations and equations of state with broader applicability in terms of temperature and pressure ranges.

6.1.12.Model to predict Watson characterization of hydrocarbon components

A group contribution technique has been successfully developed for estimation of the Watson characterization. In addition, a comprehensive dataset of experimental Watson characterization data was used to develop a general group contribution correlation. A number of chemical substructures was implemented as model inputs. Using this approach, three models were developed having 10, 15 and 200 adjustable parameters with AARD of 14.6, 17.36, and 16.9, respectively. As a result, the model proposed here is reliable and also has appropriate capability for predict and modeling the physical property.

6.2. Recommendations

Most of petroleum reservoirs fluid properties available were investigated in this study. The main recommendations suggested by current study for future works are summarized as follows:

6.2.1. Potential Alternative Applications of New Methods

The methods developed in this study have a broader application within petroleum engineering and can be used to investigate other properties associated with reservoir rock. Furthermore, they offer alternatives for application in the investigation of a broad range of issues related to petroleum engineering such as, fluid flow through porous media, enhanced and improved oil recovery, well-testing and well log data analyses, drilling technology, production and operation, EOR screening, reserve estimation, and production performance prediction.

6.2.2. New Methods Provide Reliable and Simple Predictive Techniques

The determination of reservoir fluid properties using laboratory experiments are complex and can be time-consuming and expensive. Furthermore, existing thermodynamics models have some short comings in the phase behavior modeling of petroleum reservoir fluid properties. Consequently, reservoir fluid properties, in the absence of experimental measurements, can be determined through empirical methods. Therefore, the models presented in the current study provide useful, reliable and simple predictive techniques, which are easier than existing approaches, less complicated, and with fewer computations.

6.2.3. New methods Avoid an Over-fitting Problem

Although artificial intelligence methods are powerful methods in solving regression and classification problems, the main problem associated with such techniques is the over-fitting issue. As a consequence, the development of methods/algorithms on the basis of artificial intelligence with a low number of adjustable parameters is necessary in order to avoid the over-fitting problem.

6.2.4. New methods Allow for Rapid Investigations

Petroleum engineers seek a rapid way to obtain accurate values for petroleum reservoir fluid properties, taking into account both economic and technical issues. The methods and models developed in this study are appropriate for utilization in the design of processes, and software for reservoir simulation, relating to petroleum reservoirs.

REFERENCES

Fluid Properties Package, Shell Oil Company, 2003.

A. González, M.A.B., R. Startzman,, 2003, Improved neural-network model predicts dewpoint pressure of retrograde gases. Journal of Petroleum Science and Engineering, 37 183-194.

Abdul-Majeed, G.H., Salman, N.H. and Scarth, B., 1988, An empirical correlation for oil FVF prediction. Journal of Canadian Petroleum Technology, 27.

Ahmed, T., 2006, Reservoir engineering handbook. Access Online via Elsevier

Akbari, M. and Jalali, F., Year, Dew point pressure estimation of gas condensate reservoirs, using Artificial Neural Network (ANN). EUROPEC/EAGE Conference and Exhibition, Society of Petroleum Engineers, London, UK

Al-Anazi, A. and Gates, I., 2010, Support vector regression for porosity prediction in a heterogeneous reservoir: A comparative study. Computers & Geosciences, 36: 1494-1503.

Al-Anazi, A. and Gates, I., 2012, Support vector regression to predict porosity and permeability: Effect of sample size. Computers & Geosciences, 39: 64-76.

Al-Anazi, B.D., Pazuki, G., Nikookar, M. and Al-Anazi, A.F., 2011, The Prediction of the Compressibility Factor of Sour and Natural Gas by an Artificial Neural Network System. Petroleum Science and Technology, 29: 325-336.

Al-Bulushi, N., Araujo, M. and Kraaijveld, M., 2007, Predicting water saturation using artificial neural networks (ANNS). Neural Networks, 549: 57.

Al-Bulushi, N., King, P.R., Blunt, M.J. and Kraaijveld, M., 2009, Development of artificial neural network models for predicting water saturation and fluid distribution. Journal of Petroleum Science and Engineering, 68: 197-208.

Al-Fattah, S.M. and Al-Naim, H.A., 2009, Artificial-intelligence technology predicts relative permeability of giant carbonate reservoirs. SPE Reservoir Evaluation & Engineering, 12: 96-103.

Al-Khafaji, A., Abdul-Majeed, G. and Hassoon, S., 1987, Viscosity correlation for dead, live and undersaturated crude oils. J. Pet. Res, 6: 1-16.

Al-Marhoun, M.A., 1988, PVT correlations for Middle East crude oils. Journal of Petroleum Technology, 40: 650-666.

Al-Marhoun, M.A., 1992, New Correlation for formation Volume Factor of oil and gas Mixtures. Journal of Canadian Petroleum Technology, 31: 22-26.

Al-Marhoun, M.A., 2004, Evaluation of empirically derived PVT properties for Middle East crude oils. Journal of Petroleum Science and Engineering, 42: 209-221.

Al-Shammasi, A., Year, Bubble point pressure and oil formation volume factor correlations. SPE Middle East oil show & conference, 241-256.

Al Adasani, A. and Bai, B., 2011, Analysis of EOR projects and updated screening criteria. Journal of Petroleum Science and Engineering, 79: 10-24.

Alavi, A.H., Gandomi, A.H., Mousavi, M. and Mollahasani, A., 2010, High-Precision Modeling of Uplift Capacity of Suction. Geomechanics and Engineering, 2: 253-280.

Albahri, T., 2014, Specific Gravity, RVP, Octane Number, and Saturates, Olefins, and Aromatics Fractional Composition of Gasoline and Petroleum Fractions by Neural Network Algorithms. Petroleum Science and Technology, 32: 1219-1226.

Aleme, H.G., Costa, L.M. and Barbeira, P.J., 2009, Determination of ethanol and specific gravity in gasoline by distillation curves and multivariate analysis. Talanta, 78: 1422-1428.

Alimadadi, F., Fakhri, A., Farooghi, D. and Sadati, H., 2011, Using a committee machine with artificial neural networks to predict PVT properties of Iran crude oil. SPE Reservoir Evaluation & Engineering, 14: 129-137.

Alizadeh, N., Mighani, S., Hashemi kiasari, H., Hemmati-Sarapardeh, A. and Kamari, A., Year, Application of Fast-SAGD in Naturally Fractured Heavy Oil Reservoirs: A Case Study. 18th Middle East Oil & Gas Show and Conference (MEOS)

Alkhasawneh, M.S., Ngah, U.K., Tay, L.T., Mat Isa, N.A. and Al-Batah, M.S., 2014, Modeling and Testing Landslide Hazard Using Decision Tree. Journal of Applied Mathematics, 2014.

Alomair, O.A., Elsharkawy, A.M. and Alkandari, H.A., Year, Viscosity prediction of kuwaiti heavy crudes at elevated temperatures. SPE Heavy Oil Conference and Exhibition, Kuwait City, Kuwait

Amin, J.S., Nikooee, E., Ayatollahi, S. and Alamdari, A., 2010, Investigating wettability alteration due to asphaltene precipitation: Imprints in surface multifractal characteristics. Applied Surface Science, 256: 6466-6472.

Aminian, K. and Ameri, S., 2005, Application of artificial neural networks for reservoir characterization with limited data. Journal of Petroleum Science and Engineering, 49: 212-222.

Aminzadeh, F., 2005, Applications of AI and soft computing for challenging problems in the oil industry. Journal of Petroleum Science and Engineering, 47: 5-14.

Arabloo, M., Shokrollahi, A., Gharagheizi, F. and Mohammadi, A.H., 2013, Toward a Predictive Model for Estimating Dew Point Pressure in Gas Condensate Systems. Fuel Processing Technology, (in press).

Arabloo, M., Amooie, M.-A., Hemmati-Sarapardeh, A., Ghazanfari, M.-H. and Mohammadi, A.H., 2014, Application of constrained multi-variable search methods for

prediction of PVT properties of crude oil systems. Fluid Phase Equilibria, 363: 121-130.

Arulampalam, G. and Bouzerdoum, A., 2003, A generalized feedforward neural network architecture for classification and regression. Neural Networks, 16: 561-568.

Asadisaghandi, J. and Tahmasebi, P., 2011, Comparative evaluation of back-propagation neural network learning algorithms and empirical correlations for prediction of oil PVT properties in Iran oilfields. Journal of Petroleum Science and Engineering, 78: 464-475.

Ashoori, S., Abedini, A., Abedini, R. and Nasheghi, K.Q., 2010, Comparison of scaling equation with neural network model for prediction of asphaltene precipitation. Journal of Petroleum Science and Engineering, 72: 186-194.

Assis, J.C., Teixeira, J.S., Pontes, L.A., Guimarães, P.R., Vianna, R.F., Bezerra, M.A. and Teixeira, L.S., 2013, Using the Doehlert matrix as a tool for studying the influence of gasoline components on octane numbers. Fuel, 113: 744-749.

Atiqullah, M.M. and Rao, S., 1993, Reliability optimization of communication networks using simulated annealing. Microelectronics Reliability, 33: 1303-1319.

Austad, T., Year, A Review of Retention Mechanisms of Ethoxylated Sulfonates in Reservoir Cores. SPE International Symposium on Oilfield Chemistry

Azizi, N., Behbahani, R. and Isazadeh, M., 2010, An efficient correlation for calculating compressibility factor of natural gases. Journal of Natural Gas Chemistry, 19: 642-645.

Balabin, R.M., Safieva, R.Z. and Lomakina, E.I., 2007, Comparison of linear and nonlinear calibration models based on near infrared (NIR) spectroscopy data for gasoline properties prediction. Chemometrics and Intelligent Laboratory Systems, 88: 183-188.

Bandyopadhyay, P. and Sharma, A., 2011, Development of a new semi analytical model for prediction of bubble point pressure of crude oils. Journal of Petroleum Science and Engineering, 78: 719-731.

Baniasadi, H., Kamari, A., Heidararabi, S., Mohammadi, A.H. and Hemmati-Sarapardeh, A., 2015, Rapid method for the determination of solution gas-oil ratios of petroleum reservoir fluids. Journal of Natural Gas Science and Engineering, 24: 1-10.

Banki, R., Hoteit, H. and Firoozabadi, A., 2008, Mathematical formulation and numerical modeling of wax deposition in pipelines from enthalpy–porosity approach and irreversible thermodynamics. International Journal of Heat and Mass Transfer, 51: 3387-3398.

Barnum, R.S., Brinkman, F.P., Richardson, T.W., Spillette, A.G., 1995, Gas condensate reservoir behaviour: productivity and recovery reduction due to condensation. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers, Inc., Dallas, Texas.

Beal, C., 1946, The Viscosity of Air Water Natural Gas Crude Oil and Its Associated Gases at Oil Field Temperatures and Pressures. Transactions of the AIME, 165: 94-115.

Beggs, D.H. and Brill, J.P., 1973, A study of two-phase flow in inclined pipes. Journal of Petroleum Technology, 25: 607-617.

Beggs, H.D. and Robinson, J., 1975, Estimating the viscosity of crude oil systems. Journal of Petroleum Technology, 27: 1,140-141,141.

Bello, O., Reinicke, K. and Patil, P., 2008, Comparison of the performance of empirical models used for the prediction of the PVT properties of crude oils of the niger delta. Petroleum Science and Technology, 26: 593-609.

Benedict, M., Webb, G.B. and Rubin, L.C., 1942, 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, 10: 747.

Bennison, T., Year, Prediction of heavy oil viscosity. Presented at the IBC Heavy Oil Field Development Conference, London

Bergman, D.F. and Sutton, R.P., Year, An update to viscosity correlations for gas-saturated crude oils. SPE Annual Technical Conference and Exhibition, Anaheim, California, USA

Bhatt, A. and Helle, H.B., 2002, Committee neural networks for porosity and permeability prediction from well logs. Geophysical Prospecting, 50: 645-660.

Boukadi, F., Al-Alawi, S., Al-Bemani, A. and Al-Qassabi, S., 1999, Establishing PVT correlations for Omani oils. Petroleum Science and Technology, 17: 637-662.

Statistical, B.P, Review of World Energy. June 2006.

Buckley, J.S. and Wang, J., 2002, Crude oil and asphaltene characterization for prediction of wetting alteration. Journal of Petroleum Science and Engineering, 33: 195-202.

Buxton, T. and Campbell, J., 1967, Compressibility factors for lean natural gas-carbon dioxide mixtures at high pressure. Old SPE Journal, 7: 80-86.

Carnahan, N.F. and Starling, K.E., 1969, Equation of state for nonattracting rigid spheres. The Journal of Chemical Physics, 51: 635.

Carvalho, M., Lozano, M.A., Serra, L.M. and Wohlgemuth, V., 2012, Modeling simple trigeneration systems for the distribution of environmental loads. Environmental Modelling & Software, 30: 71-80.

Chamkalani, A., Chamkalani, R. and Mohammadi, A.H., 2013a, Hybrid of Two Heuristic Optimizations with LSSVM to Predict Refractive Index as Asphaltene Stability Identifier. Journal of Dispersion Science and Technology.

Chamkalani, A., Mae'soumi, A. and Sameni, A., 2013b, An intelligent approach for optimal prediction of gas deviation factor using particle swarm optimization and genetic algorithm. Journal of Natural Gas Science and Engineering, 14: 132-143.

