Universitetet i Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation: Master of Science in Petroleum Engineering/ Reservoir Engineering	Spring semester, 2017 Open
Author: Jørgen Bergsagel Møller	…… (signature of author)
Programme coordinator: Runar Bøe Supervisor(s): Knut Kristian Meisingset, Runar Bøe	
Title of master's thesis: An Improved Correlation Approach to Predict Viscosity of Crude Oil Systems on the NCS	
Credits: 30	
Keywords: Viscosity Empirical Correlations Particle Swarm Optimization Radial Basis Function Network	Number of pages: 61 + supplemental material/other: 12 Stavanger 15/06/2017 date/year

An Improved Correlation Approach to Predict Viscosity of Crude Oil Systems on the NCS

By
Jørgen Bergsagel Møller

Faculty of Science and Technology
University of Stavanger

A thesis submitted for the degree of
Master of Science

June 2017

Abstract

An accurate estimation of viscosity values is imperative for an optimal production and transport design of hydrocarbon fluids. Based on this requirement, cost efficient and reliable empirical correlation models are highly profitable. While there are numerous correlation models from literature, a consistent correlation model is still needed, as most models are inadequate to predict an accurate oil viscosity using unbiased data. This study aims to develop new and improved empirical viscosity correlations through available field measurements on the NCS. The performance of the proposed models is then studied through a comparative analysis with published correlations from literature.

Three new correlations are developed for dead, gas saturated and undersaturated oils. The models are based on available field data from the NCS, where all PVT reports systematically were quality assured and controlled. A high-quality database is the single most important success criterion to develop an accurate and recognized prediction model. Two different correlation models are developed in this study using Particle Swarm Optimization (PSO) and Radial Basis Function Network (RBFN). The first technique is a computational optimization algorithm that improves a function with respect to a specified objective function, while the latter is an artificial neural network model that utilizes different radial basis functions as activation functions.

The optimization algorithm was used to re-calculate the coefficients of established viscosity correlations, while maintaining the functional pattern. The results show that the modified correlations are more in agreement with the test data for all three oil types, compared to the established correlations and RBFN, using the defined parameters from literature. The new correlations provide a mean absolute percentage error (MAPE) of 15.08%, 17.41% and 3.35%, for dead, saturated and undersaturated oil viscosity, respectively. However, the Kriging method proves higher accuracy than the modified saturated correlation, when including the saturated density as input parameter.

The RBFN demonstrates a secondary estimation performance compared to the modified correlations; nevertheless, the algorithm is considered to present a satisfactory low percentage error in dead and undersaturated oils, where the empirical response is superior to all discussed correlations from literature.

The results of this study make it reasonable to conclude that the proposed correlation methods are more in-line with the measured viscosity on the NCS, compared to the established correlation models that were analyzed.

Acknowledgements

I would like to express my sincere gratitude to my supervisor at Statoil ASA, Knut Kristian Meisingset, for giving me the opportunity to write my Master's Thesis for Statoil ASA, his support and interest in my work have been instrumental for the quality of this thesis.

Special thanks go to Ibnu Hafidz Arief for his guidance to develop the machine learning algorithms. I am exceedingly grateful for his enthusiasm throughout this semester, the feedback and discussions have been greatly appreciated.

Lastly, my appreciation goes to my faculty supervisor, Professor Runar Bøe, for his valuable input, and for letting me pursue my thesis in co-operation with Statoil ASA.

Objectives

The objectives of this study are presented in the following bullet-points:

- Quality assure the provided PVT-database with the aim to develop new and improved empirical correlation models using two different correlation techniques.

- Develop and adapt the Particle Swarm Optimization and Radial Basis Function Network algorithms to create the optimum correlation performance, based on the provided field measurements.

- Conduct a comparative analysis between the proposed and the established correlation models, with respect to different statistical estimation parameters.

Table of Contents

List of Figures

List of Tables

1 Introduction

The fluid viscosity presents a considerable impact on fluid flow in porous medium, which makes it imperative in most calculations related to the extraction and transportation of hydrocarbons, such as reserves estimation, enhanced oil recovery processes, production strategy, etc. A precise oil viscosity value is therefore crucial to obtain a successful production design. Viscosity data is usually quantified by laboratory experiments, but these are usually time consuming, expensive and conducted at specific conditions.

Accurate numerical approximations are therefore essential to quantify viscosity when laboratory experiments are unreliable; PVT-analyses are not accessible, because of cost or time limitations; or when its required to quantify viscosity at different temperatures and pressures, e.g. calculate pressure development in a gas lift design. Numerous correlations have been published since the 1950s, where most correlations are developed based on samples from a specific area; consequently, highly inaccurate results are often observed when the correlations are applied outside the reference region.

The viscosity of crude oil is highly affected by the composition, therefore, distinct types of correlations have been developed for dead oil viscosity, when there is no gas in solution; gas saturated viscosity; and undersaturated oil viscosity.

The types of oil are further divided into compositional and empirical viscosity correlations, based on the required input. The compositional viscosity correlation is recognized as the most accurate prediction tool. The model is based on material balance of compositional information, which implies that a comprehensive PVT-report is required. Such PVT-reports, usually includes oil viscosity, which makes the correlation model redundant, in many cases. Often, the only information available related to the fluid property is solution gas-oil ratio, temperature, oil API gravity and pressure.

With this in mind, accurate empirical viscosity correlation models based on the available field-measurements are highly requested. This motivates a study into the relation between field parameters and viscosity, with the aim to improve the viscosity correlation accuracy on the NCS.

2 Underlying Theory

A fundamental understanding of the viscosity parameter is required to fully comprehend the characteristics of a fluid flowing through a porous medium and the mobilization of hydrocarbons. Essential theory concerning viscosity correlations are therefore outlined in the following chapter:

2.1 Fluid Viscosity

Viscosity is defined by Finnemore et al.[1] as the internal resistance for a fluid to shear. An external shear stress applied to a fluid generates a movement in the molecules, in the given shear direction. The mobilized molecules will further induce a movement in the neighboring particles; consequently, the frictional interaction generates a force that oppose the fluid flow.

The absolute (dynamic) viscosity is defined by the following equation for Newtonian fluids[2]:

$$\eta = -\frac{\tau_{xy}}{\frac{\partial v_x}{\partial y}} \tag{1}$$

Where,

$$\tau_{xy} = Shear\ stress\ per\ unit\ area, \text{ Pa}$$
$$v_x = Velocity\ of\ the\ fluid\ in\ the\ applied\ stress\ direction, \text{ m/s}$$
$$\frac{\partial v_x}{\partial y} = The\ gradient\ of\ v_x\ perpinducular\ to\ the\ stress\ direction, \text{ m/s}$$

Absolute viscosity is generally defined in oil field units as centipoise, which is equivalent to 1 mPa· s.

Kinematic viscosity is defined by the ratio of the absolute viscosity and the fluid mass density, the relation is expressed in equation 2. The parameter is usually given in centistokes (cSt); however, the viscosity may also be reported in SI as mm^2/s, which is numerically equivalent to centistokes[2].

$$v = \frac{\mu}{\rho} \tag{2}$$

Where,

$$v = Kinematic\ viscosity, cSt$$
$$\mu = Absolute\ viscosity, cP$$
$$\rho = Fluid\ mass\ density, g/cm^3$$

Crude oil viscosities usually range from 0.1cP to >100cP, representing near critical-oils to heavy crudes, where the near-critical oils represent light volatile fluids. The

viscosity is one of the most demanding parameters to estimate as it is governed by a set of key properties, such as temperature, gas in solution, STO density, pressure and composition. Where the oil viscosity is characteristically decreasing with an increase in oil API gravity, temperature and gas dissolved[3].

2.1.1 Kinetic Gas Theory

The kinetic gas theory is the study of how the interactions of molecules on a microscopic level affect the transport of molecules on a macroscopic level. By introducing a shearing stress to the already existing random velocity vector of the gas molecules, an additional bulk motion is generated, because of molecular collisions. The velocity vector is greatest near the shearing source, where the velocity monotonically decreases with increasing distance from the source, which makes the surrounding molecules move in the same direction. This microscopic behavior constitutes the fundamental theory of gas viscosity[4]. Equations 3 through 6 is an approximate expression of a diluted gas for the absolute viscosity, μ, used to express the viscosity reducing paramter[5].

$$\mu = \frac{1}{3} n\, v\, MWL \tag{3}$$

Where,

$$n = Number\ of\ molecules\ per\ unit\ volume$$
$$v = Average\ molecular\ speed, m/s$$
$$MW = Molecular\ weight, kg/mol$$
$$L = Mean\ free\ path\ between\ two\ molecules\ (m)$$

The variables that comprise the analytical expression in equation 3 is demanding to calculate. Nevertheless, the equation is applied in several gas viscosity correlations by describing the viscosity reducing parameter, ξ.

The parameter is expressed by taking the average molecular velocity in terms of temperature and molecular weight. It is further possible to show that the velocity relates to $\left(\frac{RT}{MW}\right)^{\frac{1}{2}}$, when the mean free path is proportional to $1/(n\sigma^2)$. Here, σ is the molecular hard sphere diameter. The viscosity may then be expressed as:

$$\mu = const. T^{\frac{1}{2}} M^{\frac{1}{2}} / \sigma^2 \tag{4}$$

σ^3 is often known as the critical molar volume, and it is further assumed that the volume is proportional to RT_c/P_c . Thus, viscosity at the critical point may be obtained by the following expression:

$$\mu_c = const. M^{1/2} P_c^{2/3} T_c^{1/6} / \sigma^2 \tag{5}$$

Even though the kinetic gas theory is not applicable in the near critical region, it still plays a significant part in the viscosity correlation calculations. The viscosity reducing parameter, ξ, may ultimately be expressed as:

$$\xi = M^{1/2} P_c^{2/3} T_c^{1/6} \tag{6}$$

2.1.2 Corresponding-State Principle

The corresponding-state principle is the most reliable and universal molecular theory to explain the properties of a substance. The principle considers the universal behavior of properties that are governed by intermolecular forces, which relates to the critical properties in the same way[6].

The theory is valid for all pure substances, whose PVT-properties can be explained by a two-parameter equation, the theory is exemplified in Figure 1. The left side of the figure displays a graphical illustration of the macroscopic co-existents curve of three different substances, compared to absolute temperature and absolute density, while the right side presents the same three substances; however, plotted in terms of reduced temperature, $T_r = T/T_c$, vs. reduced density, $\rho_r = \rho/\rho_c$. The response between the two plots is evident, the left figure presents three distinct curves, while the response between the three substances compared to the reduced properties corresponds almost perfectly, as the behavior is collapsed down to demonstrate essentially one curve[6]. The corresponding-state principle constitutes the basis of the compositional correlation models discussed in the subsequent section.