Chamkalani, A., Zendehboudi, S., Chamkalani, R., Lohi, A., Elkamel, A. and Chatzis, I., 2013c, Utilization of support vector machine to calculate gas compressibility factor. Fluid Phase Equilibria, 358: 189-202.

Chamkalani, A., 2015, Application of LS-SVM Classifier to Determine Stability State of Asphaltene in Oilfields by Utilizing SARA Fractions. Petroleum Science and Technology, 33: 31-38.

Chandra, B., Kothari, R. and Paul, P., 2010, A new node splitting measure for decision tree construction. Pattern Recognition, 43: 2725-2731.

Chen, G., Fu, K., Liang, Z., Sema, T., Li, C., Tontiwachwuthikul, P. and Idem, R., 2014, The genetic algorithm based back propagation neural network for MMP prediction in CO 2-EOR process. Fuel, 126: 202-212.

CHEN, W. and ZHAO, Z., 2006, Thermodynamic modeling of wax precipitation in crude oils. Chinese Journal of Chemical Engineering, 14: 685-689.

Chen, W., Zhao, Z., Zhang, X. and Wang, L., 2007, Thermodynamic phase equilibria of wax precipitation in crude oils. Fluid phase equilibria, 255: 31-36.

Chew, J.-N. and Connally Jr, C.A., 1959, A viscosity correlation for gas-saturated crude oils. Trans. AIME.

Choisy, C. and Belaid, A., Year, Handwriting recognition using local methods for normalization and global methods for recognition. Document Analysis and Recognition, 2001. Proceedings. Sixth International Conference on, 23-27.

Chuang, Y.-F., Lee, H.-T. and Lai, Y.-C., 2012, Item-associated cluster assignment model on storage allocation problems. Computers & Industrial Engineering, 63: 1171-1177.

Chylinski, K., Cebola, M., Meredith, A., Saville, G. and Wakeham, W., 2002, Apparatus for phase equilibrium measurements at high temperatures and pressures. The Journal of Chemical Thermodynamics, 34: 1703-1728.

Coulibaly, P. and Baldwin, C.K., 2005, Nonstationary hydrological time series forecasting using nonlinear dynamic methods. Journal of Hydrology, 307: 164-174.

Coutinho, J.A. and Daridon, J.-L., 2005, The limitations of the cloud point measurement techniques and the influence of the oil composition on its detection. Petroleum science and technology, 23: 1113-1128.

Cramer, N.L., Year, A representation for the adaptive generation of simple sequential programs. Proceedings of the First International Conference on Genetic Algorithms, 183-187.

Cristianini, N. and Shawe-Taylor, J., 2000, An introduction to support vector machines and other kernel-based learning methods. Cambridge university press

Crogh, A., 1996, Improved Correlations for Retrograde Gases. Texas A & M University, College Station.

D. Afidick, N.J.K., S.Bette, 1994, Productionperformanceofaretrogradegasreservoir: a case study of the Arunfield. SPE Asia Pacific Oil and Gas conference Society of Petroleum Engineers, Inc.,Melbourne, Australia, Copyright 1994, , 1994.

Danesh, A., 1998, PVT and phase behaviour of petroleum reservoir fluids. Elsevier

Darabi, H., Kavousi, A., Moraveji, M. and Masihi, M., 2010, 3D fracture modeling in Parsi oil field using artificial intelligence tools. Journal of Petroleum Science and Engineering, 71: 67-76.

Daridon, J.-L., Pauly, J. and Milhet, M., 2002, High pressure solid–liquid phase equilibria in synthetic waxes. Physical Chemistry Chemical Physics, 4: 4458-4461.

De Ghetto, G. and Villa, M., Year, Reliability analysis on PVT correlations. European Petroleum Conference

de Oliveira, F.S., Gomes Teixeira, L.S., Ugulino Araujo, M.C. and Korn, M., 2004, Screening analysis to detect adulterations in Brazilian gasoline samples using distillation curves. Fuel, 83: 917-923.

Dindoruk, B. and Christman, P.G., 2004, PVT properties and viscosity correlations for Gulf of Mexico oils. SPE Reservoir Evaluation & Engineering, 7: 427-437.

Doble, P., Sandercock, M., Du Pasquier, E., Petocz, P., Roux, C. and Dawson, M., 2003, Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. Forensic science international, 132: 26-39.

Dokla, M.E. and Osman, M.E., 1992, Correlation of PVT properties for UAE crudes. SPE formation evaluation, 7: 41-46.

Dowman, A. and Woolf, A., 1995, Prediction of the ability of hydrofluorocarbons to dissolve hydrocarbons by means of neural networks. Journal of fluorine chemistry, 74: 207-210.

Dranchuk, P., Purvis, R. and Robinson, D., Year, Computer calculation of natural gas compressibility factors using the Standing and Katz correlation. Annual Technical Meeting

Dranchuk, P. and Kassem, H., 1975, Calculation of Z factors for natural gases using equations of state. Journal of Canadian Petroleum Technology, 14.

Dudek, A.Z., Arodz, T. and Galvez, J., 2006, Computational methods in developing quantitative structure-activity relationships (QSAR): a review. Combinatorial chemistry & high throughput screening, 9: 213-228.

E. Organick, B.G., () 1952, Prediction of saturation pressures for condensate-gas and volatile-oil mixtures. Transactions of AIME, 195: 135.

Eberhart, R.C. and Kennedy, J., Year, A new optimizer using particle swarm theory. Proceedings of the sixth international symposium on micro machine and human science, 39-43.

Egbogah, E.O. and Ng, J.T., 1990, An improved temperature-viscosity correlation for crude oil systems. Journal of Petroleum Science and Engineering, 4: 197-200.

Eilerts, K., Smith,R.V, 1942, Specific volumes and phase-boundary properties of separator-gas and liquid-hydrocarbon mixtures. ,Bureau of Mines, Bartlesville,

Okla. (USA): 74.

El-Sebakhy, E.A., 2009a, Data mining in forecasting PVT correlations of crude oil systems based on Type1 fuzzy logic inference systems. Computers & Geosciences, 35: 1817-1826.

El-Sebakhy, E.A., 2009b, Forecasting PVT properties of crude oil systems based on support vector machines modeling scheme. Journal of Petroleum Science and Engineering, 64: 25-34.

Elam, F.M., 1957, Prediction of Bubble Point Pressures and Formation Volume Factors from Field Data, University of Texas at Austin, Texas

Elsharkawy, A. and Alikhan, A., 1999, Models for predicting the viscosity of Middle East crude oils. Fuel, 78: 891-903.

Elsharkawy, A.M., Elgibaly, A.A. and Alikhan, A.A., 1995, Assessment of the PVT correlations for predicting the properties of Kuwaiti crude oils. Journal of Petroleum Science and Engineering, 13: 219-232.

Elsharkawy, A.M. and Alikhan, A.A., 1997, Correlations for predicting solution gas/oil ratio, oil formation volume factor, and undersaturated oil compressibility. Journal of Petroleum Science and Engineering, 17: 291-302.

Elsharkawy, A.M., Year, Modeling the properties of crude oil and gas systems using RBF network. Asia Pacific oil & gas conference, 35-46.

Elsharkawy, A.M. and Foda, S.G., 1998, EOS simulation and GRNN modeling of the constant volume depletion behavior of gas condensate reservoirs. Energy & Fuels, 12: 353-364.

Elsharkawy, A.M., 2002a, Predicting the dew point pressure for gas condensate reservoirs: empirical models and equations of state. Fluid phase equilibria, 193: 147-165.

Elsharkawy, A.M., 2002b, Predicting the dew point pressure for gas condensate reservoirs : empirical models and equations of state. Fluid Phase Equilibria, 193 147-165.

Elsharkawy, A.M., 2004, Efficient methods for calculations of compressibility, density and viscosity of natural gases. Fluid Phase Equilibria, 218: 1-13.

Elsharkwy, A. and Gharbi, R., 2000, Comparing classical and neural regression techniques in modeling crude oil viscosity. Advances in Engineering Software, 32: 215-224.

Erdogan, M., Mudford, B., Chenoweth, G., Holeywell, R. and Jakubson, J., Year, Optimization of decision tree and simulation portfolios: A comparison. SPE Hydrocarbon Economics and Evaluation Symposium

Esfahani, S., Baselizadeh, S. and Hemmati-Sarapardeh, A., 2015, On determination of natural gas density: Least square support vector machine modeling approach. Journal of Natural Gas Science and Engineering, 22: 348-358.

Eslamimanesh, A., Gharagheizi, F., Mohammadi, A.H. and Richon, D., 2011a, Artificial Neural Network modeling of solubility of supercritical carbon dioxide in 24 commonly used ionic liquids. Chemical Engineering Science, 66: 3039-3044.

Eslamimanesh, A., Gharagheizi, F., Mohammadi, A.H. and Richon, D., 2011b, Phase equilibrium modeling of structure H clathrate hydrates of methane+ water "insoluble" hydrocarbon promoter using QSPR molecular approach. Journal of Chemical & Engineering Data, 56: 3775-3793.

Eslamimanesh, A., Gharagheizi, F., Illbeigi, M., Mohammadi, A.H., Fazlali, A. and Richon, D., 2012a, Phase equilibrium modeling of clathrate hydrates of methane, carbon dioxide, nitrogen, and hydrogen+ water soluble organic promoters using Support Vector Machine algorithm. Fluid Phase Equilibria, 316: 34-45.

Eslamimanesh, A., Gharagheizi, F., Mohammadi, A.H. and Richon, D., 2012b, A statistical method for evaluation of the experimental phase equilibrium data of simple clathrate hydrates. Chemical Engineering Science.

F. Kurata, D.L.V.K., 1942, Critical properties of volatile hydrocarbon mixtures, . American Institute of Chemical Engineers AICHE Journal, 38 995-1021.

Fabian, V., 1997, Simulated annealing simulated. Computers & Mathematics with Applications, 33: 81-94.

Fan, L., Harris, B.W., Jamaluddin, A.J., Kamath, J., Mott, R., Pope, G.A., Shandrygin, A., and Whitson, C.H., 1998, Understanding gas-condensate reservoirs. ,

. Oilfield Rev., 10: 16-25.

Fang, W., Lei, Q. and Lin, R., 2003, Enthalpies of vaporization of petroleum fractions from vapor pressure measurements and their correlation along with pure hydrocarbons. Fluid Phase Equilibria, 205: 149-161.

Farasat, A., Shokrollahi, A., Arabloo, M., Gharagheizi, F. and Mohammadi, A.H., 2013, Toward an intelligent approach for determination of saturation pressure of crude oil. Fuel Processing Technology, 115: 201-214.

Fathinasab, M., Ayatollahi, S. and Hemmati-Sarapardeh, A., 2015, A Rigorous Approach to Predict Nitrogen-Crude Oil Minimum Miscibility Pressure of Pure and Nitrogen Mixtures. Fluid Phase Equilibria.

Fazavi, M., Hosseini, S.M., Arabloo, M., Shokrollahi, A. and Amani, M., 2013, Applying a Smart Technique for Accurate Determination of Flowing Oil/Water Pressure Gradient in Horizontal Pipelines. Journal of Dispersion Science and Technology, (in press).

Ferreira, C., 2001, Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems, 13: 87-129.

Ferreira, C., 2006, Gene Expression Programming: Mathematical Modeling by an Artificial Intelligence (Studies in Computational Intelligence). Springer-Verlag New York, Inc., Secaucus, NJ

Ferrell, H., King, D. and Sheely Jr, C., 1988, Analysis of the Low-Tension Pilot at Big Muddy Field Wyoming. SPE formation evaluation, 3: 315-321.

Frashad, F., LeBlanc, J., Garber, J. and Osorio, J., Year, Empirical PVT correlations for Colombian crude oils. SPE Latin America/Caribbean petroleum engineering conference, Port-of-Spain,Trinidad

Gandomi, A.H., Fridline, M.M. and Roke, D.A., 2013, Decision tree approach for soil liquefaction assessment. The Scientific World Journal, 2013.

Gao, D., Zhou, J. and Xin, L., Year, SVM-based detection of moving vehicles for automatic traffic monitoring. Intelligent Transportation Systems, 2001. Proceedings. 2001 IEEE, 745-749.

García, M.d.C. and Carbognani, L., 2001, Asphaltene-paraffin structural interactions. Effect on crude oil stability. Energy & fuels, 15: 1021-1027.

Garrett, H.E., 1972, Surface active chemicals. Pergamon

Ghanaei, E., Esmaeilzadeh, F. and Kaljahi, J.F., Year, Evaluation of Activity Coefficient Models in Prediction of Wax Appearance Temperature. Abu Dhabi International Petroleum Exhibition and Conference

Ghanaei, E., Esmaeilzadeh, F. and Kaljahi, J.F., 2007, A new predictive thermodynamic model in the wax formation phenomena at high pressure condition. Fluid phase equilibria, 254: 126-137.

Ghanaei, E., Esmaeilzadeh, F. and Fathikalajahi, J., 2012, Wax formation from paraffinic mixtures: A simplified thermodynamic model based on sensitivity analysis together with a new modified predictive UNIQUAC. Fuel, 99: 235-244.

Ghanaei, E., Esmaeilzadeh, F. and Fathikalajahi, J., 2014, High pressure phase equilibrium of wax: A new thermodynamic model. Fuel, 117: 900-909.