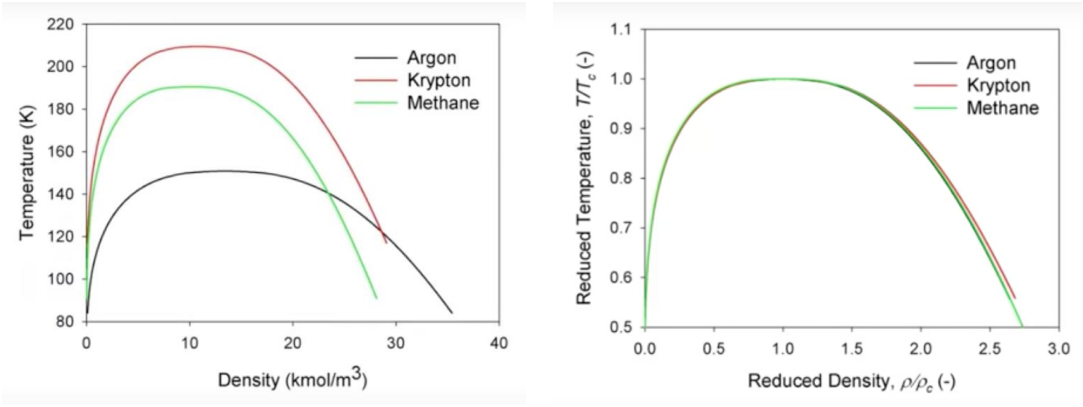

Figure 1 – Left side presents the macroscopic behavior of Argon, Krypton and Methane in terms of absolute temperature and absolute density, while the right side show the same three substances in terms of reduced temperature and reduced density[7].

However, the principle has its limitations, as unrelated molecules fail to present the analogous behavior, as observed in Figure 2. Thus, the success criterion of using the corresponding-state principle is governed by the similarity between substances.

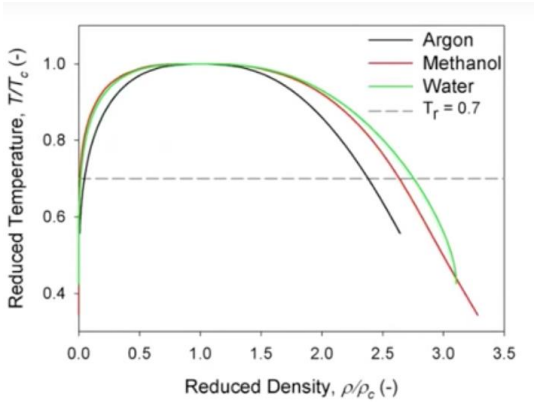

Figure 2 - Reduced temperature vs. reduced density of Argon, Methanol and Water to demonstrate the behavior of three non-similar molecules[7].

3 Viscosity Measurements and Correlations

The following section outlines different experimental measurements and correlation techniques to determine the oil viscosity:

3.1 Experimental Viscosity Measurements

The "true" viscosity is defined by experimental measurements, where the viscosity is measured in well-defined mixtures of reservoir fluids, either from bottom hole or recombined separator samples. There are three different measurement techniques related to this study: Electro-Magnetic, Rolling-ball and Gravimetric Capillary Principle.

Electro-Magnetic Viscometer

Electro-Magnetic Viscometer measures viscosity by initiating a constant force on a piston inside a chamber through electromagnetism. Viscosity is measured as the viscous forces impede the motion of the piston, while flowing in the annulus between the piston and the measurement chamber wall. The motion is dictated by the viscosity of the fluid, i.e. the more viscous fluids, the slower the piston moves[8].

Rolling-Ball Principle

Rolling ball experiment is one of the oldest and most basic techniques to measure viscosity. The experiment is carried out by calculating the terminal velocity of a sphere falling through a fluid; thus, distance and time is required, in addition to the mass and diameter of the sphere. The more viscous the fluid, the slower the sphere rolls, as the buoyancy of the fluid opposes the gravitational force. The expression to determine dynamic viscosity through the rolling-ball experiment relates to the following[9]:

$$\eta = K_p * (\rho_s - \rho) * t \tag{7}$$

Where,

$$\eta = Dynamic\ Viscosity,\ cP$$
$$K_p = \text{Proportionality Constant}$$
$$\rho_S = Sphere\ Density,\ g/cm^3$$
$$\rho = Fluid\ Density, g/cm^3$$
$$t = Time, s$$

Gravimetric Capillary Principle

Viscosity is determined with the Gravimetric Capillary principle by quantifying the time a fluid need to traverse through a capillary tube with a defined diameter and length. A minimum flow time is defined to ensure that the flow conditions inside the capillary do not contradict the assumption of laminar flow. The principle of the experimental methods relates to the kinematic viscosity as the technique utilizes time and gravity as driving force, as seen in equation (8[9]:

$$\upsilon = K_C * t_f \tag{8}$$

Where,

$$\upsilon = Kinematic\ Viscosity,\ \ cSt$$
$$K_C = Capillary\ Factor, mm/s^2$$
$$t_f = Time, s$$

3.2 Theoretical Viscosity Correlations

While the kinetic gas theory is the theoretical basis to correlate gas viscosity at low pressures, there is no equivalent theory for liquid viscosity calculations. The correlation of liquid viscosity is essentially divided into three categories: dead oil, gas saturated and undersaturated oil viscosity. However, the correlations do not allow calculations of viscosity from molecular structure[3]; therefore, alternative techniques have been developed. The following section outlines three different types of correlation models: compositional correlations, correlations using surrogate models, and empirical correlations, presented successively in the order given.

3.2.1 Compositional Viscosity Correlation Models

The following presents the two most recognized compositional correlations models, where both models are based on the corresponding-state principle.

Lohrenz, Bray and Clark[10]

The compositional correlation method of Lohrenz, Bray and Clark (LBC) is a continuation and modification of the published work by Jossi et al[11]. The model is established on the investigation of viscosity of pure substances, such as Argon, Nitrogen, Oxygen, Carbon Dioxide, Sulphur Dioxide, Methane and Ethane. The correlations are developed based on a dimensional analysis of each distinctive substance, and the assumption that the residual viscosity, $\mu - \mu_{atm}$, is a function of density.

The authors later adapted the work to apply for hydrocarbon mixtures by comparing experimental and calculated results of 260 crude oils. The work is now recognized as one of the foremost important compositional viscosity correlation used in liquid flow models. The correlation of LBC is related to a fourth-degree polynomial in the reduced density, presented as follows:

$$[(\mu - \mu^*)\xi + 10^{-4}]^{\frac{1}{4}} = a_1 + a_2\rho_r + a_3\rho_r^2 + a_4\rho_r^3 + a_5\rho_r^4 \quad (9)$$

Where,

$$a_1 = 0.1023$$
$$a_2 = 0.023364$$
$$a_3 = 0.058533$$
$$a_4 = -0.040758$$
$$a_5 = 0.0093324$$
$$\mu^* = Low\ pressure\ gas\ mixture$$

And, ξ is the viscosity-reducing parameter given by:

$$\xi = \left[\sum_{i=1}^{N} x_i T_{ci}\right]^{\frac{1}{6}} \left[\sum_{i=1}^{N} x_i M_i\right]^{-\frac{1}{2}} \left[\sum_{i=1}^{N} x_i P_{ci}\right]^{-\frac{2}{3}} \quad (10)$$

Where,

$$N = Number\ of\ components\ in\ mixture$$
$$x_i = Mole\ fraction\ of\ component\ i$$

Pedersen et al.[12]

Pedersen et al. presented a model to determine viscosity at gaseous and liquid phases, based on the corresponding-state principle, using methane as a reference substance. The method is more comprehensive than the method of LBC[10]; thus, more computing power is required to obtain satisfactory viscosity values. However, the model is advantageous as it provides more consistent viscosity values in the near-critical region of the respective fluid, compared to other correlation methods.

The concept is developed based on methane as reference fluid, as methane is one of the most reviewed substances in terms of viscosity and density in both liquid and gaseous phases. Later, in 1988, the authors continued their work to develop the following viscosity relation[12]:

$$\mu = \left(\frac{T_C}{T_{c,methane}}\right)^{-\frac{1}{6}} \left(\frac{P_c}{P_{c,methane}}\right)^{\frac{2}{3}} \left(\frac{M_w}{M_{w,methane}}\right)^{\frac{1}{2}} \frac{\propto}{\propto_{methane}} \mu_{methane}[P', T] \quad (11)$$

Where,

$$P' = \frac{P * P_{c,methane} * \propto_{methane}}{P_c * \propto}$$

$$T' = \frac{T * T_{c,methane} * \propto_{methane}}{T_c * \propto}$$

$$\propto = 1 + 7.378 * 10^{-3} * \rho_r^{1.847} * M_w^{0.5173}$$

$$\propto_{methane} = 1 + 0.031 \rho_{r_{methane}}^{1.847}$$

3.2.2 Surrogate Models

The traditional correlation models present a mathematical model based on field measurements to predict viscosity. While this section discusses the use of surrogate models as an alternative method to predict the viscosity properties. The main difference is that the surrogate models do not present a mathematical correlation, as

the models use a more statistical approach with more consideration of the variation in input variables, compared to the traditional correlations. Three models are considered in this study: Universal Kriging, Artificial Neural Network and Radial Basis Function Network.

The models are to mimic the behavior of regular simulation models while being computationally less expensive. The algorithms are defined as black box modelling, meaning that the inner part of the model is assumed to be unknown; the system is based on the transfer characteristics, i.e. only the input and output variables have any influence to the model[13]. The aim of the surrogate model is to replace time consuming and cumbersome simulations and experiments to predict the outcome variables of a known input set[14].

Universal Kriging

The method is a stochastic interpolation technique strongly related to regression analysis of surrounding data points. The method was initially applied in the petroleum industry as a geostatistical technique to determine field properties, such as porosity and rock permeability, based on the input of several wells in a particular field[15].

The method essentially weights data-points through a semi-data driven function rather than an arbitrary function, i.e. some points are more important than others. The technique considers not only distance, but also orientation and direction of each data point to determine the unknown data. The underlying idea is that the sampling points in the vicinity of the objective point will be weighted more than the farther points. The concept is expressed by the following equation[16]:

$$Z(s_0) = \sum_{i=1}^{n} w_i Z_i \tag{12}$$

Where,

$$Z(s_o) = Estimated\ value\ of\ unsampled\ region\ from\ regionalized\ variable, Z_i$$
$$w_i = Weight\ of\ corresponding\ regionalized\ value, Zi$$
$$Z_i = Regionalized\ variable$$

Artificial Neural Network

Neural network is a conceptual model inspired by the neurons in the human brain. The computational model is based on a single unit that receives input information to create an output. Neural networks provide the computers a sense of intuition to solve complex problems. Put in simple terms, the neural network adjusts some of the original set of values to develop a best-fit model, by creating a prediction of the output values based on the input pattern. While most computational systems are procedural, i.e. execute and process code linearly from the first to the last line, the Neural Network computes information collectively by a series of parallel network nodes[17].

Figure 3 illustrates a two-layer neural network. In this case, there are three cells in the input layer, four hidden cells in the intermediate layer and two cells in the output layer. The concept is modeled in an acyclic graph, as a cyclic network would involve an infinity loop. The neurons are further fully pairwise connected between two adjacent layers, but not within a layer.

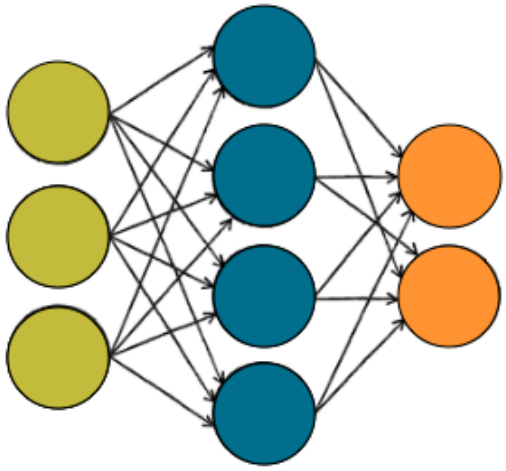

Figure 3 - A graphic illustration of a two-layer artificial neural network, where there are synapses between all neurons across the layers, but not within a layer [18].

Radial Basis Function Network

The Radial Basis Function Network (RBFN) is a continued product from the artificial neural network that uses radial basis functions as activation functions, i.e. it uses a real function which only depends on the distance from the origin, to define the output of a node from a certain set of inputs. The function carries out a comparative analysis between the input signals and a set of reference vectors, by computing the Euclidian distance, i.e. the regular straight-line distance to approximate the input function[19]. Simply put, if the input signal is more analogous to class X than class Y, the input signal will be classified as class X. This dissertation utilizes three of the most common types of radial basis functions that relates to the following expressions[20]:

Gaussian:

$$\varphi(r) = e^{-(\varepsilon r)^2} \tag{13}$$

Multiquadric:

$$\varphi(r) = \sqrt{1 + (\varepsilon r)^2} \tag{14}$$

Inverse quadratic:

$$\varphi(r) = \frac{1}{1 + (\varepsilon r)^2} \tag{15}$$

Where,

$$r = \|x - \mu\|$$

The μ is defined as the average distribution of the sample, i.e. the reference vector at the center of the bell curve, as shown in Figure 4. The double bar notation that defines r, simply denotes the Euclidean distance between μ and x.

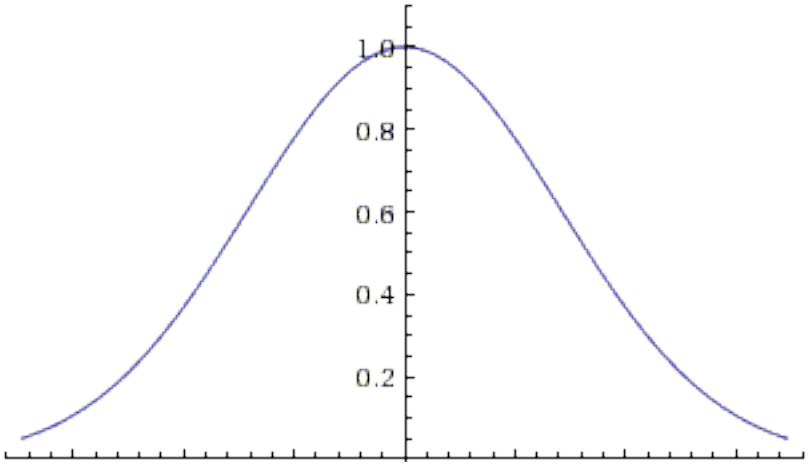

Figure 4 - The RBF activation function illustrates the resemblance between the input vector and reference vectors. An analogous match generates a function shape approaching 1.

Figure 4 show a typical shape response of RBFN, where the function is made up by an input vector, a hidden layer with RBFN-reference neurons, and an output layer comprised by a set of single classified node. The hidden layer identifies a classification to the N-dimensional input vector, where each neuron compares the input vector to a reference function, to compute the degree of correlation ranging from 0 to 1, whereas 1 represents a uniform fit[19]. The RBFN is conceptually illustrated in Figure 5.

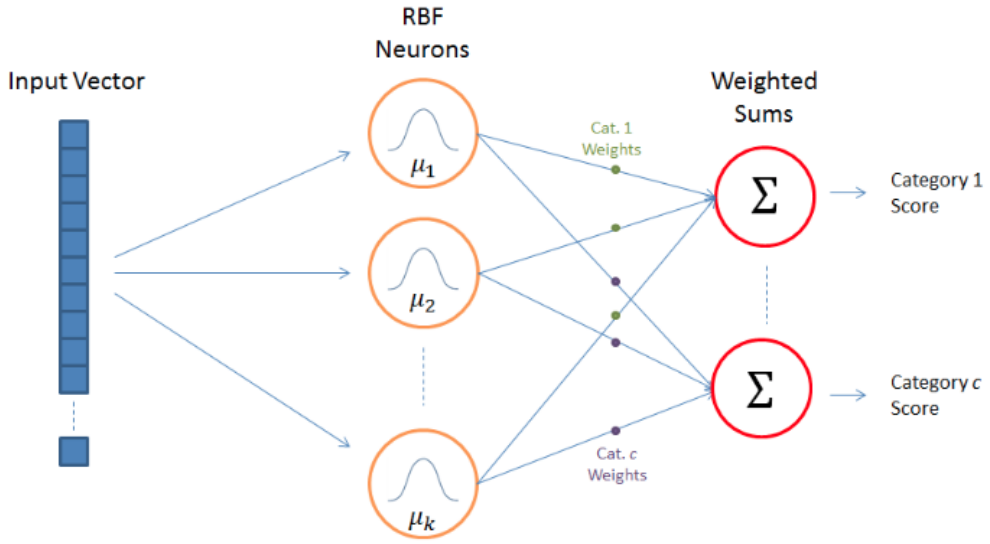

Figure 5 - Traditional RBF Network architecture with a single input vector, a hidden layer where the input signal is compared against the RBF reference function, and a categorized output layer[19].

Particle Swarm Optimization

The method is a stochastic optimization technique inspired by the flocking and schooling patterns of birds and fish. PSO is comprised of a swarm of particles moving through the problem space. Each particle is termed as unintelligent, meaning that they as an individual are unable to compute a solution; however, the particles are able to obtain a solution through interactions with neighboring particles.

The concept of PSO relates to the principles of learning and communication. The particles have a memory of their personal best (pbest) solution in the problem space, relative to the defined objective function. Each particle further has a memory of the global best (gbest) solution, which is the best position any particle has achieved at each time-step. Over a number of iterations, each particle interacts with other particles to change its position according to the objective function, based on the initial position, velocity vector, pbest and the gbest[21].

Figure 6 illustrates the traditional flow of the algorithm. The PSO is first initialized by generating a set of random numbers representing the search particles, where each particle is a potential solution to the defined objective function.

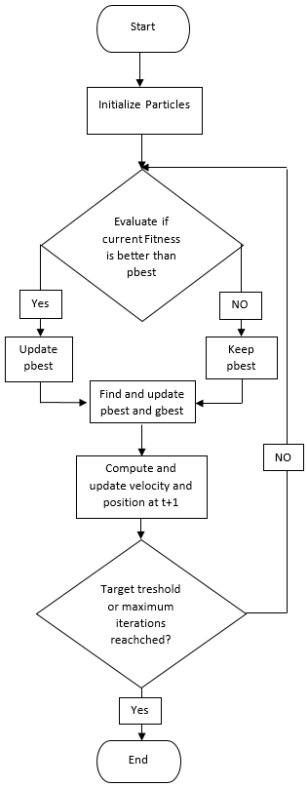

Figure 6 – Flowchart of standard PSO algorithm. First, the PSO is populated with a random set of numbers representing the particles. The random valued particles move through the search space to optimize a defined objective function by communicating their pbest and following gbest at each iteration.

The algorithm then finds the pbest and gbest, before the particles updates their respective velocity and position through the following formulas:

$$x_i(t+1) = x_i(t) + v_i(t+1) \tag{16}$$

Where,

$$x_i = Position\ of\ particle\ i$$
$$v_i = Velocity\ of\ particle\ i$$

The new velocity of the i^{th} particle at time step t+1 relates to the following expression:

$$v_i(t+1) = wv_i(t) + c_1\big(p_i(t) - x_i(t)\big) + c_2(g(t) - x_i(t)) \quad (17)$$

Where,

$$w = Intertia\ Value$$
$$c_1 = Cognitive\ acceleration\ Component$$
$$p_i = Personal\ best\ of\ particle\ i$$
$$c_2 = Social\ acceleration\ component$$
$$g(t) = Global\ best$$

The inertia term plays a crucial role in the success criteria in the PSO method, as it provides a flexibility to enhance the global and local exploration and exploitation of the particles. The cognitive acceleration coefficient, C_1, relates to the particle's ability to perceive its own personal best position and the tendency to return to this position. The social acceleration coefficient, C_2, describes the capability each particle has to communicate the global best position[22].

3.2.3 Empirical Viscosity Correlation

Empirical viscosity correlations are widely implemented in the petroleum industry as an alternative to cumbersome and expensive laboratory experiments. Most correlations are based on laboratory results or field data, as presented in the following literature review. The models are developed by treating the oil as a two-component system, i.e. considering the STO and gas collected at the surface. The correlations are often termed as "black oil correlations" as the calculations are not based on compositional input. This simplification of the system is based on the assumption that crude oils generally are comprised of approximately 40% methane and C_{7+}-components. The remaining components are believed to present only a minor impact on the system. Most empirical correlation models for viscosity are based on four input parameters: oil API gravity, solution gas-oil ratio, reservoir temperature and pressure[23].