Gharagheizi, F., 2007, QSPR analysis for intrinsic viscosity of polymer solutions by means of GA-MLR and RBFNN. Computational materials science, 40: 159-167.

Gharagheizi, F., Eslamimanesh, A., Ilani-Kashkouli, P., Mohammadi, A.H. and Richon, D., 2012a, QSPR molecular approach for representation/prediction of very large vapor pressure dataset. Chemical Engineering Science, 76: 99-107.

Gharagheizi, F., Eslamimanesh, A., Sattari, M., Tirandazi, B., Mohammadi, A.H. and Richon, D., 2012b, Evaluation of thermal conductivity of gases at atmospheric pressure through a corresponding states method. Industrial & Engineering Chemistry Research, 51: 3844-3849.

Gharagheizi, F., Ilani-Kashkouli, P., Acree, W.E., Mohammadi, A.H. and Ramjugernath, D., 2013, A group contribution model for determining the vaporization enthalpy of

organic compounds at the standard reference temperature of 298K. Fluid Phase Equilibria, 360: 279-292.

Gharbi, R., 1997, Estimating the isothermal compressibility coefficient of undersaturated Middle East crudes using neural networks. Energy & fuels, 11: 372-378.

Gharbi, R. and Elsharkawy, A., Year, Neural Network Model for Estimating The PVT Properties of Middle East Crude Oils. Middle East Oil Show and Conference

Gharbi, R.B. and Elsharkawy, A.M., Year, Universal Neural Network Based Model for Estimating The PVT Properties of Crude Oil Systems. SPE Asia Pacific Oil and Gas Conference and Exhibition

Ghetto, G.D., Paone, F. and Villa, M., Year, Reliability Analysis on PVT Correlations. European Petroleum Conference

Ghiasi, M.M., Bahadori, A., Zendehboudi, S., Jamili, A. and Rezaei-Gomari, S., 2013, Novel methods predict equilibrium vapor methanol content during gas hydrate inhibition. Journal of Natural Gas Science and Engineering, 15: 69-75.

Glaso, O., 1980, Generalized pressure-volume-temperature correlations. Journal of Petroleum Technology, 32: 785-795.

Glover, C., Puerto, M., Maerker, J. and Sandvik, E., 1979, Surfactant Phase Behavior and Retention in Porous Media. Society of Petroleum Engineers Journal, 19: 183-193.

Goodall, C.R., 1993, 13 Computation using the QR decomposition. Handbook of Statistics, 9: 467-508.

Gopal, V., 1977, Gas z-factor equations developed for computer. Oil and Gas Journal (Aug. 8, 1977): 58-60.

Gramatica, P., 2007, Principles of QSAR models validation: internal and external. QSAR & combinatorial science, 26: 694-701.

Green, D.W. and Willhite, G.P., 1998, Enhanced oil recovery. Richardson, Tex.: Henry L. Doherty Memorial Fund of AIME, Society of Petroleum Engineers

H. Reamer, B.S., 1950, Volumetric behavior of oil and gas from a Louisianafield. Transactions of AIME, 189: 261-268.

Hadi Rostami-Hosseinkhani , F.E., Dariush Mowla, 2014, Application of expert systems for accurate determination of dew-point pressure of gas condensate reservoirs. Journal of Natural Gas Science and Engineering, 18: 296e303.

Hall, K.R. and Yarborough, L., 1973, A new equation of state for Z-factor calculations. Oil Gas J, 71: 82-92.

Healy, W.C., Maassen, C.W. and Peterson, R.T., 1959, A new approach to blending octanes. API Div. Refin., 39: 132-192.

Hegeman, P.S., Dong, C., Varotsis, N. and Gaganis, V., Year, Application of artificial neural networks to downhole fluid analysis. International Petroleum Technology Conference

Heidaryan, E., Moghadasi, J. and Rahimi, M., 2010a, New correlations to predict natural gas viscosity and compressibility factor. Journal of Petroleum Science and Engineering, 73: 67-72.

Heidaryan, E., Salarabadi, A. and Moghadasi, J., 2010b, A novel correlation approach for prediction of natural gas compressibility factor. Journal of Natural Gas Chemistry, 19: 189-192.

Heinze, L.R., Winkler, H.W. and Lea, J.F., Year, Decision Tree for selection of Artificial Lift method. SPE Production Operations Symposium

Hemmati-Sarapardeh, A., Mahmoudi, B., Ramazani, S.A. and Mohammadi, A.H., Experimental measurement and modeling of saturated reservoir oil viscosity. Korean Journal of Chemical Engineering: 1-12.

Hemmati-Sarapardeh, A., Alipour-Yeganeh-Marand, R., Naseri, A., Safiabadi, A., Gharagheizi, F., Ilani-Kashkouli, P. and Mohammadi, A.H., 2013a, Asphaltene precipitation due to natural depletion of reservoir: Determination using a SARA fraction based intelligent model. Fluid Phase Equilibria, 354: 177-184.

Hemmati-Sarapardeh, A., Khishvand, M., Naseri, A. and Mohammadi, A.H., 2013b, Toward reservoir oil viscosity correlation. Chemical Engineering Science, 90: 53-68.

Hemmati-Sarapardeh, A., Shokrollahi, A., Tatar, A., Gharagheizi, F., Mohammadi, A.H. and Naseri, A., 2013c, Reservoir Oil Viscosity Determination Using an Intelligent Approach. Fuel, (in press).

Hemmati-Sarapardeh, A., Majidi, S.-M.-J., Mahmoudi, B., Ahmad Ramazani, S.A. and Mohammadi, A., 2014a, Experimental measurement and modeling of saturated reservoir oil viscosity. Korean Journal of Chemical Engineering, 31: 1253-1264.

Hemmati-Sarapardeh, A., Shokrollahi, A., Tatar, A., Gharagheizi, F., Mohammadi, A.H. and Naseri, A., 2014b, Reservoir oil viscosity determination using a rigorous approach. Fuel, 116: 39-48.

Holland, J.H., 1975, Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. U Michigan Press

Hossain, M.S., Sarica, C., Zhang, H.-Q., Rhyne, L. and Greenhill, K., Year, Assessment and development of heavy oil viscosity correlations. SPE International Thermal Operations and Heavy Oil Symposium, Calgary, Canada

Hosseinzadeh, M. and Hemmati-Sarapardeh, A., 2014, Toward a predictive model for estimating viscosity of ternary mixtures containing ionic liquids. Journal of Molecular Liquids, 200: 340-348.

Hu, Y.-F. and Guo, T.-M., 2001, Effect of temperature and molecular weight of n-alkane precipitants on asphaltene precipitation. Fluid Phase Equilibria, 192: 13-25.

Huang, Y., Huang, G., Dong, M. and Feng, G., 2003, Development of an artificial neural network model for predicting minimum miscibility pressure in CO< sub> 2</sub> flooding. Journal of Petroleum Science and Engineering, 37: 83-95.

Humoud, A.A., Al-Marhoun, M.A.,, 2001, A new correlation for gas-condensate dewpoint pressure prediction. In: SPE Middle East Oil Show. Society of Petroleum Engineers Inc., Bahrain.

I. Marruffo, J.M., J. Him, G. Rojas,, 2001, Statistical forecast models to determine retrograde dew pressure and C7+ percentage of gas condensates on basis of production

test data of eastern Venezuelan reservoirs. SPE Latin American and Caribbean Petroleum Engineering Conference, Buenos Aires, Argentina.

Iglauer, S., Wu, Y., Shuler, P., Tang, Y. and Goddard III, W.A., 2010, New surfactant classes for enhanced oil recovery and their tertiary oil recovery potential. Journal of Petroleum Science and Engineering, 71: 23-29.

Ikiensikimama, S.S. and Ogboja, O., Year, New bubblepoint pressure empirical PVT correlation. Nigeria Annual International Conference and Exhibition

Irani, R. and Nasimi, R., 2011, Application of artificial bee colony-based neural network in bottom hole pressure prediction in underbalanced drilling. Journal of Petroleum Science and Engineering, 78: 6-12.

Jafari Kenari, S.A. and Mashohor, S., 2013, Robust committee machine for water saturation prediction. Journal of Petroleum Science and Engineering, 104: 1-10.

Jahanandish, I., Salimifard, B. and Jalalifar, H., 2011, Predicting bottomhole pressure in vertical multiphase flowing wells using artificial neural networks. Journal of Petroleum Science and Engineering, 75: 336-342.

Jang, J.-S.R., 1993, ANFIS: adaptive-network-based fuzzy inference system. Systems, Man and Cybernetics, IEEE Transactions on, 23: 665-685.

Ji, H.-Y., Tohidi, B., Danesh, A. and Todd, A.C., 2004, Wax phase equilibria: developing a thermodynamic model using a systematic approach. Fluid Phase Equilibria, 216: 201-217.

Johnson, S., 1991, Svrcek, WY. J. Can. Pet. Technol., 26: 60.

Johnson, S.E., Svrcek, W.Y. and Mehrotra, A.K., 1987, Viscosity prediction of Athabasca bitumen using the extended principle of corresponding states. Industrial & Engineering Chemistry Research, 26: 2290-2298.

Ju, B., Fan, T. and Jiang, Z., 2013, Modeling asphaltene precipitation and flow behavior in the processes of CO 2 flood for enhanced oil recovery. Journal of Petroleum Science and Engineering, 109: 144-154.

Junior, L.C.R., Ferreira, M.S. and da Silva Ramos, A.C., 2006, Inhibition of asphaltene precipitation in Brazilian crude oils using new oil soluble amphiphiles. Journal of Petroleum Science and Engineering, 51: 26-36.

Kamari, A., Hemmati-Sarapardeh, A., Mirabbasi, S.-M., Nikookar, M. and Mohammadi, A.H., 2013a, Prediction of sour gas compressibility factor using an intelligent approach. Fuel Processing Technology, 116: 209-216.

Kamari, A., Hemmati-Sarapardeh, A., Mirabbasi, S.-M., Nikookar, M. and Mohammadi, A.H., 2013b, Prediction of sour gas compressibility factor using an intelligent approach. Fuel Processing Technology, 116: 209-216.

Kamari, A., Khaksar-Manshad, A., Gharagheizi, F., Mohammadi, A.H. and Ashoori, S., 2013c, Robust Model for the Determination of Wax Deposition in Oil Systems. Industrial & Engineering Chemistry Research, 52: 15664−15672.

Kamari, A., Bahadori, A., Mohammadi, A.H. and Zendehboudi, S., 2014a, Evaluating the Unloading Gradient Pressure in Continuous Gas-lift Systems During Petroleum Production Operations. Petroleum Science and Technology, 32: 2961-2968.

Kamari, A., Bahadori, A., Mohammadi, A.H. and Zendehboudi, S., 2014b, A reliable model for estimating a hydrate inhibitor injection rate. Journal of Natural Gas Science and Engineering.

Kamari, A., Gharagheizi, F., Bahadori, A. and Mohammadi, A.H., 2014c, Determination of the equilibrated calcium carbonate (calcite) scaling in aqueous phase using a reliable approach. Journal of the Taiwan Institute of Chemical Engineers, 45: 1307-1313.

Kamari, A., Mohammadi, A., Bahadori, A. and Zendehboudi, S., 2014d, A Reliable Model for Estimating the Wax Deposition Rate During Crude Oil Production and Processing. Petroleum Science and Technology, 32: 2837-2844.

Kamari, A. and Mohammadi, A.H., 2014, Screening of Enhanced Oil Recovery Methods. Nova Science Publishers, Inc.

Kamari, A., Mohammadi, A.H., Bahadori, A. and Zendehboudi, S., 2014e, Prediction of Air Specific Heat Ratios at Elevated Pressures Using a Novel Modeling Approach. Chemical Engineering & Technology, 37: 2047-2055.

Kamari, A., Nikookar, M., Hemmati-Sarapardeh, A., Sahranavard, L. and Mohammadi, A.H., 2014f, Screening of Potential Application of EOR Processes in an Naturally Fractured Oil Reservoir. Book Chapter, In: Enhanced Oil Recovery: Methods, Economic Benefits and Impacts on the Environment, Nova Science Publishers, Inc., USA 2014.

Kamari, A., Nikookar, M., Sahranavard, L. and Mohammadi, A.H., 2014g, Rigorous Modeling for Efficient Enhanced Oil Recovery Screening Methods and Predictive Economic Analysis. Neural Computing and Applications.

Kamari, A., Nikookar, M., Sahranavard, L. and Mohammadi, A.H., 2014h, Efficient screening of enhanced oil recovery methods and predictive economic analysis. Neural Computing and Applications, 25: 815-824.

Kamari, A., Arabloo, M., Shokrollahi, A., Gharagheizi, F. and Mohammadi, A.H., 2015a, Rapid method to estimate the minimum miscibility pressure (MMP) in live reservoir oil systems during CO 2 flooding. Fuel, 153: 310-319.

Kamari, A., Bahadori, A., Mohammadi, A.H. and Zendehboudi, S., 2015b, New tools predict monoethylene glycol injection rate for natural gas hydrate inhibition. Journal of Loss Prevention in the Process Industries, 33: 222-231.

Kamari, A., Hemmati-Sarapardeh, A., Mohammadi, A.H., Hashemi-Kiasari, H. and Mohagheghian, E., 2015c, On the evaluation of Fast-SAGD process in naturally fractured heavy oil reservoir. Fuel, 143: 155-164.

Kamari, A., Safirii, A. and Mohammadi, A.H., 2015d, Compositional Model for Estimating Asphaltene Precipitation Conditions in Live Reservoir Oil Systems. Journal of Dispersion Science and Technology, 36: 301-309.

Kamyab, M., Sampaio, J.H., Qanbari, F. and Eustes, A.W., 2010a, Using artificial neural networks to estimate the z-factor for natural hydrocarbon gases. Journal of Petroleum Science and Engineering, 73: 248-257.

Kamyab, M., Sampaio Jr, J.H., Qanbari, F. and Eustes III, A.W., 2010b, Using artificial neural networks to estimate the< i> z</i>-factor for natural hydrocarbon gases. Journal of Petroleum Science and Engineering, 73: 248-257.

Karambeigi, M., Zabihi, R. and Hekmat, Z., 2011, Neuro-simulation modeling of chemical flooding. Journal of Petroleum Science and Engineering, 78: 208-219.

Kartoatmodjo, T. and Schmidt, Z., 1994a, Large data bank improves crude physical property correlations. Oil and Gas Journal;(United States), 92.

Kartoatmodjo, T. and Schmidt, Z., 1994b, Large data bank improves crude physical property correlations. Oil and Gas Journal, 92: 51-55.

Kartoatmodjo, T. and Schmidt, Z., 1994c, Large data bank improves crude physical property correlations. Oil & Gas Journal.

Katz, D.L., 1942, Prediction of the shrinkage of crude oils. Drilling and Production Practice: 137-147.

Kaviani, D., Bui, T., Jensen, J.L. and Hanks, C., 2008, The Application of Artificial Neural Networks With Small Data Sets: An Example for Analysis of Fracture Spacing in the Lisburne Formation Northeastern Alaska. SPE Reservoir Evaluation & Engineering, 11: 598-605.

Kaye, S., 1985, Offshore California viscosity correlations. COFRC, TS85000940,(Aug. 1985).

Kazemzadeh, Y., Malayeri, M., Riazi, M. and Parsaei, R., 2015, Impact of Fe 3 O 4 nanoparticles on asphaltene precipitation during CO 2 injection. Journal of Natural Gas Science and Engineering, 22: 227-234.

Kennedy, J., 2010, Particle swarm optimization. Encyclopedia of Machine Learning. Springer760-766.

Khairy, M., El-Tayeb, S. and Hamdallah, M., 1998, PVT correlations developed for Egyptian crudes. Oil and Gas Journal, 96.

Khan, S., Al-Marhoun, M., Duffuaa, S. and Abu-Khamsin, S., 1987, Viscosity correlations for Saudi Arabian crude oils, Manama, Bahrain.

Khoukhi, A., 2012, Hybrid soft computing systems for reservoir PVT properties prediction. Computers & Geosciences, 44: 109-119.

Koch, G., 2005, Discovering multi-core: extending the benefits of Moore's law. Technology, 1.

Koza, J.R., 1992, Genetic programming: on the programming of computers by means of natural selection. MIT press, Cambridge

Kumar, N., 2005, Compressibility factors for natural and sour reservoir gases by correlations and cubic equations of state

L. Nemeth, H.K., e,, 1967, A correlation of dewpoint pressure withfluid composition

and temperatur. SPE Journal, 7: 99-104.

Labedi, R., 1992, Improved correlations for predicting the viscosity of light crudes. Journal of Petroleum Science and Engineering, 8: 221-234.

Labedi, R.M., Year, Use of Production Data to Estimate the Saturation Pressure Solution Gor and Chemical Composition of Reservoir Fluids. SPE Latin America Petroleum Engineering Conference, Brazil

Lasater, J., 1958, Bubble point pressure correlation. Journal of Petroleum Technology, 10: 65-67.

Laughton, D.G., Joe, G., Paduada, M. and Samis, M., Year, Complete Decision-Tree Analysis Using Simulation Methods: Illustrated With an Example of Bitumen Production in Alberta Using Steam Injection. SPE Annual Technical Conference and Exhibition

Lawal, A.S., 1999, Application of the Lawal-Lake-SilberbergEquation-of-State to Thermodynamic and Transport Properties of Fluid and Fluid Mixtures. Lubbock: Department of Petroleum Engineering, Texas Tech University. Technical Report TR-4-99.

Lei, H., Pingping, S., Ying, J., Jigen, Y., Shi, L. and Aifang, B., 2010, Prediction of asphaltene precipitation during CO 2 injection. Petroleum Exploration and Development, 37: 349-353.

Li, B., Song, S. and Li, K., 2012, Improved conjugate gradient implementation for least squares support vector machines. Pattern Recognition Letters, 33: 121-125.

Li, H. and Jing, G., Year, The Effect of Pressure on Wax Disappearance Temperature and Wax Appearance Temperature of Water Cut Crude Oil. The Twentieth International Offshore and Polar Engineering Conference

Li, Z., Weida, Z. and Licheng, J., Year, Radar target recognition based on support vector machine. Signal Processing Proceedings, 2000. WCCC-ICSP 2000. 5th International Conference on, 1453-1456.

LINGO-Softwate, 2011, Optimization Modeling Software for Linear, Nonlinear, and Integer Programming, LINDO Systems, LINGO Version 11.

Litani-Barzilai, I., Sela, I., Bulatov, V., Zilberman, I. and Schechter, I., 1997, On-line remote prediction of gasoline properties by combined optical methods. Analytica chimica acta, 339: 193-199.

Liu, S., 2008, Alkaline Surfactant Polymer enhanced oil recovery process, Doctoral Thesis, Rice University. http://hdl. handle. net/1911/22224

Loh, W.Y., 2011, Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1: 14-23.

Loureiro, T.S., Palermo, L.C.M. and Spinelli, L.S., 2015, Influence of precipitation conditions (n-heptane or carbon dioxide gas) on the performance of asphaltene stabilizers. Journal of Petroleum Science and Engineering, 127: 109-114.

M.R. Carlson, W.B.C., 1996, Obtaining PVT data for very sour retrograde condensate gas and volatile oil reservoirs: a multi-disciplinary approach. SPE Gas Technology

Symposium, Calgary, Alberta, Canada.

Ma, C., Randolph, M.A. and Drish, J., Year, A support vector machines-based rejection technique for speech recognition. Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on, 381-384.

Macary, S. and El-Batanoney, M., 1993, Derivation of PVT correlations for the Gulf of Suez crude oils. Sekiyu Gakkai Shi, 36: 472-478.

Maerker, J. and Gale, W., 1992, Surfactant flood process design for Loudon. SPE reservoir engineering, 7: 36-44.

Mahmood, M.A. and Al-Marhoun, M.A., 1996a, Evaluation of empirically derived< i> PVT</i> properties for Pakistani crude oils. Journal of Petroleum Science and Engineering, 16: 275-290.

Mahmood, M.A. and Al-Marhoun, M.A., 1996b, Evaluation of empirically derived PVT properties for Pakistani crude oils. Journal of Petroleum Science and Engineering, 16: 275-290.

Mazandarani, M.T. and Asghari, S.M., 2007, Correlations for Predicting Solution Gas-Oil Ratio, Bubblepoint Pressure and Oil Formation Volume Factor at Bubblepoint of Iran Crude Oils, Copenhagen.

McCain, W.D., 1990, The properties of petroleum fluids. PennWell Books

McLeod, W.R., 1968, Applications of molecular refraction to the principle of corresponding states, University of Oklahoma

Mendes, G., Aleme, H.G. and Barbeira, P.J., 2012, Determination of octane numbers in gasoline by distillation curves and partial least squares regression. Fuel, 97: 131-136.

Metivaud, V., Rajabalee, F., Oonk, H.A., Mondieig, D. and Haget, Y., 1999, Complete determination of the solid (RI)-liquid equilibria of four consecutive n-alkane ternary systems in the range C14H30-C21H44 using only binary data. Canadian journal of chemistry, 77: 332-339.

Miao, Z., 2011, Discussion of optimize method of fire alarm dispatching based on operation research principle. Procedia Engineering, 11: 689-694.

Milhet, M., Pauly, J., Coutinho, J., Dirand, M. and Daridon, J.-L., 2005, Liquid–solid equilibria under high pressure of tetradecane+ pentadecane and tetradecane+ hexadecane binary systems. Fluid Phase Equilibria, 235: 173-181.

Mitchell, M., 1998, An introduction to genetic algorithms (complex adaptive systems).

Modesty Kelechukwu, E., Said Al-Salim, H. and Saadi, A., 2013, Prediction of wax deposition problems of hydrocarbon production system. Journal of Petroleum Science and Engineering, 108: 128-136.

Moghadam, J.N., Salahshoor, K. and Kharrat, R., 2011, Introducing a new method for predicting PVT properties of Iranian crude oils by applying artificial neural networks. Petroleum Science and Technology, 29: 1066-1079.

Mohaghegh, S., Arefi, R., Ameri, S. and Hefner, M.H., A Methodological Approach for Reservoir Heterogeneity Characterization Using Artificial Neural Networks. Society of Petroleum Engineers.

Mohammadi, A.H., Ji, H., Burgass, R.W., Ali, A.B. and Tohidi, B., Year, Gas hydrates in oil systems. SPE Europec/EAGE Annual Conference and Exhibition

Mohammadi, A.H. and Richon, D., 2007, New predictive methods for estimating the vaporization enthalpies of hydrocarbons and petroleum fractions. Industrial & Engineering Chemistry Research, 46: 2665-2671.

Mohammadi, A.H. and Richon, D., 2010, Hydrate phase equilibria for hydrogen+ water and hydrogen+ tetrahydrofuran+ water systems: Predictions of dissociation conditions using an artificial neural network algorithm. Chemical Engineering Science, 65: 3352-3355.

Mohammadi, A.H., Eslamimanesh, A. and Richon, D., 2011, Wax solubility in gaseous system: Thermodynamic consistency test of experimental data. Industrial & Engineering Chemistry Research, 50: 4731-4740.

Mohammadi, A.H., Eslamimanesh, A., Gharagheizi, F. and Richon, D., 2012a, A novel method for evaluation of asphaltene precipitation titration data. Chemical Engineering Science, 78: 181-185.

Mohammadi, A.H., Eslamimanesh, A. and Richon, D., 2012b, Monodisperse Thermodynamic Model Based on Chemical+ Flory–Hüggins Polymer Solution Theories for Predicting Asphaltene Precipitation. Industrial & Engineering Chemistry Research, 51: 4041-4055.

Mohammadi, A.H., Gharagheizi, F., Eslamimanesh, A. and Richon, D., 2012c, Evaluation of experimental data for wax and diamondoids solubility in gaseous systems. Chemical Engineering Science.

Mohammadi, A.H., Gharagheizi, F., Eslamimanesh, A. and Richon, D., 2012d, Evaluation of experimental data for wax and diamondoids solubility in gaseous systems. Chemical Engineering Science, 81: 1-7.

Monger, T. and Trujillo, D., 1991, Organic deposition during CO2 and rich-gas flooding. SPE reservoir engineering, 6: 17-24.

Montgomery, D.C., 2008, Design and analysis of experiments. John Wiley & Sons

Moradi, G., Khoshmaram, A. and Riazi, M., 2011, Estimation of properties distribution of C 7+ by using artificial neural networks. Journal of Petroleum Science and Engineering, 76: 57-62.

Moradi, G., Mohadesi, M. and Mokhtari, M., 2013a, A New Correlation for Prediction of Wax Disappearance Temperature of Hydrocarbon Mixtures at Various Pressures. Journal of Chemical and Petroleum Engineering, 47: 27-38.

Moradi, G., Mohadesi, M. and Moradi, M.R., 2013b, Prediction of wax disappearance temperature using artificial neural networks. Journal of Petroleum Science and Engineering, 108: 74-81.

Murty, B. and Rao, R., 2004, Global optimization for prediction of blend composition of gasolines of desired octane number and properties. Fuel Processing Technology, 85: 1595-1602.

Naseri, A., Nikazar, M. and Mousavi Dehghani, S., 2005, A correlation approach for prediction of crude oil viscosities. Journal of Petroleum Science and Engineering, 47: 163-174.

Nejatian, I., Kanani, M., Arabloo, M., Bahadori, A. and Zendehboudi, S., 2014, Prediction of natural gas flow through chokes using support vector machine algorithm. Journal of Natural Gas Science and Engineering, 18: 155-163.

Nemeth, L. and Kennedy, H., 1967, A correlation of Dewpoint pressure with fluid composition and temperature. SPE Journal, 7: 99-104.

Nemeth, L.K., 1966, A correlation of dew-point pressure with reservoir fluid composition and temperature, Texas A & M University.

Novosad, J., Baxter, L. and Parker, G., 1981, Surfactant Retention on Berea Sandstone: Effects of Phase Behavior and Temperature. Petroleum Recovery Institute

Nowroozi, S., Ranjbar, M., Hashemipour, H. and Schaffie, M., 2009, Development of a neural fuzzy system for advanced prediction of dew point pressure in gas condensate reservoirs. Fuel Processing Technology, 90: 452-457.

Ø. Fevang , C.H.W., 1996, Modeling gas-condensate well deliverability SPE reservoir engineering, 11: 221-230

Obanijesu, E. and Araromi, D., 2008, Predicting bubble-point pressure and formation-volume factor of Nigerian crude oil system for environmental sustainability. Petroleum Science and Technology, 26: 1993-2008.