Numerous correlations have been developed for the petroleum industry, where they differ mostly based on the range of data used in the calculations. Thus, it is important to be aware of the different methods restrictions, as they as often are limited to be valid within the reference data, or that they only present adequate results for one specific geologic region, such as the Middle East and the North Sea. The following summarizes briefly three different oil classifications related to the following literature review.

Dead Oil Viscosity

Dead oil viscosity is defined when there is no gas dissolved in the fluid at standard conditions. All dead oil viscosity correlations have stated the viscosity as a function of oil API gravity and temperature. Figure 7 demonstrates the typical shape curve between oil API gravity and dead oil viscosity; nevertheless, the dead oil viscosity is one of the more problematic parameter to correlate. The difficulty attributes in the variables dependency on asphaltic, paraffinic and aromatic components[3].

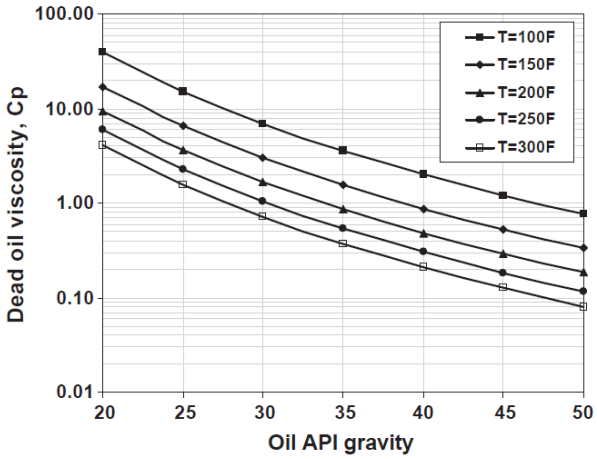

Figure 7 – Typical relation between dead oil viscosity and oil API gravity and different temperatures. A correlation approach for prediction of crude oil viscosities[24].

Saturated-oil Viscosity

The saturated oil viscosity is determined when the reservoir pressure is less than, or equal to the saturation pressure, at a specified temperature. All discussed correlation models presents live oil viscosity correlation as a function of dead oil viscosity and the solution gas oil ratio, excluding the proposed model of Labedi[25], which expresses the viscosity in terms of saturation pressure instead of GOR.

Later, Abu-Khamsin and Al-Marhoun [26] conveniently identified that the saturated oil viscosity, μ_{ob}, corresponds nicely to the saturated oil density, ρ_{ob}, which is further implemented in the compositional correlation method of LBC[10]. The viscosity further proves a reasonable prediction of the undersaturated oil density and dead oil density, the latter is; however, restricted to higher temperatures[3].

Undersaturated oil Viscosity

When the reservoir pressure increases beyond the saturation pressure the oil becomes undersaturated. Characteristically, the viscosity of a crude oil will be reduced when saturated with a gas under pressure. The viscosity is predicted as a function of bubble point viscosity, reservoir pressure and bubble point pressure, in all the presented empirical correlation models[3].

4 Literature Review

There are numerous publications in the petroleum industry trying to establish equitable empirical correlations of oil viscosity. The purpose is to design an accurate prediction of viscosity by the use of available field measurement data. The following reviews the empirical correlation models, presented in the following comparative analysis in Chapter 6. The publications are presented successively, with respect to time:

Beal[27]

In 1946, Beal developed one of the first viscosity correlations when he presented two charts to determine dead oil viscosity, as a function of oil API gravity and temperature. The first chart was established at temperatures up to 100°F, from 655 data points gathered from 492 fields in the United States, while 90 samples were used to develop a model at temperatures exceeding 100°F. The higher temperature correlation gave an average relative error of 29.0%, while the other model provided an average relative error of 25.6%. The author presented, moreover, a correlation to predict viscosity of undersaturated oil, based on 52 data points from 26 different crude oils. The correlations presented an overall average relative error of 24.2%. The study is, however, considered to have its limitations as no analytical expressions corresponding to the graphical correlations were published.

Chew & Connally[28]

In 1959, Chew & Connally presented a correlation expression together with a graphical interpretation, to predict the saturated oil viscosity. The method was developed using 457 oil samples from the most prominent producing fields of Canada, USA and South America. Their work revealed that the relationship between dead oil and saturated oil viscosities, at a constant dissolved GOR, provides a straight line on a logarithmic scale. The published correlation equation for saturated oil viscosity is as follows:

$$\mu_{ob} = A\mu_{od}^{B} \tag{18}$$

Where the coefficients, A and B, represent functions of solution gas-oil ratio.

Beggs & Robinson[29]

Beggs & Robinson developed two correlations to predict viscosity from dead and saturated crude oils as a function of temperature and API gravity. The dead oil viscosity model is based on 460 data points obtained from 93 oil samples, while the saturated oil viscosity correlation was developed from 2073 samples. The authors modified the concept of Chew & Connally[28] stating that a Cartesian plot of log(T) versus $log[\log(\mu_{od} + 1)]$ provides a linear relation, where each line represents different oil API gravities. However, the method demonstrates a considerable error when tested against samples, other than the reference samples. The authors never clarified the source of error, but advised that an extrapolation of the method outside

the range of data used in the publication should be with care. The correlation equations are presented below.

Dead oil viscosity

$$\mu_{od} = -1 + 10^{C} \tag{19}$$

Where,

$$C = 10^{3.0324-0.02023\gamma_{API}} * T^{-1.163}$$

With T in °F, μ in cp and $\gamma_{API}\ in\ °API$, for equation 19 through 43.

Saturated oil viscosity

$$\mu_{ob} = A\mu_{od}^{B} \tag{20}$$

Where,

$$A = 10.715(R_S + 100)^{-0.515}$$
$$B = 5.440(R_S + 150)^{-0.338}$$

An average error of -0.64% was observed for the dead oil correlation compared to measured viscosity data, while the latter correlation presented an average error of -1.83%.

Standing [30]

Standing published in 1977 new correlations based on Beals[27] graphical approach to predict dead oil and undersaturated oil viscosity. The new correlation demonstrated an average statistical error of -1.58%, compared to the error of 24% from the original method. The author presented, moreover, a modified prediction model based on the work of Chew & Connally[28] to predict the saturated oil viscosity. The following correlation equations presents the findings for dead oil, saturated oil and undersaturated oil:

Dead Oil:

$$\mu_{od} = \left(0.32 + \frac{1.8 * 10^{7}}{\gamma_{API}^{4.53}}\right)\left(\frac{360}{T + 200}\right)^{D} \tag{21}$$

Where,

$$D = antilog(0.43 + \frac{8.33}{\gamma_{API}})$$

Saturated Oil:

$$\mu_{ob} = A\mu_{od}^{B} \tag{22}$$

Where,

$$A = 10^{\left(2.2*10^{-7}R_s - 7.4*10^{-4}\right)R_s}$$

$$B = \left(\frac{0.68}{10^{8.62*10^{-5}*R_s}}\right) + \left(\frac{0.25}{10^{1.10*10^{-3}*R_s}}\right) + \left(\frac{0.062}{10^{3.74*10^{-3}*R_s}}\right)$$

Undersaturated Oil:

$$\mu_o = \mu_{ob} + 0.001(p - p_b)(0.024 * \mu_{ob}^{1.6} + 0.038 * \mu_{ob}^{0.56}) \quad (23)$$

Glaso[31]

Glaso presented in 1980 a generalized mathematical expression to predict dead oil viscosity by analyzing 29 data points from six crude oil samples. The correlation was developed on temperatures ranging from 50 to 300 °F, with an oil API gravity varying from 20 to 48°. The correlation is developed on North Sea crudes; thus, predicting viscosity should mainly be conducted on samples from this region. The author claimed, however, that the model could be extended to be valid for all compositions by a paraffinicity correction, i.e. correct for varying amounts of paraffinic oil components. The following equation presents the proposed relation to predict dead oil viscosity:

$$\mu_{oD} = (3.141 \times 10^{10})T^{-3.444}log(\gamma_{API})^{[10.313(logT)-36.447]} \quad (24)$$

Al-khafaji et al.[32]

Al-Khafaji et al. modified the viscosity correlations of Beal[27] and Chew & Connally[28] to develop a prediction model for dead oil, saturated and undersaturated oil. The empirical correlation expressions were established from 1270 data points of different crude oil compositions from the Middle East. The correlations were developed as a function of temperature, oil gravity and GOR, ranging from 60-300 °F , 15-51 °API and $0 - 2100$ scf/STB. The model presented an absolute average percentage error of 4.8% when tested against the data of the dead oil, while the saturated and unsaturated crudes reported an absolute average percentage error of 2.7% and 0.44%, précised in the order given.

Al-khafaji et al. gave the following correlation expressions:

Dead oil:

$$\mu_{od} = \frac{10^{4.9563 - 0.00488T}}{(\gamma_{API} + T/30\ - 14.29)^{2.709}} \quad (25)$$

The modified correlation for viscosity at the saturation pressure extends to a GOR of 2000 scf/stb, while the original expression was primarily developed on samples less than 1000 scf/stb.

Saturated Oil:

$$\mu_{ob} = A\mu_{od}^{B} \tag{26}$$

$$A = 0.247 + 0.2824A_o + 0.5657A_o^2 - 0.4065A_o^3 + 0.0631A_o^4$$
$$B = 0.894 + 0.0546A_o + 0.07667A_o^2 - 0.0736A_o^3 + 0.01008A_o^4$$

Where,

$$A_o = \log(R_s)$$

Egbogah & Ng[33]

In 1983, Egbogah & Ng presented a method to predict dead oil viscosity. The model is based on the empirical correlation of Beggs & Robinson[29]. The modified correlation was developed on 394 oil systems, and revealed an average percentage error of 5.85%, between the measured and calculated values. The authors later proposed a second correlation including pour point, i.e. the lowest temperature point of which a fluid can maintain its flow characteristics, to investigate the impact of oil composition on the viscosity. Pour point is, however, cumbersome to measure and is rarely included in the typical PVT-report; thus, the latter correlation is considered unprofitable, bearing in mind the aim of this study.

Dead Oil:

$$\log(\log(\mu_{oD} + 1) = 1.8653 - 0.025086\gamma_o - 0.5644\log(T) \tag{27}$$

Dead Oil including pour point:

$$\log((\log(\mu_{oD} + 1)) = -1.7095 - 0.0087917T_p + 2.7523\gamma_o - \left(1.2943 - 0.0033214T_p - 0.9581957\gamma_o * \log\left(\frac{T + 32}{1.8} + T_p\right)\right) \tag{28}$$

Where,

$$T_p = Pour\ point\ temperature, ℃$$

Note that the correlation is valid for pour point temperatures ranging from -50 to 15 ℃.

Labedi[25]

In 1992, Labedi introduced a set of new correlations for dead oil viscosity, saturated and undersaturated oil viscosity, through the use of multiple regression analysis of 100 oil samples from Libya. The correlation of dead oil viscosity revealed an average error of -2.61%, while the saturated oil viscosity presented an average error of

-2.38%. Even though the model is developed based on samples from a relatively small region, the author claims the correlation to be valid for other areas, e.g. Middle East and North Sea. However, the correlation should not be conducted on oil samples having an oil API gravity less than 32°. The empirical correlation expressions are presented as follows:

Dead Oil:

$$\mu_{od} = \frac{10^{9.224}}{\gamma_{API}^{4.7013} * T^{0.6739}} \tag{29}$$

Saturated Oil:

$$\mu_{ob} = (10^{2.344-0.03542*\gamma_o}) * \frac{\mu_{od}^{0.6447}}{p_b^{0.426}} \tag{30}$$

Undersaturated Oil:

$$\mu_o = \mu_{ob} - \left[(1 - \frac{p}{p_b})(\frac{10^{-2.488} * \mu_{od}^{0.9036} * p_b^{0.6151}}{10^{0.01976*\gamma_{API}}}\right] \tag{31}$$

Bergman[3]

Bergman developed two correlations to predict dead oil and saturated oil viscosities based on new data, and data taken from the publication of Beggs & Robinson [29]. The author used the concept of Beggs & Robinson to develop a new correlation for dead oil viscosity, while the published work of Chew & Connally [28] constituted the basis of the saturated viscosity correlation. The Bergman correlation equations are presented as follows:

Dead oil:

$$\mu_{od} = -1 + e^{e^{22.33-0.194*\gamma_{API}+0.00033*\gamma_{API}^2-(3.20-0.0185*\gamma_{API}^2)\ln(T+310)}} \tag{32}$$

Saturated oil:

$$\mu_{ob} = A\mu_{od}^B \tag{33}$$

Where,

$$A = \exp(4.768 - 0.8359\ln(R_s + 300))$$

$$B = 0.555 + \frac{133.5}{R_s + 300}$$

Kartoatmodjo & Schmidt[34]

In 1994, Kartoatmodjo & Schmidt developed a new empirical correlation to predict viscosity of dead, saturated and undersaturated oil samples. Their model is one of the most comprehensive viscosity correlation studies using a large databank from reservoirs worldwide. The models are a continuation and revision of previous work,

where the dead oil viscosity model is a modified version of the Glaso[31] correlation, while the saturated oil and undersaturated oil viscosities was developed by modifying the correlation concept of Chew & Connally[28] and Beal[27], respectively.

The dead oil correlation method was developed using 661 data points from 26 heavy dead oil samples, where an average absolute percentage relative error of 39.60% was reported. The saturated viscosity correlation was developed using 27 samples, which provided an average percentage relative error ranging from -12.82 to 14.34%. The undersaturated viscosity calculations was developed using non-linear regression of 3588 data points obtained from 321 undersaturated oil samples, the new correlation revealed an average absolute percentage relative error of 2.64%. The respective correlation calculations are presented as follows:

Dead oil:

$$\mu_{od} = 16.0 * 10^8 * T^{-2.8177} * log(\gamma_{API})^{5.7526log(T)-26.9718} \quad (34)$$

Saturated oil:

$$\mu_{ob} = -0.06821 + 0.9824 * H + 0.0004034 * H^2 \quad (35)$$

Where,

$$H = (0.2001 + 0.8428 * 10^{-0.000845*R_s})\mu_{od}{}^{(0.43+0.5165*I)}$$

$$I = 10^{-0.00081*R_s}$$

Undersaturated oil:

$$\mu_o = 1.0081 * \mu + 0.001127(p - p_b)(-0.006517 * \mu^{1.8148} + 0.038 * \mu^{1.590}) \quad (36)$$

Later, the authors compared the model to an unbiased databank to investigate the validity and statistical accuracy of the new modified models. The study showed that the proposed model provided the best empirical response for all three oil types, relative to the compared established correlations.

Petrosky & Farshad[35]

In 1995, Petrosky & Farshad developed three empirical correlations to estimate the viscosity of dead, saturated and undersaturated oils from the Gulf of Mexico. The presented correlations were developed using non-linear multiple regression analysis. The publication presented an average absolute error of 14.47% and 2.91% for saturated oil and undersaturated oil, respectively. The correlation proved to be a significant improvement in the respective reference area. The authors claimed, moreover, that the published correlation is applicable in other regions, as long as the

correlations are exercised within the reference parameter. The respective correlations for oil viscosities are according to the following relations:

Dead Oil:

$$\mu_{od} = 2.3511 * 10^{7} * T^{-2.10255} \log(API)^{x} \quad (37)$$

Where,

$$x = 4.59388 \log(T) - 22.827022$$

Saturated Oil:

$$\mu_{ob} = A\mu_{od}^{B} \quad (38)$$

Where,

$$A = 0.1651 + 0.6165 * 10^{-6.0866*10^{-4}*R_s}$$
$$B = 0.5131 + 0.5109 * 10^{-1.1831*10^{-3}*R_s}$$

Undersaturated Oil:

$$\mu_o = \mu_{ob} + 1.3449 * 10^{-3}(p - p_b) * 10^{k} \quad (39)$$

Where,

$$K = -1.0146 + 1.3322 \log(\mu_{ob}) - 0.4876 \log(\mu_{ob})^2 - 1.15036 \log(\mu_{ob})^3$$

Bennison[36]

Bennison presented a new correlation based on only 16 heavy crude oil samples from the North Sea. The correlation revealed an 13% mean error compared to the measured data. The author recommends the model to be used on heavy crude oils, at API gravities <20° and at temperatures <250°F. However, the model should be used with care as it is established on a very limited amount of data. The correlation expression is presented below:

Dead Oil:

$$\mu_{od} = 10^{(0.10231*\gamma_{API}^2 - 3.9464*\gamma_{API} + 46.5037)} * T^{(-0.04542*\gamma_{API}^2 + 1.70405*\gamma_{API} - 19.18)} \quad (40)$$

Elsharkawy & Alikhan[37]

In 1999, Elsharkawy & Alikhan presented a study to predict dead, saturated and undersaturated oil viscosity. The correlation model was developed based on 254 crude oil reference samples from the Middle East, using multiple regression analyses. The authors conducted, furthermore, a comparative analysis between the proposed models and the models provided by Chew & Connaly[28], Beggs & Robsinon[29], Labedi[25] and Kartoatmodjo & Schmidt[34]. Discretion is advised when using the correlation models outside the reference region, as no information is provided by the authors regarding the validity of the correlation outside the Middle East. The proposed correlation equations are provided as follows:

Dead oil:

$$\log(\log(\mu_{od} + 1) = 2.16924 - 0.02525 * API - 0.68875Log(T) \quad (41)$$

Saturated Oil:

$$\mu_{ob} = A * \mu_{od}^{B} \quad (42)$$

Where,

$$A = 1241.932(R_S + 641.026)^{-1.12410}$$
$$B = 1768.841(R_S + 1180.335)^{-1.06622}$$

Undersaturated Oil

$$\mu_o = \mu_{ob} + 10^{-2.0771}(p - p_b)(\mu_{od}^{1.19279} * \mu_{ob}^{-0.40712} * p_b^{-0.7941}) \quad (43)$$

Arief et al.[14]

In 2017, Arief et al. proposed a new method to predict the saturated oil viscosity using two different surrogate models: Universal Kriging and Neural Network. The aim of the study was to replace the traditional correlation methods. The study showed that the Universal Kriging was superior to the established correlations using 100 testing points, where the surrogate model demonstrated a mean absolute error of 20.7%. Nevertheless, the use of surrogates demonstrates an evidently disadvantage as the method require a large PVT database in the machine learning process, and that the models do not generate a mathematical correlation; thus, the models are only available for internal use.

5 Method

The following is an outline of the methodology related to the study. The presented machine learning algorithms are developed using the Python programming language[38].

5.1 Prediction Criteria

Statistical and graphical error analyses are the most common method to evaluate the prediction accuracy of oil viscosity correlation expressions. There are three statistical parameters used in this thesis as prediction criteria: mean absolute error (MAE), mean absolute percentage error (MAPE), R squared (R^2) and mean percentage error (MPE). The criteria are related to the following expressions:

$$MAE = \frac{1}{n}\sum_{i=1}^{N}\left|\mu_{estimated,i} - \mu_{measured,i}\right| \tag{44}$$

$$\text{MAPE}=\frac{100}{N}\sum_{i=1}^{N}\left|\frac{\mu_{estimated,i}-\mu_{measured,i}}{\mu_{measured,i}}\right| \tag{45}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{100}\left(\mu_{estimated,i} - \mu_{measured,i}\right)^2}{\sum_{i=1}^{100}(\bar{\mu}_{measured,i} - \mu_{measured,i})^2} \tag{46}$$

$$MPE = \frac{100}{N}\sum_{i=1}^{N}\frac{\mu_{estimated,i} - \mu_{measured,i}}{\mu_{measured,i}} \tag{47}$$

Where $\mu_{estimated}$ and $\mu_{measured}$ are the calculated and the measured viscosity, while N is the number of samples of which the correlations are tested against.

MAE provides an indication of how well the calculated values fit to the real values. The parameter provides actual rather than percentage values as output, i.e. it measures the accuracy of a forecast value, with respect to the unit of measure. The severity of the error is, however, not always obvious, as the parameter does not provide any information about the relative degree of deviation between the calculated values and the real values.

MAPE is widely used in tracking the relative forecast accuracy in percentage terms. MAPE presents, however, some limitations, as the true value is located in the denominator; thus, parameters may take on extreme values when the true value approach zero. The MAPE can be computed in different ways, depending on the denominator. There are two relevant methods for this thesis, either using the actual value, or the forecast value as the denominator. This study is carried out using the actual value, as seen in equation (45. Using the forecast value in the denominator makes it possible to measure performance against forecast; however, the actual value

is preferred, as the correlation performance is more easily compared. Using the forecast value makes it difficult to evaluate the statistical accuracy between predictions, as the value is not fixed. The measured value is further the target of the study, and is therefore chosen to serve as the baseline for the measurements.

The coefficient of determination, denoted by R^2, describes how well one dependent variable is able to explain an independent variable, i.e. how accurate the outcome values are simulated by the model. A good forecast model should have R^2 values close to 1.