Obanijesu, E. and Omidiora, E., 2008, Artificial neural network's prediction of wax deposition potential of Nigerian crude oil for pipeline Safety. Petroleum Science and Technology, 26: 1977-1991.

Obanijesu, E. and Omidiora, E., 2009, The artificial neural network's prediction of crude oil viscosity for pipeline safety. Petroleum Science and Technology, 27: 412-426.

Obomanu, D. and Okpobiri, G., 1987, Correlating the PVT properties of Nigerian crudes. Journal of energy resources technology, 109: 214-217.

Omar, M. and Todd, A., Year, Development of new modified black oil correlations for Malaysian crudes. SPE Asia Pacific oil and gas conference

Orbey, H. and Sandler, S.I., 1993, The prediction of the viscosity of liquid hydrocarbons and their mixtures as a function of temperature and pressure. The Canadian Journal of Chemical Engineering, 71: 437-446.

Osei-Bryson, K.-M., 2004, Evaluation of decision trees: a multi-criteria approach. Computers & Operations Research, 31: 1933-1945.

Ostermann, R., Ehlig-Economides, C. and Owolabi, O., 1983, Correlations for the reservoir fluid properties of Alaskan crudes. in, University of Alaska.

Ostermann, R. and Owolabi, O., Year, Correlations for the reservoir fluid properties of Alaskan crudes. SPE California Regional Meeting

Parhizgar, H., Dehghani, M.R. and Eftekhari, A., 2013, Modeling of vaporization enthalpies of petroleum fractions and pure hydrocarbons using genetic programming. Journal of Petroleum Science and Engineering, 112: 97-104.

Parsa, S., Javanmardi, J., Aftab, S. and Nasrifar, K., 2014, Experimental measurements and thermodynamic modeling of wax disappearance temperature for the binary systems nC< sub> 14</sub> H< sub> 30</sub>+ n-C< sub> 16</sub> H< sub> 34</sub>, nC< sub> 16</sub> H< sub> 34</sub>+ n-C< sub> 18</sub> H< sub>

38</sub> and nC< sub> 11</sub> H< sub> 24</sub>+ n-C< sub> 18</sub> H< sub> 38</sub>. Fluid phase equilibria.

Patel, N.C. and Teja, A.S., 1982, A new cubic equation of state for fluids and fluid mixtures. Chemical Engineering Science, 37: 463-473.

Pelckmans, K., Suykens, J.A.K., Van Gestel, T., De Brabanter, J., Lukas, L., Hamers, B., De Moor, B. and Vandewalle, J., 2002, LS-SVMlab: a Maltab/c toolbox for least squares support vector machines. Tutorial. KULeuven-ESAT. Leuven, .

Peng, D.-Y. and Robinson, D.B., 1976, A new two-constant equation of state. Industrial & Engineering Chemistry Fundamentals, 15: 59-64.

Petrosky, G.E., 1990, PVT correlations for gulf of mexico crude oils, University of Southwestern Louisiana

Petrosky Jr, G. and Farshad, F., Year, Pressure-volume-temperature correlations for Gulf of Mexico crude oils. SPE Annual Technical Conference and Exhibition

Petrosky Jr, G. and Farshad, F., 1998, Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils. SPE Reservoir Evaluation & Engineering, 1: 416-420.

R. Olds, B.S., W. Lacey, , () 1945, Volumetric and phase behavior of oil and gas from Paloma

Field, Trans. AIME, 160: 77-99.

R. Olds, B.S., W. Lacey, , () 1949, Volumetric and viscosity studies of oil and gas from a San Joaquin valleyfield. Transactions of AIME, 179: 287.

R. Sarkar, A.D., A. Todd,, 1991, Phase behavior modeling of gas-condensatefluids

using an equation of state,. SPE Annual Technical Conference and Exhibition.

Rafiee-Taghanaki, S., Arabloo, M., Chamkalani, A., Amani, M., Zargari, M.H. and Adelzadeh, M.R., 2013, Implementation of SVM framework to estimate PVT properties of reservoir oil. Fluid Phase Equilibria, 346: 25-32.

Rafiq, M., Bugmann, G. and Easterbrook, D., 2001, Neural network design for engineering applications. Computers & Structures, 79: 1541-1552.

Rahimzadeh Kivi, I., Ameri Shahrabi, M. and Akbari, M., 2013, The development of a robust ANFIS model for predicting minimum miscibility pressure. Petroleum Science and Technology, 31: 2039-2046.

Rassamdana, H., Dabir, B., Nematy, M., Farhani, M. and Sahimi, M., 1996, Asphalt flocculation and deposition: I. The onset of precipitation. AIChE journal, 42: 10-22.

Rassamdana, H. and Sahimi, M., 1996, Asphalt flocculation and deposition: II. Formation and growth of fractal aggregates. AIChE journal, 42: 3318-3332.

Redlich, O. and Kwong, J., 1949, On the Thermodynamics of Solutions. V. An Equation of State. Fugacities of Gaseous Solutions. Chemical reviews, 44: 233-244.

Reid, R.C., Prausnitz, J.M. and Poling, B.E., 1987, The properties of gases and liquids.

Riazi, M. and Daubert, T., 1980, Simplify property predictions. Hydrocarbon Processing, 60: 115-116.

Riazi, M.R., 1989, Distribution model for properties of hydrocarbon-plus fractions. Industrial & Engineering Chemistry Research, 28: 1731-1735.

Riazi, M.R. and Al-Sahhaf, T.A., 1996, Physical properties of heavy petroleum fractions and crude oils. Fluid Phase Equilibria, 117: 217-224.

Riazi, M.R., 1997, A continuous model for C7+ fraction characterization of petroleum fluids. Industrial & Engineering Chemistry Research, 36: 4299-4307.

Robinson Jr, R. and Jacoby, R., 1965, Better compressibility factors. Hydrocarbon Processing, 44: 141-145.

Robles, L., Espeau, P., Mondieig, D., Haget, Y. and Oonk, H., 1996, Polymorphism and Molecular Alloys in the Binary-System C17H36-C19H40. Thermochimica Acta, 274: 61-72.

Rollins, J.B., McCain, W.D., Jr. and Creeger, T.J., 1990, Estimation of Solution GOR of Black Oils. Journal of Petroleum Technology, 42: 92-94.

Rousseeuw, P.J. and Leroy, A.M., 2005, Robust regression and outlier detection. Wiley. com

Saemi, M., Ahmadi, M. and Varjani, A.Y., 2007, Design of neural networks using genetic algorithm for the permeability estimation of the reservoir. Journal of Petroleum Science and Engineering, 59: 97-105.

Saleh, A., Mahgoub, I. and Asaad, Y., 1987, Evaluation of Empirically Derived PVT Properties for Egyptians Oils. Middle East Oil Show.

Sanjari, E. and Lay, E.N., 2012a, Estimation of natural gas compressibility factors using artificial neural network approach. Journal of Natural Gas Science and Engineering, 9: 220-226.

Sanjari, E. and Lay, E.N., 2012b, An accurate empirical correlation for predicting natural gas compressibility factors. Journal of Natural Gas Chemistry, 21: 184-188.

Sethi, I.K. and Chatterjee, B., 1977, Efficient decision tree design for discrete variable pattern recognition problems. Pattern Recognition, 9: 197-206.

Seyed Mohammad Javad Majidi , A.S., Milad Arabloo , Ramin Mahdikhani-Soleymanloo , Mohsen Masihi, 2013, Evolving an accurate model based on machine learning approach for prediction of dew-point pressure in gas condensate reservoirs. chemical engineering research and design: 1347-1359.

Shafiei, A., Dusseault, M.B., Zendehboudi, S. and Chatzis, I., 2013, A new screening tool for evaluation of steamflooding performance in Naturally Fractured Carbonate Reservoirs. Fuel, 108: 502-514.

Shafiie, Moghadasi, Shahbazian and Zargani, 2012, Optimization of Formation Volume Factor and Solution Gas-Oil Ratio Correlations for Southern Iranian Oilfields Using Genetic Algorithm. Journal of American Science.

Shahebrahimi, Y. and Zonnouri, A., 2013, A new combinatorial thermodynamics model for asphaltene precipitation. Journal of Petroleum Science and Engineering, 109: 63-69.

Shateri, M., Ghorbani, S., Hemmati-Sarapardeh, A. and Mohammadi, A.H., 2014, Application of Wilcoxon generalized radial basis function network for prediction of natural gas compressibility factor. Journal of the Taiwan Institute of Chemical Engineers.

Shi, Y. and Eberhart, R., Year, A modified particle swarm optimizer. Evolutionary Computation Proceedings, 1998. IEEE World Congress on Computational Intelligence., The 1998 IEEE International Conference on, 69-73.

Shokir, E.M.E.-M., 2008, Dewpoint Pressure Model for Gas Condensate Reservoirs Based on Genetic Programming. Energy & Fuels, 22: 3194-3200.

Shokir, E.M.E.-M., El-Awad, M.N., Al-Quraishi, A.A. and Al-Mahdy, O.A., 2012, Compressibility factor model of sweet, sour, and condensate gases using genetic programming. Chemical Engineering Research and Design, 90: 785-792.

Shokrollahi, A., Arabloo, M., Gharagheizi, F. and Mohammadi, A.H., 2013, Intelligent model for prediction of CO< sub> 2</sub>–Reservoir oil minimum miscibility pressure. Fuel.

Simon, R. and Briggs, J.E., 1964, Application of Benedict-Webb-Rubin equation of state to hydrogen sulfide-hydrocarbon mixtures. AIChE journal, 10: 548-550.

Soave, G., 1972, Equilibrium constants from a modified Redlich-Kwong equation of state. Chemical Engineering Science, 27: 1197-1203.

Solairaj, S., 2011, New method of predicting optimum surfactant structure for EOR. MSc Thesis. The University of Texas at Austin.

Solairaj, S., Britton, C., Kim, D.H., Weerasooriya, U. and Pope, G.A., Year, Measurement and Analysis of Surfactant Retention. SPE Improved Oil Recovery Symposium

Soorghali, F., Zolghadr, A. and Ayatollahi, S., 2014, Effect of resins on asphaltene deposition and the changes of surface properties at different pressures: a microstructure study. Energy & Fuels, 28: 2415-2421.

Srinivas, H., Srinivasan, K. and Umesh, K., 2010, Application of Artificial Neural Network and Wavelet Transform for Vibration Analysis of Combined Faults of Unbalances and Shaft Bow. Adv. Theor. Appl. Mech, 3: 159-176.

Srivastava, R., Huang, S. and Dong, M., 1999, Asphaltene deposition during CO2 flooding. SPE production & facilities, 14: 235-245.

Standing, M., 1947a, A pressure-volume-temperature correlation for mixtures of California oils and gases. Drilling and Production Practice: 275-285.

Standing, M., 1947b, A pressure-volume-temperature correlation for mixtures of California oils and gases. Drilling and Production Practice.

Standing, M.B. and Katz, D.L., 1942, Density of natural gases. Trans. AIME, 146: 140-149.

Standnes, D.C. and Austad, T., 2000, Wettability alteration in chalk: 2. Mechanism for wettability alteration from oil-wet to water-wet using surfactants. Journal of Petroleum Science and Engineering, 28: 123-143.

Sutton, R.P. and Farshad, F., 1990, Evaluation of empirically derived PVT properties for Gulf of Mexico crude oils. SPE reservoir engineering, 5: 79-86.

Suykens, J.A. and Vandewalle, J., 1999, Least squares support vector machine classifiers. Neural processing letters, 9: 293-300.

Suykens, J.A., Vandewalle, J. and De Moor, B., 2001, Intelligence and cooperative search by coupled local minimizers. International Journal of Bifurcation and Chaos, 11: 2133-2144.

Suykens, J.A., De Brabanter, J., Lukas, L. and Vandewalle, J., 2002a, Weighted least squares support vector machines: robustness and sparse approximation. Neurocomputing, 48: 85-105.

Suykens, J.A.K., Van Gestel, T., De Brabanter, J., De Moor, B. and Vandewalle, J., 2002b, Least Squares Support Vector Machines. World Scientific Publishing Company.

Taber, J., Martin, F. and Seright, R., 1997, EOR screening criteria revisited-Part 1: Introduction to screening criteria and enhanced recovery field projects. SPE reservoir engineering, 12: 189-198.

Taghanaki, S.R., Arabloo, M., Chamkalani, A., Amani, M., Zargari, M.H. and Adelzadeh, M.R., 2013, Implementation of SVM framework to estimate PVT properties of reservoir oil. Fluid Phase Equilibria.

Taghaz, A., Eltaeb, N. and Alakhdar, S., 2008, Comparison Study of Published PVT Correlations and Its Application to Estimate Reservoir Fluid Properties for Libyan Oil Reservoirs, Libya.

Tahmasebi, P. and Hezarkhani, A., 2012, A fast and independent architecture of artificial neural network for permeability prediction. Journal of Petroleum Science and Engineering, 86: 118-126.

Talebi, R., Ghiasi, M.M., Talebi, H., Mohammadyian, M., Zendehboudi, S., Arabloo, M. and Bahadori, A., 2014, Application of soft computing approaches for modeling saturation pressure of reservoir oils. Journal of Natural Gas Science and Engineering, 20: 8-15.

Tan, P., Steinbach, M. and Kumar, V., 2006, Introduction to data mining. Classification: basic concepts, decision trees and model evaluation. Addison–Wesley, Boston.