MPE reflects the bias of the forecast error, i.e. if the forecasted values present a consistent deviation from the actual value in either a low or high direction. A negative MPE indicates that the forecast is underpredicted compared to the actual value. In this study, the MPE is included to investigate if any of the established correlations from the literature demonstrates a bias trend towards any of the datasets, to reflect the quality of the provided data points.

5.2 Preparing the PVT-database

The provided database was originally comprised of more than 1300 reports from the Norwegian continental shelf (NCS). The database was used to develop new correlation methods; to investigate the validity of the established correlations from literature, not accounting for their reference samples; and to conduct a comparative analysis between the discussed models.

First, a quality assurance of the entire dataset had to be administered, to ensure that the measured field data was physically legitimate. The single most important success criterion to create an accurate correlation model for fluid properties is to have a high-quality dataset. Each PVT report was therefore investigated thoroughly to remove any errors, demonstrative outliers or missing data. Subsequently, the database was checked for any data duplications in both the input properties and the output. This measure is required as repetition of the same data may confuse the surrogate models.

Table 1 presents the data utilized in the different correlations, within the range of input parameters in temperature, oil API gravity, GOR, reservoir pressure, saturation pressure, saturated viscosity and dead oil viscosity. The true viscosity reflects the output viscosity value that the correlation models aim to predict. The data intends to create an authentic representation of the different oil viscosities on the NCS.

Table 1 – Parameter range from the provided Statoil fluid database.

Range of PVT-properties	Dead Oil	Saturated Oil	Undersaturated Oil
Temperature, °F	85 to 338	85 to 347	-
Tank-oil Gravity, API	18 to 55	18 to 52	18 to 55
Solution GOR, scf/stbf	-	79 to 3770	-
Res Pressure, psia	-	-	1704 to 13146
Sat Pressure, psia	-	754 to 8456	880 to 8455
Saturated Viscosity, cP	-	-	0.07 to 7.62
Dead oil Viscosity, cP	-	-	0.41 to 19.58
True Viscosity, cP	0.412 to 19.58	0.08 to 14.35	0.06 to 8.46

5.3 Development of Computational Methods

The following section discuss the relevant types of supervised machine learning algorithms, i.e. systems that require two sets of data. The algorithms analyze a set of training data with known solutions, to produce a function based on inference, to predict the output of the test data. In order to create a reliable correlation model and to avoid biases, the data points used in the training data and the testing data were selected on random.

5.3.1 Radial Basis Function Network

Training the RBFN involves selecting three different sets of parameters: the RBF neuron activation functions; the prototype vector for each of the RBF neurons; and the matrix of the output weights, computed by the Euclidean distance between the RBF neurons and the output nodes.

RBF Neuron Activation Function

The radial basis function is advantageous as the model takes the weighted sum of all neurons, and therefore all points present an influence on the system. There are three common radial basis activation functions, where the function behavior differs relative to the Euclidean distance between the reference and the input vectors. At small distances the functions provide a similar response behavior, as $(\varepsilon r)^2$ approaches zero, while the difference is more evidently displayed at larger distances. The Gaussian function (13) experience an exponentially response towards zero as the Euclidean distance increases. The behavior is favorable when dealing with extreme outlier, as it makes the contribution of the farther points infinitesimal.

The function response at large distances in the multi-quadratic and inverse quadratic is more obviously explained by rewriting the expression in terms of an exponential function, as seen below:

Multiquadric:

$$\varphi(r) = \sqrt{1 + (\varepsilon r)^2} = \exp(\frac{1}{2}\log(1 + \varepsilon r)^2) \tag{48}$$

Inverse quadratic:

$$\varphi(r) = \frac{1}{1 + (\varepsilon r)^2} = \exp(-\log(1 + \varepsilon r)^2) \tag{49}$$

The most distinct difference between the multi-quadratic and the inverse function relates to the sign orientation of the exponent. The exponent of a negative number, as observed in the quadratic function, results in a monotonically inclined response as the distance between the centers and the input increases. The increasing function response relates to a global search pattern, where the algorithm treats the featured search space as one, as soon as the best solution is found. In contrast, the Gaussian and the inverse quadratic function exhibit a local response, i.e. localized function, as the function decreases at greater distances. Compared to the Gaussian, the inverse quadratic function presents a slower response decline at greater distances, as a result to the logarithm in the exponent. A local search pattern will initially try to find the best solution within the vicinity of the starting point, and iteratively begin to find a better solution, relative to the objective function. The use of the different search techniques is dependent on the objective function and the input data; local search will potentially stop when it encounters an extreme outlier, while the global search often is a time-consuming process[39]. The response characteristics are graphically illustrated in Figure 8.

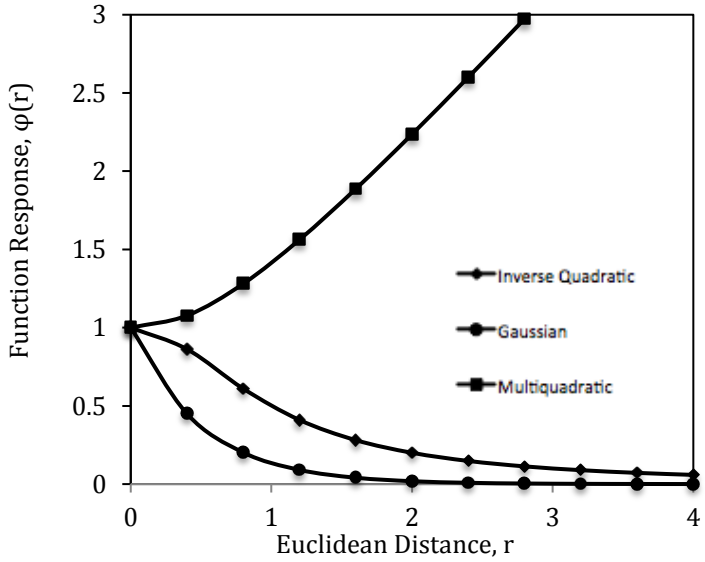

Figure 8 - Illustration of the conceptual search pattern in the radial basis response for each function.

Shape Parameter

The discussed radial basis functions include a shape parameter, ε, that reflects how the model aims to fit the training data. Large values fit the data more closely with respect to the localized functions, while smaller values generate smoother results, and vice versa for the multiquadratic function. A parameter value approaching zero is usually referred to as the functions flat limit[40], as it generates a constant value for the radial basis functions. Figure 9 demonstrates the different responses of the Gaussian function, with various shape parameters.

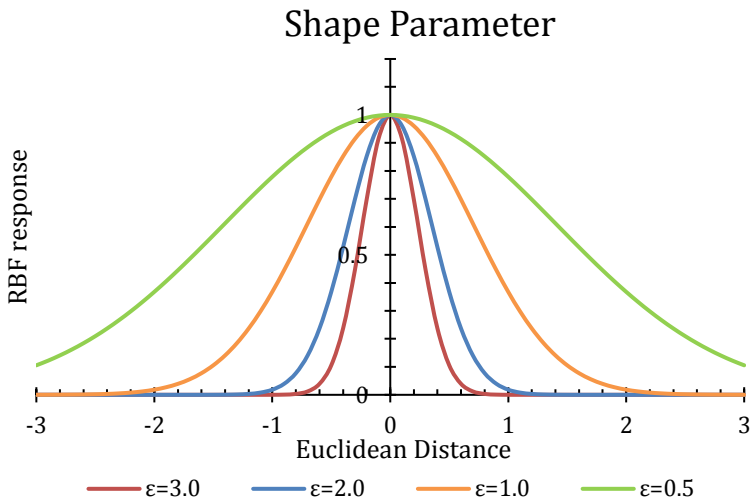

Figure 9 – Illustrated function response with increasing shape parameter.

The best value of ε is still a topic of discussion, as the different values and the corresponding impact on the statistical accuracy is highly dependent on the data sample, and choice of radial basis function[40].

RBF Neurons

The more RBF neurons used in the algorithm, the more accurate response, but with more neurons the more expensive the model becomes. The RBFN is, however, regarded as a universal predictor model; thus, given enough RBF neurons the model is able to compute, and state any arbitrary boundary of a continuous function[41]. K-means clustering has, therefore, been used as an intelligent approach to optimize the algorithm in terms of RBF neurons. The technique is based on performing k-means clustering on the training set and to utilize the cluster centers as neurons.

K-means is one of the simpler unsupervised learning algorithms, i.e. the technique does not require any training with known output variables. The concept is to define data points through non-comparable clusters, one for each k-center. Different center locations create different solutions; thus, an intelligent placement is preferable, e.g. each center should be positioned as far from the other centers as possible. The algorithm is initialized when all data points are linked with the nearest center; subsequently, new k-centroids are computed as barycenter of the cluster from the previous step, then new associations between the same data and the new nearest center are formed. This loop continues until the maximum number of iterations is completed, or until the Euclidean distance is less than a defined threshold. The algorithm aims to minimize a squared error function, related to the following expression[42]:

$$f(\mu) = \sum_{i=1}^{k} \sum_{j=1}^{k_i} (\|x_i - \mu\|)^2 \tag{50}$$

Where,

$$\|x_i - \mu\| = The\ Euclidean\ distance\ between\ x_i\ and\ \mu$$
$$k_i = Number\ of\ data\ points\ in\ the\ i^{th}\ cluster$$
$$k = Number\ of\ cluster\ centres$$

The technique requires that there are no duplications in the input vector, as it is highly unfavorable to have clusters comprised of the same data points from multiple vectors. The algorithm is further very sensitive to any outliers; thus, a high-quality data set is required.

5.4 Development of Particle Swarm Optimization

The goal of any optimization tool is to maximize or minimize an objective function, $f(\vec{x})$, here, $\vec{x}$ is the decision vector that produces the global minimum in a system. This study aims to minimize a function, but the algorithm is nonetheless without loss of generality, as maximizing $f(\vec{x})$ is equivalent to the minimization of $-f(\vec{x})$. The objective of the PSO was to modify the coefficients of the published correlations, while maintaining the functional form[21].

R^2 and MAPE were utilized as objective functions. Both are easily implemented, as MAPE (49) behaves as an optimization function in its original state, while R^2 (50) performs as an objective function by excluding the first term.

The PSO algorithm takes basis in the original coefficients of the respective correlations, while trying to optimize the empirical correlations. The new and modified correlations were then tested on a set of different data points. The first particle in the initial population makes up the original coefficients, while the remaining particles are a random set of numbers, where each particle potentially represents the solution in terms of new coefficients. This randomness makes the PSO response vary considerably from run to run; nevertheless, the unpredictability is regarded as a crucial success criterion for the algorithm. The re-calculations of coefficients using PSO was therefore run 5 times with a number of 1000 iterations, using only the best run to reduce the performance inconsistency.

The success of PSO lies, furthermore, very much in the ability to balance for a global and local search in the algorithm. By changing the inertia weight from a large to a small value promotes the particles to go from a global search to a more local search pattern. The simulations are therefore carried out using a time-varying inertia weight, which is so named for its dependency on the iteration number. The idea is to change the inertia weight dynamically with each iteration so that one facilitates a more progressive search. The inertia weight is therefore set to change according to the proposed scheme of Feng et al[43]. The authors recommended the use of an inertia weight that linearly decreases with each iteration with an additional chaotic term, to present an evolutionary search pattern in-line with the principles presented above. The inertia relates to the following expression:

$$w(\alpha) = (w_{max} - w_{min}) * \frac{\alpha_{max} - \alpha}{\alpha_{max}} + w_{min} * z \quad (51)$$

Where,

$$\alpha = Iteration\ number$$
$$z = 4 * z * (1 - z)$$

The initial z value is chosen randomly as a number between 0 and 1.

Lastly, the particle search space was defined individually for each correlation model, to circumvent any mathematical restrictions, e.g. denominator approaching zero, logarithm of a negative number, and math range error in expressions involving multiple exponents.

6 Results and Discussion

The following presents the empirical response and discussion of the different correlations using the Statoil PVT-database. The results are ordered successively from low to high values, with respect to the MAPE score.

6.1 Correlation Performance of Established Correlations:

Table 2 show that most empirical models demonstrate highly inaccurate results when correlating for dead oil viscosity, where the erroneous behavior could potentially be linked to the provided fluid database. The dead oil viscosity is defined at temperature and pressure at standard conditions, while the database presents dead oil viscosity at atmospheric pressure and reservoir temperature. The PVT-data of dead oil are therefore considered questionable as the presence of gas is not clearly accounted for, and could vary considerably from the authors' reference samples; however, the MPE% column indicates that there is no obvious bias trend, as 5 out of 11 correlations predict a lower forecast than the actual value. The most accurate model is provided by the work of Petrosky & Farshad[35], presenting a MAPE of 25.43%, 0.88 cP MAE and an R^2 value of 0.39. The correlation of Glaso[31] displays bordering results of 27.63 MAPE and 0.33 in R^2. The worst correlation is presented by Bennsion[36], where the correlation presents an extreme error of 3855.52% MAPE. The high error is somewhat expected as the correlation is based on a wide range of viscosity values from a very few heavy crude oil samples; thus, the model is believed to be highly questionable outside the reference range, as suggested by the author.

For saturated oils, the three best correlation models are presented by the work of Beggs & Robinson[29] and Petrosky & Farshad[35], providing a statistical accuracy error in the area of 23.70% MAPE, which is regarded as a relatively high percentage error. The three worst correlation results are presented by the work of Elsharkawy[37], Standing[30] and Labedi[25], where the two latter correlations present notably erroneous results of 80.54% MAPE and 146.29% MAPE, respectively. Again, there is no definite trend reflecting the correlation models display a bias trend to the dataset.

The correlation results of undersaturated oil are considered extremely accurate, where the work of Labedi[25] presents the highest prediction accuracy of 4.34% MAPE, 0.03cP MAE and 0.99 R^2. The correlation results are furthermore considered to be relatively uniform reflected by the estimation criteria only ranging in the area of 4.34 to 7.95% MAPE, 0.97 to 0.99 in R^2 and 0 to -3.56% MPE.

Table 2 - Results of established dead oil viscosity correlations using the Statoil fluid database.

Correlation	R^2	MAE (cP)	MPE (%)	MAPE (%)
Petrosky	0.39	0.88	3.38	25.43
Glaso	0.33	0.97	-8.69	27.63
Bergmann	0.20	1.10	7.72	28.77
Beggs-Robinson	-0.63	1.01	3.19	28.78
Standing	0.34	1.05	-35.55	33.99
Kartoatmodjo	0.15	1.11	-22.12	34.17
Alkhafaji	0.46	1.01	-68.02	37.56
Egbogah	0.20	1.22	22.63	40.12
Elsharkawy	-0.47	1.67	26.13	49.53
Labedi	-10.35	4.02	38.43	111.04
Bennison	-180049.75	188.76	-79987920	3855.52

Table 3 - Results of published saturated oil viscosity correlations using the Statoil fluid database.

Correlation	R^2	MAE (cP)	MPE (%)	MAPE (%)
Beggs-Robinson	0.73	0.33	-7.42	23.65
Petrosky	0.93	0.19	9.08	23.68
Bergman	0.68	0.33	9.47	25.88
Alkhafaji	0.74	0.30	-9.22	25.93
Kartoatmodjo	0.79	0.31	-25.73	28.11
Elsharkawy	0.85	0.28	15.00	28.73
Standing	0.66	0.39	3.84	80.54
Labedi	-6.84	2.09	40.25	146.29

Table 4 - Results of the published undersaturated oil viscosity correlations using the Statoil fluid database.

Correlation	R^2	MAE (cP)	MPE (%)	MAPE (%)
Labedi	0.99	0.03	0.01	4.34
Standing	0.98	0.05	-3.58	5.31
Petrosky	0.99	0.04	0.00	5.50
Elsharkawy	0.97	0.07	0.00	6.95
Kartoatmodjo	0.97	0.07	0.00	7.95

6.2 Empirical Response Performance by Modified Correlations

The results obtained from the published correlations shows a high average error in estimating viscosity values for both dead and saturated oils, which recognizes the need to create more accurate empirical correlation models.

The following section presents the performance of the modified correlations, where the new correlation models are obtained by re-calculating the coefficients through PSO. The re-calculation is computed using the original coefficients as starting point, so that the new correlations have an improved, or the same, fit to the training data, relative to the objective function.

The modified correlations are developed and tested on the number of PVT-reports presented in Table 5, where all reports are selected on random to avoid biases.

Table 5 – Number of PVT reports in the training and testing set, used to develop and evaluate the discussed correlations.

Fluid Property	Number of reports in the training set	Number of reports in the testing set
Dead oil	149	64
Saturated oil	221	94
Undersaturated oil	150	57

6.2.1 Objective Functions

The purpose of the following section is to determine the best choice of objective function to create a correlation model that provides an improvement in all three estimation criteria. The results are ordered successively from low to high MAPE. The RI column presents the relative improvement between the modified and the original coefficients, with respect to the MAPE. The following modified correlations are specified as "M-*Correlation*".

Performance Using R^2 as Objective Function.

Table 6 presents the results of the modified correlations developed through PSO using R^2 as objective function. The table show that Kartoatmodjo & Schmidt[34], Bergman[3] and Labedi[25] provides the three best correlations in terms of MAPE. The modified models present an improvement in 9 out of 11 correlations.

Table 6 – Modified correlations of dead oil using PSO.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Kartoatmodjo	0.85	0.53	18.19	0.47
M-Bergman	0.88	0.52	20.62	0.28
M-Labedi	0.83	0.55	20.64	0.81
M-Petrosky	0.84	0.54	20.88	0.18
M-Alkhafaji	0.80	0.63	23.45	0.38
M-Elsharkawy	0.79	0.65	25.20	0.49
M-Egbogah	0.79	0.65	25.20	0.37
M-Beggs	0.79	0.65	25.20	0.12
M-Glaso	0.79	0.73	34.23	-0.24
M-Standing	0.20	1.43	69.59	-1.05
M-Bennison	-0.09	2.07	90.82	0.98

Table 7 presents the results in estimating the saturated oil viscosity. Note that the correlations are modified by fixing the re-calculated coefficients of dead oil. The most prominent correlation model provided by the work of Elsharkawy[37], presents a MAPE value of 17.76%, R^2 value of 0.97, which corresponds to an average absolute difference of 0.25 cP. The use of PSO with R^2 as objective function provides an improvement in 6 out of 8 correlations.

Table 7 – Modified correlations of saturated oil using PSO using the R^2 as the objective function.

Correlation	R^2	MAE (cP)	MAPE %	RI
M-Elsharkawy	0.97	0.25	17.76	0.38
M-Beggs	0.97	0.25	18.48	0.23
M-Petrosky	0.96	0.24	20.78	0.12
M-Kartoatmodjo	0.96	0.25	21.71	0.23
M-Labedi	0.92	0.36	31.90	0.78
M-Standing	0.90	0.34	32.03	0.60
M-Alkhafaji	0.91	0.35	35.52	-0.35
M-Bergman	0.87	0.44	44.00	-0.68

The correlation performance related to the undersaturated oil viscosity provided by the modified models are presented in Table 8. All correlations display excellent estimation results in all prediction criteria. The model based on the work of Kartoatmodjo & Schmidt[34] provides the best correlation, by demonstrating an improvement of 23%, compared to the original correlation. The modified model provides estimation results well within the measurement uncertainty by 3.36% MAPE, 0.03cP MAE and 0.99 in R^2.

Table 8 - Modified correlations of undersaturated oil using PSO with R^2 as objective function.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Labedi	0.99	0.03	3.36	0.23
M-Kartoatmodjo	0.99	0.04	3.85	0.52
M-Petrosky	0.99	0.04	4.01	0.27
M-Standing	0.98	0.04	4.08	0.23
M-Elsharkawy	0.99	0.04	5.09	0.27

Performance using MAPE as objective function

Table 9 presents the correlation results of dead oil viscosity using MAPE as the objective function. The three most prominent correlations are the modified equations provided by the work of Bergman[3], Glaso[31] and Kartoatmodjo & Schmidt[34], demonstrating a statistical error of 15.08%, 15.38% and 15.58%, in terms of MAPE. The three mentioned correlations show an improvement of 40%, 44% and 83%, respectively. Furthermore, applying the MAPE as objective function shows an improvement in all correlations, excluding the prediction model of Standing[30]. Again, the proposed model of Bennison[36], presents a noticeably erroneous result. Where the negative R^2 reflects that the model fits the data points less than the null hypothesis, which is a horizontal straight line.