Teixeira, L.S., Souza, J.C., dos Santos, H.C., Pontes, L.A., Guimarães, P.R., Sobrinho, E.V. and Vianna, R.F., 2007, The influence of Cu, Fe, Ni, Pb and Zn on gum formation in the Brazilian automotive gasoline. Fuel Processing Technology, 88: 73-76.

Teixeira, L.S., Dantas, M.S., Guimarães, P.R., Teixeira, W., Vargas, H. and Lima, J.A., 2009, Correlation of PVR, Octane Numbers and Distillation Curve of Gasoline with Data from a Thermal Wave Interferometer. Computer Aided Chemical Engineering, 27: 759-764.

Teja, A. and Rice, P., 1981, Generalized corresponding states method for the viscosities of liquid mixtures. Industrial & Engineering Chemistry Fundamentals, 20: 77-81.

Teodorescu, L. and Sherwood, D., 2008, High energy physics event selection with gene expression programming. Computer Physics Communications, 178: 409-419.

Torabi, F., Abedini, A. and Abedini, R., 2011, The development of an artificial neural network model for prediction of crude oil viscosities. Petroleum Science and Technology, 29: 804-816.

Twu, C.H., 1984, An internally consistent correlation for predicting the critical properties and molecular weights of petroleum and coal-tar liquids. Fluid Phase Equilibria, 16: 137-150.

Vafaie-Sefti, M., Mousavi-Dehghani, S.A. and Bahar, M.M.-Z., 2000, Modification of multisolid phase model for prediction of wax precipitation: a new and effective solution method. Fluid Phase Equilibria, 173: 65-80.

Valderrama, J.O., 2003, The state of the cubic equations of state. Industrial & Engineering Chemistry Research, 42 (8), 1603-1618.

Valko, P. and McCain Jr, W., 2003, Reservoir oil bubblepoint pressures revisited; solution gas–oil ratios and surface gas specific gravities. Journal of Petroleum Science and Engineering, 37: 153-169.

van der Waals, J.D., 2004, On the continuity of the gaseous and liquid states. DoverPublications. com

Van Gestel, T., Suykens, J.A., Baestaens, D.-E., Lambrechts, A., Lanckriet, G., Vandaele, B., De Moor, B. and Vandewalle, J., 2001, Financial time series prediction using least squares support vector machines within the evidence framework. Neural Networks, IEEE Transactions on, 12: 809-821.

Vasan, A. and Raju, K.S., 2009, Comparative analysis of simulated annealing, simulated quenching and genetic algorithms for optimal reservoir operation. Applied Soft Computing, 9: 274-281.

Vazquez, M. and Beggs, H.D., 1980, Correlations for fluid physical property prediction. Journal of Petroleum Technology, 32: 968-970.

Velarde, J., Blasingame, T. and McCain Jr, W., Year, Correlation of Black Oil Properties At Pressures Below Bubble Point Pressure-A New Approach. Annual Technical Meeting

Velarde, J., Blasingame, T. and McCain Jr, W., 1999, Correlation of black oil properties at pressures below bubblepoint pressure–a new approach. Journal of Canadian Petroleum Technology, 36: 1-6.

Vetere, A., 1979, New correlations for predicting vaporization enthalpies of pure compounds. The Chemical Engineering Journal, 17: 157-162.

Vetere, A., 1995, Methods to predict the vaporization enthalpies at the normal boiling temperature of pure compounds revisited. Fluid Phase Equilibria, 106: 1-10.

Vidal, D., Blobel, J., Pérez, Y., Thormann, M. and Pons, M., 2007, Structure-based discovery of new small molecule inhibitors of low molecular weight protein tyrosine phosphatase. European journal of medicinal chemistry, 42: 1102-1108.

Wang, K.-S., Wu, C.-H., Creek, J.L., Shuler, P.J. and Tang, Y., 2003, Evaluation of effects of selected wax inhibitors on paraffin deposition. Petroleum science and technology, 21: 369-379.

Wang, X. and Economides, M.J., 2009, Advanced Natural Gas Engineering. Gulf Publishing Company Houston, TX

Wang, X., Liu, X., Pedrycz, W. and Zhang, L., 2015, Fuzzy rule based decision trees. Pattern Recognition, 48: 50-59.

Whitson, C. and Torp, S., Year, Evaluating constant volume depletion data. SPE Annual Technical Conference and Exhibition

Whitson, C.H., 1983, Characterizing hydrocarbon plus fractions. Society of Petroleum Engineers Journal, 23: 683-694.

Whitson, C.H., Soreide, I. and Anderson, T., 1989, C 7+ characterization of related equilibrium fluids using the gamma distribution, Miami, FL.

Wichert, E. and Aziz, K., 1972, Calculate Z's for sour gases. Hydrocarbon Processing, 51: 119-122.

Won, K., 1989, Thermodynamic calculation of cloud point temperatures and wax phase compositions of refined hydrocarbon mixtures. Fluid phase equilibria, 53: 377-396.

Xavier-de-Souza, S., Suykens, J.A., Vandewalle, J. and Bollé, D., 2010, Coupled simulated annealing. Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 40: 320-335.

Xu, Z., Zhang, J., Feng, Z., Fang, W. and Wang, F., 2010, Characteristics of remaining oil viscosity in water-and polymer-flooding reservoirs in Daqing Oilfield. Science in China Series D: Earth Sciences, 53: 72-83.

Yan, K.-L., Liu, H., Sun, C.-Y., Ma, Q.-L., Chen, G.-J., Shen, D.-J., Xiao, X.-J. and Wang, H.-Y., 2013, Measurement and calculation of gas compressibility factor for condensate gas and natural gas under pressure up to 116 MPa. The Journal of Chemical Thermodynamics.

Yang, J., Bouzerdoum, A. and Phung, S.L., Year, A training algorithm for sparse LS-SVM using compressive sampling. Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on, 2054-2057.

Yassin, M.R., Arabloo, M., Shokrollahi, A. and Mohammad, A.H., 2013, Prediction of Surfactant Retention in Porous Media: A Robust Modeling Approach. Journal of Dispersion Science and Technology.

Yi, X., Year, Using wellhead sampling data to predict reservoir saturation pressure. SPE Permian Basin Oil and Gas Recovery Conference

Zabihi, R., Schaffie, M., Nezamabadi-Pour, H. and Ranjbar, M., 2011, Artificial neural network for permeability damage prediction due to sulfate scaling. Journal of Petroleum Science and Engineering, 78: 575-581.

Zanganeh, P., Dashti, H. and Ayatollahi, S., 2015, Visual investigation and modeling of asphaltene precipitation and deposition during CO 2 miscible injection into oil reservoirs. Fuel.

Zendehboudi, S., Mohammadzadeh, O. and Chatzis, I., Year, Experimental study of controlled gravity drainage in fractured porous media. Canadian International Petroleum Conference

Zendehboudi, S., Chatzis, I., Mohsenipour, A.A. and Elkamel, A., 2011, Dimensional analysis and scale-up of immiscible two-phase flow displacement in fractured porous media under controlled gravity drainage. Energy & Fuels, 25: 1731-1750.

Zendehboudi, S., Ahmadi, M.A., Bahadori, A., Shafiei, A. and Babadagli, T., 2013a, A developed smart technique to predict minimum miscible pressure—eor implications. The Canadian Journal of Chemical Engineering, 91: 1325-1337.

Zendehboudi, S., Shafiei, A., Bahadori, A., James, L.A., Elkamel, A. and Lohi, A., 2013b, Asphaltene precipitation and deposition in oil reservoirs–Technical aspects, experimental and hybrid neural network predictive tools. Chemical Engineering Research and Design.

Zendehboudi, S., Shafiei, A., Bahadori, A., James, L.A., Elkamel, A. and Lohi, A., 2014, Asphaltene precipitation and deposition in oil reservoirs–Technical aspects, experimental and hybrid neural network predictive tools. Chemical Engineering Research and Design, 92: 857-875.

Zhang, D., Liu, S., Yan, W., Puerto, M., Hirasaki, G.J. and Miller, C.A., Year, Favorable attributes of alkali-surfactant-polymer flooding. SPE/DOE Symposium on Improved Oil Recovery

Zuo, J.Y., Zhang, D.D. and Ng, H.-J., 2001, An improved thermodynamic model for wax precipitation from petroleum fluids. Chemical Engineering Science, 56: 6941-6947.

APPENDIX A

Least Square Support Vector Machine Algorithm

The MATLAB code developed in this study for the least square support vector machine algorithm is presented as follows. In the beginning, the original LSSVM toolbox for MATLAB should be installed, after that the directory of the LSSVM toolbox should be inserted as the main directory in the MATLAB environment. The encoded MATLAB file (VE.m) is available upon reader request. Following example provides a step-by-step instruction for using the proposed models.

Example: Calculation of vaporization enthalpy (VE) using the input variables (*e.g.* boiling point, specific gravity, and molecular weight). Then, vaporization enthalpy is calculated easily applying the below codes in the command window:

```
clc;clear;

Data=[355.5 0.7015 95]; %Input vector

%% Prediction of LS-SVM model based on kernel function

%% Calculation of VE

load 'VE.mat'

VE_clac = simlssvm({trainX,trainY,type,gam,sig2,'RBF_kernel','preprocess'},{alpha,b},Data) % Calculated

VE
```

APPENDIX B

Artificial Neural Network Algorithm

The MATLAB code developed in this study for the artificial neural network algorithm is presented as follows:

```
%% Start of Program
clc
clear
%close all

%% Data Loading
Data = xlsread('Data.xls');

X = Data(:,1:end-1);
Y = Data(:,end);

DataNum = size(X,1);
InputNum = size(X,2);
OutputNum = size(Y,2);

%% Normalization
MinX = min(X);
MaxX = max(X);

MinY = min(Y);
MaxY = max(Y);
```

```
XN = X;

YN = Y;


for ii = 1:InputNum

   XN(:,ii) = Normalize_Fcn(X(:,ii),MinX(ii),MaxX(ii));

end


for ii = 1:OutputNum

   YN(:,ii) = Normalize_Fcn(Y(:,ii),MinY(ii),MaxY(ii));

end


%% Test and Train Data

TrPercent = 80;

TrNum = round(DataNum * TrPercent / 100);

TsNum = DataNum - TrNum;


R = randperm(DataNum);

trIndex = R(1 : TrNum);

tsIndex = R(1+TrNum : end);


Xtr = XN(trIndex,:);

Ytr = YN(trIndex,:);


Xts = XN(tsIndex,:);

Yts = YN(tsIndex,:);


%% Network Structure

pr = [-1 1];
```

```
PR = repmat(pr,InputNum,1);


Network = newff(PR,[5 OutputNum],{'tansig' 'tansig'});


%% Training

Network = TrainUsing_PSO_Fcn(Network,Xtr,Ytr);


%% Assesment

YtrNet = sim(Network,Xtr')';

YtsNet = sim(Network,Xts')';


MSEtr = mse(YtrNet - Ytr)

MSEts = mse(YtsNet - Yts)


%% Denormalization


Denormalization_Ytr = (Ytr./2+0.5) * (MaxY-MinY) + MinY;

Denormalization_Yts = (Yts./2+0.5) * (MaxY-MinY) + MinY;

Denormalization_YtrNet = (YtrNet./2+0.5) * (MaxY-MinY) + MinY;

Denormalization_YtsNet = (YtsNet./2+0.5) * (MaxY-MinY) + MinY;


%% Display

figure(1)

plot(Ytr,'-or');

hold on

plot(YtrNet,'-sb');

hold off
```

```
figure(2)
plot(Yts,'-or');
hold on
plot(YtsNet,'-sb');
hold off

figure(3)
t = -1:.1:1;
plot(t,t,'b','linewidth',2)
hold on
plot(Ytr,YtrNet,'ok')
hold off

figure(4)
t = -1:.1:1;
plot(t,t,'b','linewidth',2)
hold on
plot(Yts,YtsNet,'ok')
hold off
```

APPENDIX C

Decision Tree Algorithm

The MATLAB code developed in this study for the regression decision tree algorithm is presented as follows:

```
clc

clf

A=xlsread('wax');

x=A(:,1:2);

y=A(:,3);

tree = fitrtree(x,y);

Yfit = predict(tree,x)
```

APPENDIX D

Gene Expression Programming Algorithm

The MATLAB code developed in this study for the gene expression programming algorithm associated with each property studied is presented as follows.

Surfactant Retention:

```
G1C1 = -2.23953344600004;
G1C5 = -6.4152152106317;
G1C3 = -8.57843269407123;
G2C7 = 9.34891432853352;
G3C5 = 0.622676265417228;
G3C3 = -5.28553342496036;
G3C4 = -30.7231165312331;

syms K;
syms TAN;
syms T;
syms Co_solvent;
syms Polymer;
syms MaxPH;
syms Mobility;
syms MW;

TAN_10 = TAN*10;
Co_solvent_1E3 = Co_solvent*1E3;
Mobility_1E2 = Mobility*1E2;
```

```
Polymer_1E_3 = Polymer*1E-3;
K_1E_1 = K*1E-1;

y = 0.0;

y = (((Co_solvent_1E3)+((G1C5+G1C3)^2))+((G1C1*(T))/((K_1E_1)-(MaxPH))));
y = y + ((-((-(((((T)^2)*sqrt((MW)))*(G2C7-(MaxPH))))))))^(1/3);
y = y + ((sqrt(((Mobility_1E2)-(Co_solvent_1E3)-G3C4))-(G3C5/(Polymer_1E_3)))*((-(G3C3))-
(TAN_10)));
```

Dew Point Pressure:

```
G1C1 = 1367.41291856528;
G1C0 = 1.02137231558021;
G1C2 = 1909.69508955175;
G1C4 = 358.093397489431;
G1C3 = 933.348585036562;
G2C0 = 1.024023;
G2C1 = -1697.64421139709;
G2C3 = 9.97999003998998;
G2C2 = -5096.84992009682;
G3C2 = 13.1454000006988;
G3C0 = 9.43977272271795e-02;
G3C1 = 3.19035556629968e-03;
G3C3 = 1961.70118405183;
G3C4 = 2096.95260530625;
G3C6 = 39335.0733508395;
G3C5 = 201.934160542236;
G4C0 = 8.56837764914087e-02;
G4C2 = -6.52240974576734e-02;
G4C1 = 3.31393632912371e-03;
G4C3 = -4.94222387656948;
G4C4 = -3451.17091475473;
G4C5 = -6212.70961095208;

d(1)=sym('T');
T = 1;

d(2)=sym('C1');
```

```
C1 = 2;

d(3)=sym('C2');
C2 = 3;

d(4)=sym('C3');
C3 = 4;

d(5)=sym('C4');
C4 = 5;

d(6)=sym('C5');
C5 = 6;

d(7)=sym('C6');
C6 = 7;

d(8)=sym('C7');
C7 = 8;

d(9)=sym('N2');
N2 = 9;

d(10)=sym('CO2');
CO2 = 10;

d(11)=sym('H2S');
H2S = 11;

d(12)=sym('SGC7');
SGC7 = 12;

d(13)=sym('MWC7');
MWC7 = 13;

y = 0.0;
```

```
y =
(((((log(d(C7))*G1C4)+(G1C3*d(CO2)))+(G1C2*d(N2)))+G1C1+((((d(C1)*d(MWC7))*G2C3)+(G2C2*d(C7
)))+(G2C1*d(C3))))/(G1C0-d(SGC7)));
y = y +
(((((((d(C4)^2)*G3C6)+(G3C5*d(H2S)))+(G3C4*d(C5)))+(G3C3*d(C2)))+G3C2+((((G4C4*d(C6))+(G4C5*
d(C4)))+(G4C3*d(C1)))+(G4C2*d(T))))/((G3C1*d(T))+G3C0));
```

Bubble Point Pressure:

```
G1C7 = -31.5611233086022;

G1C4 = 45.3186644685036;

G2C6 = -6.6033163791293;

G2C5 = 4.84791197058782;

G3C3 = 7.77581429969672;


d(1)=sym('Rsi');

Rsi = 1;


d(2)=sym('GG');

GG = 2;


d(3)=sym('API');

API = 3;


d(4)=sym('TR');

TR = 4;
```

```
y = 0.0;

y = (((d(Rsi)+d(Rsi)+d(Rsi))+(d(API)+d(API)+d(Rsi))+d(Rsi))+(G1C7/d(GG))+(G1C4*d(GG)));
y = y * (((G2C5^2)-(d(GG)*d(GG)*d(GG)))-((d(Rsi)/d(TR))+G2C6));
y = y * (1.0/(((log10(d(TR))^2)+sqrt(G3C3)+(d(GG)*d(API)))));
```

Oil Formation Volume Factor:

```
G1C9 = 4.90321382405149;
G1C4 = 6.94610229569658;
G1C6 = 9.40509870669927;
G2C7 = 2.73797223607466;
G2C2 = -2.65549676704016;
G3C7 = 4.84603411969359;

d(1)=sym('Rsi');
Rsi = 1;

d(2)=sym('GG');
GG = 2;

d(3)=sym('API');
API = 3;

d(4)=sym('TR');
TR = 4;
```

```
y = 0.0;


y = ((1.0/(((G1C9*G1C4)+(G1C6^2))))*d(GG));
y = y * (power((d(API)+d(TR)),(1.0/(G2C7)))+(1.0/((G2C2/d(GG)))));
y = y * (((d(Rsi)-G3C7)/sqrt(d(GG)))+sqrt((d(TR)*d(API))));
y = y/100 + 1;
```

Oil Formation Volume Factor:
Dead Oil Viscosity:

```
G1C0 = 20358822.3876518;

G1C1 = 614.823948580112;

G1C2 = 63528.8729420635;

G2C0 = 482063.84826673;



d(1)=sym('T');

T = 1;



d(2)=sym('API');

API = 2;



y = 0.0;
```

```
y = ((((d(T)*d(API))*G1C1)-(d(T)*G1C2))+G1C0);

y = y * (1.0/((((d(API)^3)*d(T))-G2C0)));
```

Saturated Oil Viscosity:

```
G1C0 = 7.97272338586754e-02;

G1C1 = 0.707434125153867;

G1C2 = 14.1932041932374;

G2C0 = 1.92690208295064;

G2C2 = 9.05243619825834;

G2C1 = 3.5226945752933;

G2C3 = 0.13056267156335;



d(1)=sym('Pb');

Pb = 1;



d(2)=sym('DOV');

DOV = 2;
```

```
y = 0.0;
```

```
y = (power(G1C1,(G1C2/(d(DOV)^2)))+G1C0);
```

```
y = y + ((G2C0+(G2C1*d(DOV)))/(G2C2+(d(Pb)-(G2C3*d(DOV)))));
```

Saturated Oil Viscosity:
Under-Saturated Oil Viscosity:

```
G1C0 = 1.11496172357196e-02;
```

```
G2C0 = 10.7121633356412;
```

```
G2C2 = 10.9261044638846;
```

```
G2C1 = 0.000999897000303;
```

```
G2C3 = 7.93724070371478e-04;
```

```
G2C4 = 1.19885535429917e-08;
```

```
d(1)=sym('Pb');
```

```
Pb = 1;
```

```
d(2)=sym('SOV');
```

```
SOV = 2;
```

```
d(3)=sym('P');
```

```
P = 3;
```

```
y = 0.0;

y = ((G1C0*d(P))/d(Pb));

y = y +

(((G2C2*d(SOV))+((G2C3*(d(SOV)*d(P)))+(G2C4*((d(SOV)^2)*(d(P)^2)))))/(G2C0+(G2C1*d(Pb))));
```

Asphaltene Precipitation:

```
G1C1 = -188.762283858578;

G1C7 = -1.10521967470357;

G1C2 = 33.3016661393978;

G1C6 = 6.12109132450331;



d(1)=sym('T');

T = 1;

d(2)=sym('Rv');

Rv = 2;

d(3)=sym('MW');

MW = 3;



y = 0.0;
```

```
y = (((G1C1+d(3))*(d(2)/G1C7))/((G1C2+d(1))+(G1C6*d(2))));
```

Heptane-Plus fractions Properties (Molecular Weight):

```
G1C8 = 11.3703978820953;
G1C0 = -12.1871296215731;



d(1)=sym('CWf');
CWf = 1;

d(2)=sym('MWAvg');
MWAvg = 2;

d(3)=sym('SGAvg');
SGAvg = 3;



y = 0.0;

y = (((d(2)+G1C8)-power(d(3),G1C0))*((d(1))^(1/3)+(d(1)^2)));
```

Heptane-Plus fractions Properties (Specific Gravity):

```
G1C6 = 17.0636230676328;
G1C1 = -5.38069093905454;
G1C3 = -3.19434379244057;



d(1)=sym('CWf');
CWf = 1;

d(2)=sym('MWAvg');
MWAvg = 2;
```

```
d(3)=sym('SGAvg');
SGAvg = 3;



y = 0.0;

y = power(((G1C6/d(1))*(d(3)/d(1))),((G1C1+G1C3)/(d(2)+d(3))));
```

Heptane-Plus fractions Properties (Boiling Point):

```
G1C9 = 5.63157492202521;
G1C0 = 1.43944467574612;
G1C2 = -396.286225847496;


d(1)=sym('CWf');
CWf = 1;

d(2)=sym('MWAvg');
MWAvg = 2;

d(3)=sym('SGAvg');
SGAvg = 3;



y = 0.0;

y = (((d(1)*d(2))*(G1C9/G1C0))+((d(1)/d(3))-(G1C2-d(2))));
```

Gas Compressibility Factor:

```
clc
clear all

a=xlsread('z-factor');
Tpr=a(:,1);
Ppr=a(:,2);
z_exp=a(:,3);

a  =  0.2625136  ;
```

```
b  =  3.126365119 ;
c  =  0.56388785  ;
d  =  -3.891636887   ;
e  =  -0.33725254 ;
f  =  -1.397645228   ;
g  =  1.055176333 ;
h  =  0.061688046 ;
i  =  0.447935053 ;
j  =  0.521752161 ;

z=a+b./Tpr+c*log(Ppr)+d./Tpr.^2+e*(log(Ppr)).^2+f*(log(Ppr))./Tpr+g./Tpr.^3+h*(log(Ppr)).^3+i*(log(
Ppr)).^2./Tpr+j*(log(Ppr))./Tpr.^2;

figure(1)
plot(z_exp,z,'*',[0.3 1.3],[0.3 1.3],'r-')
axis([0.3 1.3 0.3 1.3])
100*sum(abs(z-z_exp)./z_exp)/length(z)
figure(2)
plot(Ppr,z,'*')
```

APPENDIX E

Adaptive Neuro-Fuzzy Inference System Algorithm

The MATLAB code developed in this study for the adaptive neuro-fuzzy inference system algorithm is presented as follows.

```
clc;

clear all;

tic

%% Load Data

Data = xlsread('Data.xls');    % read the data from your excel file
```

```
x = Data(:,1:end-1);          % read the number of input columns

t = Data(:,end);              % read the number of output column

%% Generate Basic FIS

fis = genfis3(x,t);           % for different data you can use "genfis1" and "genfis2", but the best one is

"genfis3"



%% Train Using ANFIS Method

fis = anfis([x t],fis);       % training process is going to be handeled by "anfis"



%% Results

yy = evalfis(x,fis);          % evaluation or prediction is handeled by "evalfis"

ee = t - yy;                  % estimation of errors

figure;

subplot(1,2,1);

plot(t,'r:');

hold on

plot(yy,'b');

legend('Targets','Outpus');

title('Targets and Outputs');

subplot(1,2,2);
```

```
bar(ee);

legend('Errors');

toc

fuzzy(fis)
```

APPENDIX F

Definition of Statistical Error Parameters

For performance evaluation of the newly proposed artificial intelligence based models, a number of statistical deviation parameters have been used. The formulas related to the abovementioned deviation parameters are as follows:

1. Average percent relative error (APRE, or Er). It evaluates the relative deviation of estimated petroleum reservoir fluids properties data from the actual ones, which is expressed as follows:

$$E_r\% = \frac{1}{n}\sum_{i=1}^{n} E_i\% \qquad \text{(A.1)}$$

where $E_i\%$ stands for the relative deviation of a represented/predicted values from its related actual value and is defined as percent relative error:

$$E_i\% = \left| \frac{Z_{\exp} - Z_{\text{rep./pred}}}{Z_{\exp}} \right| \times 100 \Rightarrow i = 1,2,3,\ldots,n \qquad \text{(A.2)}$$

2. Average absolute percent relative error (AAPRE, or Ea), is also called average absolute relative deviation (AARD) in this study. It measures the absolute relative deviation (ARD) from the actual data and is expressed as follows:

$$E_a\% = \frac{1}{n}\sum_{i=1}^{n} |\, E_i\% \,| \tag{A.3}$$

3. Root mean square error (RMSE). It measures the data scattering around the zero deviation, expressed as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Z_{i\,\mathrm{exp}} - Z_{i\,\mathrm{rep./pred}}\right)^2} \tag{A.4}$$

4. Standard deviation (SD). It is a criterion of dispersion and a lower value exhibits a smaller degree of dispersion. It is expressed as follows:

$$SD = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{Z_{i\,\mathrm{exp}} - Z_{\mathrm{i\,rep./pred}}}{Z_{i\,\mathrm{exp}}}\right)^2} \tag{A.5}$$

5. Coefficient of Determination ($R^{2,}$ or R-squared). This parameter is a simple statistical deviation parameter which illustrations how good the model matches the data and accordantly, expresses a measure of the usefulness of the developed models. It is expressed as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(Z_{i\,\mathrm{exp}} - Z_{\mathrm{i\,rep./pred}}\right)^2}{\sum_{i=1}^{n}\left(Z_{\mathrm{i\,rep./pred}} - \overline{Z}\right)^2} \tag{A.6}$$

where $\overline{Z}$ is the mean of the actual/literature/experimental data values presented in the above equation.

APPENDIX G

The Results for the Classification of Asphaltene Stability

As mentioned earlier, to classify the asphaltene stability, the DT and ANFIS approaches have been employed in this study. Furthermore, the results obtained by the mentioned approaches are compared with the output of LSSVM algorithm (Chamkalani, 2015). **Table A. 1** summarizes a comparison between the predicted data by the ANFIS approach comparing with the LSSVM output. **Fig. A. 1** indicates the applicability ranges of most important input variables viz. aromatic + resin and asphaltene + saturates for DT algorithm. Furthermore, **Fig. A. 2** illustrates the fitting curve and relative error distribution plot for ANFIS algorithm. The results demonstrates that the output data are in satisfactory agreement with the actual field data.

Table A. 1 Comparison between the simulated output and field observation stability state.

Ar+Re	As+Sa	Data (Numeric)	Data (Symbolic)	ANFIS	LSSVM (Chamkalani, 2015)
56.16	35.09	0	No or minor	OK	OK
54.27	33.47	0	No or minor	OK	OK
63.89	28.61	0	No or minor	OK	OK
70.79	33.1	0	No or minor	OK	OK
75.97	45.69	0	No or minor	Wrong	OK
76.13	47.31	0	No or minor	OK	OK
74	45.82	0	No or minor	OK	OK
38.62	30.12	1	Mild	OK	OK
38.79	31.3	1	Mild	OK	OK
40.43	30.5	1	Mild	OK	OK
42.33	34.37	1	Mild	OK	OK
48.84	38.87	1	Mild	OK	OK
52.14	41.5	1	Mild	OK	OK
55.1	43.37	1	Mild	OK	OK
55.35	46	1	Mild	OK	OK
69.61	58.25	1	Mild	OK	OK
62.03	64.4	2	Severe	Wrong	Wrong
59.39	70.88	2	Severe	OK	Wrong
53.21	59.91	2	Severe	OK	OK
51.23	54.67	2	Severe	OK	OK
48.1	50.93	2	Severe	OK	OK
42.08	56.42	2	Severe	OK	OK
44.31	59.66	2	Severe	OK	OK

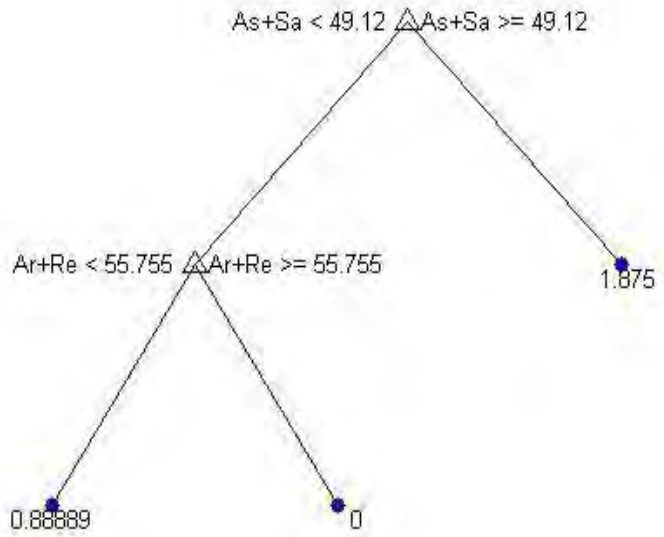

Fig. A. 1 the applicability ranges of most important input variables viz. aromatic + resin and asphaltene + saturates for DT algorithm

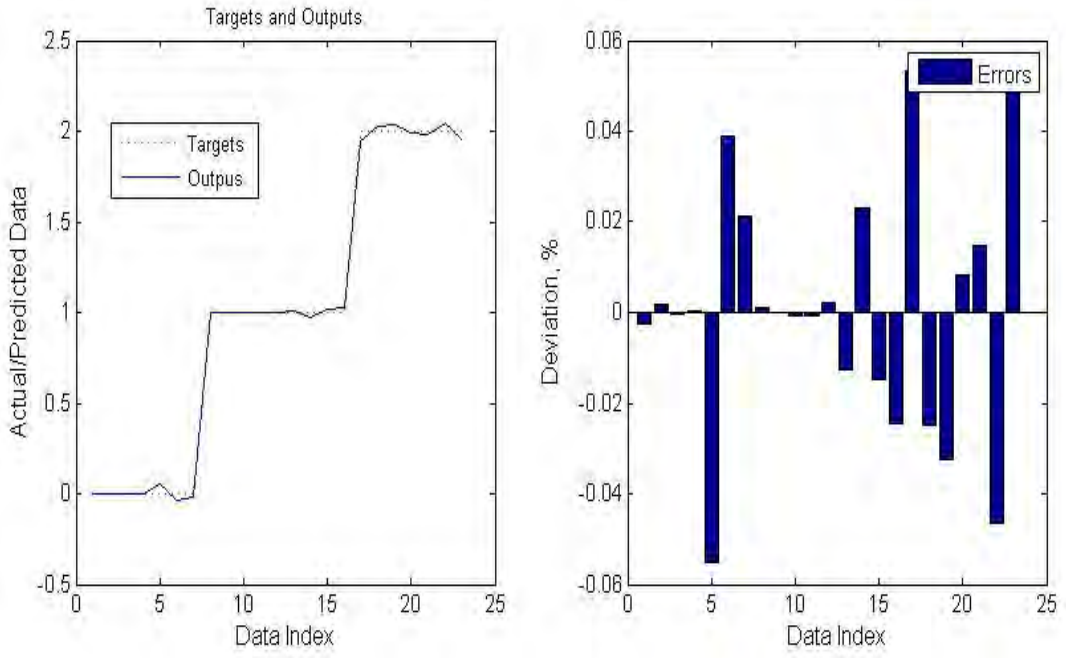

Fig. A. 2 the fitting curve and relative error distribution plot for ANFIS algorithm.

APPENDIX H

Further Study on the Classifiably Performance of the Developed Algorithms

Hydrofluorocarbons (HFCs), or "super greenhouse gases," are gases used for refrigeration and air conditioning, and known as super greenhouse gases because the combined effect of their soaring use and high global warming potential could undercut the benefits expected from the reduction of other greenhouse gases such as carbon dioxide. Used as refrigerants, they were introduced by the chemical industry to replace ozone destroying CFCs (chlorofluorocarbons). As a matter of fact, the determination of solubility of hydrocarbons in hydrofluorocarbons experimentally is expensive and time-consuming. Therefore, it is much important to develop reliable models which are fast and accurate than laboratory apparatus. To this end, three reliable predictive models have been proposed in this study for the prediction of solubility of hydrocarbons in hydrofluorocarbons.

In this study, three reliable models are developed which are based on artificial intelligence techniques viz. least square support vector machine, decision tree, and fuzzy logic system. To this end, numerical descriptors have been considered as the input variables of the models to predict the hydrofluorocarbons ability to dissolve the hydrocarbons, considering “Solvent” or soluble and “Non-solvent” or insoluble components as an output set. In other words, numerical descriptors (Dowman and Woolf, 1995) for a number of FC and HFC compounds have been considered as input variables, and the outputs are set at 1 and 0 for solvents or soluble components and non-solvents or insoluble components, respectively. The input variables are the number of C, H, and F as well as the ration C to F and H to F. Additionally, sum of α and β factors which allowed for polarity is another input in addition to R which is the ratio of number of carbons. **Fig. A. 3** is the tree diagram obtained by the DT algorithm illustrating the applicability ranges of most important input variables (R and C/F) for the prediction of solubility of hydrocarbons in hydrofluorocarbons. **Fig. A. 4** demonstrate the performance (receiver operating characteristic (ROC curve)) of the LSSVM algorithm in predicting the solubility

of hydrocarbons in hydrofluorocarbons. **Fig. A. 5** indicates the error range obtained by the ANFIS method for the output variables viz. solvents and non-solvents.

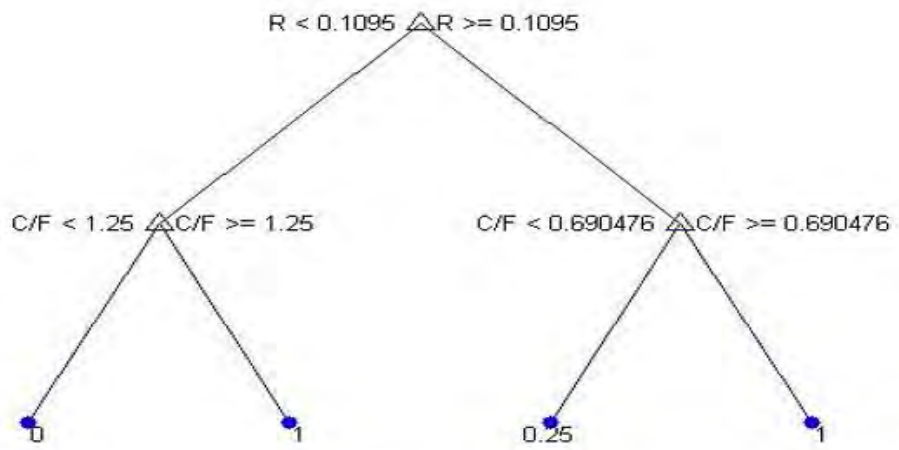

Fig. A. 3 The applicability ranges of most important input variables (R and C/F) for DT algorithm.

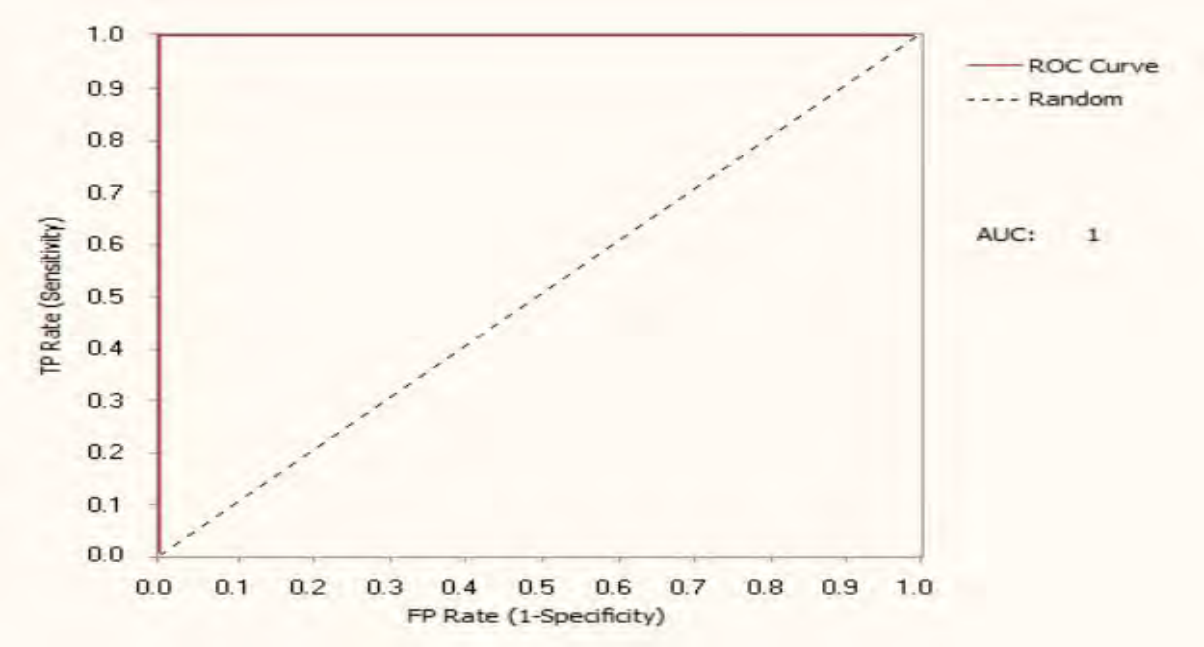

Fig. A. 4 The ROC curve which measures the performance of a classifier (LSSVM).

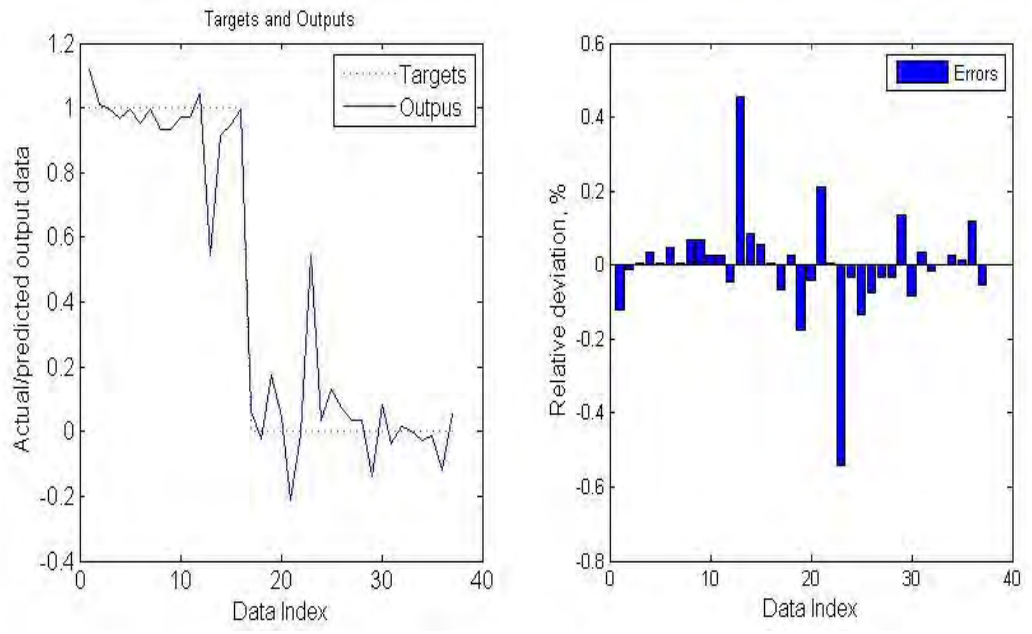

Fig. A. 5 Fitting curve and relative error distribution plot for ANFIS algorithm.