Table 9 – Results of modified dead oil viscosity correlations.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Bergman	0.87	0.48	15.08	0.40
M-Glaso	0.89	0.46	15.38	0.44
M-Kartoatmodjo	0.88	0.48	15.58	0.83
M-Petrosky	0.86	0.50	15.67	0.66
M-Elsharkawy	0.84	0.55	16.53	0.92
M-Beggs	0.84	0.55	16.53	0.93
M-Egbogah	0.84	0.55	16.54	0.59
M-Labedi	0.71	0.65	18.55	0.81
M-Alkhafaji	0.77	0.64	19.98	0.54
M-Standing	-0.06	1.30	36.42	-0.10
M-Bennison	-0.91	1.84	51.97	0.99

Table 10 shows the results of correlating saturated oil viscosity. The optimization process is carried out by fixing the re-calculated coefficients of dead oil. All modified correlations show an improvement relative to the original correlation models. The three best modified correlations display highly accurate estimation results, providing MAPE values of 17.04%, 17.20% and 17.51%, from the correlation models of Kartoatmodjo & Schmidt[34], Elsharkawy[37] and Petrosky & Farshad[35]. The results show, furthermore, that all correlations display a noticeably improvement, compared to the statistical accuracy provided by the original models.

Table 10 - Modified correlations of saturated oil using PSO with MAPE as objective function.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Kartoatmodjo	0.98	0.19	17.04	0.39
M-Elsharkawy	0.93	0.28	17.20	0.40
M-Petrosky	0.98	0.22	17.51	0.26
M-Beggs	0.94	0.27	17.62	0.25
M-Bergman	0.72	0.46	18.42	0.29
M-Alkhafaji	0.65	0.47	19.48	0.25
M-Labedi	0.74	0.42	22.22	0.85
M-Standing	0.84	0.35	27.13	0.66

The estimation results of the modified undersaturated viscosity correlations are presented in Table 11. All correlations show an extremely accurate prediction of undersaturated oil viscosity, where the most prominent correlation model is the re-calculated correlation model published by Labedi[25]. Again, the accuracy is well within the measurement uncertainty. The correlation presents an error of 3.35 MAPE, 0.023 cP MAE and 0.99 in R^2, improving the accuracy by more than 23% relative to the original model.

Table 11 - Modified correlations of undersatured oil using PSO with MAPE as objective function.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Labedi	0.99	0.03	3.35	0.23
M-Petrosky	0.99	0.04	3.87	0.30
M-Standing	0.98	0.04	4.08	0.23
M-Kartoatmodjo	0.98	0.04	4.09	0.49
M-Elsharkawy	0.99	0.04	5.09	0.27

Selecting Objective Function

The two different objective functions provide a distinct difference in the empirical test response. The results are justified by investigating the mathematical behavior between the two functions.

The denominator of the R^2 function is a fixed value; thus, only the numerator is being altered. The model seeks to fit the input to the output in the training data, so that a maximum proportion of the total variance in the output is explained by the total variance of the input, i.e. the model aims to minimize the squared value of residuals between a measured value and the corresponding estimated value. The objective function targets a mean value of the measured viscosity, as equation (46) is minimized when the estimated value approach the mean of the summed true values, as expressed in equation (52:

$$\mu_{estimated} = \frac{1}{N}\sum_i \mu_{measured,i}. \tag{52}$$

The use of R^2 as objective function demonstrates unsatisfactory estimation values, the reason is believed to be linked to the function's higher sensitivity to increasing viscosity values, because of the squared residuals expression; consequently, the model seemingly neglects the small values while prioritizing the higher range of viscosity, because of the higher cost. The behavior is exemplified in the correlations results of the training data in Figure 10. Here, the smaller viscosity values in the modified correlations present a worse fit to the data than the original model, while the higher values presents a distinctly better fit to the true values. The corresponding test results are displayed in Figure 11. The modified correlation fit the model nicely in the higher viscosity values, while the smaller viscosity values display more unsatisfactory estimation results, in-line with the empirical response behavior seen in the training set. The performance presents, however, a concern when calculating the MAPE, as the relative error of small viscosity residuals quickly become substantial, e.g. the squared difference between a predicted value, 0.3cP, and a measured value, 0.1 cP, equals to 0.2, while the relative percentage difference equals to 200%. Thus, the MAPE results are often subject to excessively high errors using the R^2 as objective function.

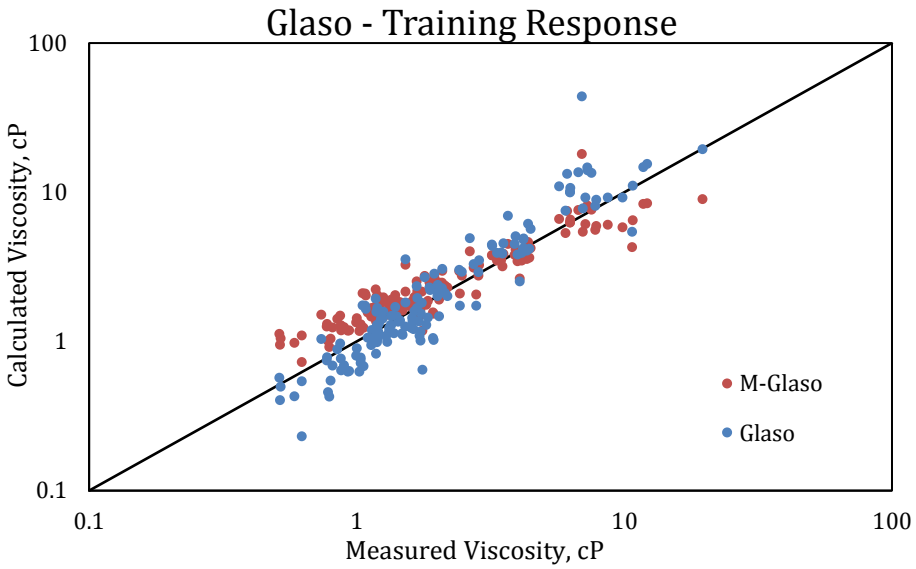

Figure 10 – Dead oil correlation results of training data using the original and the modified model of Glaso[31], developed using R^2 as objective function.

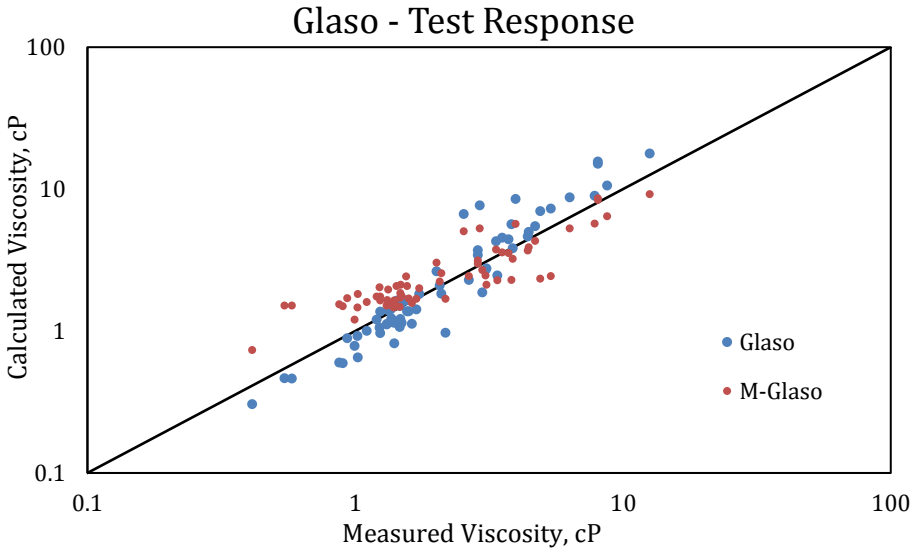

Figure 11 – Dead oil correlation results of test data using the original and the modified model of Glaso[31], developed using R^2 as objective function.

Based on the results, using MAPE as objective function provides much more promising results compared to use of R^2, mainly because MAPE as objective function is more robust to any outliers. Optimization wise, using MAPE as objective function generates a median regression analysis, where the quotient, $\frac{1}{|\mu_{estimated,i}|}$, can be considered as a fixed weight; thus, any regression analysis that aims to estimate the conditional median may be used[44]. The quotient provides a scale dependence, which generates a more robust correlation model, with respect to viscosity values in the lower range. Figure 12 illustrates the different performance between the two objective functions relative to the dead oil training and test data. The figure displays a distinct difference in the lower region, while the estimated values in the higher region presents a more correspondent response.

This study comprises roughly the whole viscosity range demonstrated on the NCS, where low viscosity values are considered equally important as the viscosity in the higher range. The selected base cases of this study are therefore the empirical performances using MAPE as objective function, as the function response is evidently more robust to predict low viscosity.

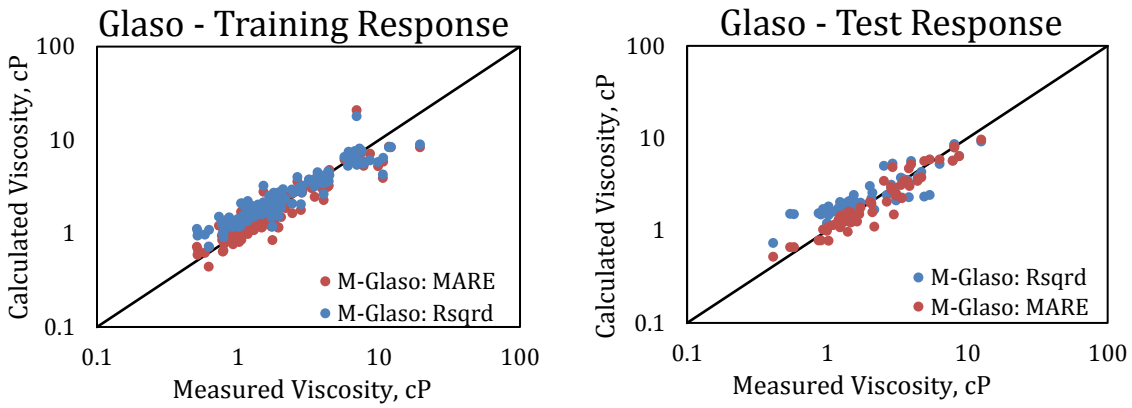

Figure 12 – Calculated dead oil viscosity obtained with the two different objective functions, presented on a log-log scale of both the training data response, and the response on a test set. The most noticeably difference is located in the lower region, while the correlation is more correspondent in the higher range.

6.2.2 Coefficient Analysis – Saturated Oil

The saturated oil viscosity correlations are modified by fixing the re-calculated coefficients from the dead oil viscosity correlations. A sensitivity study was therefore conducted to investigate the empirical response by fixing the original dead oil coefficients, or by setting all coefficients free. The results are successively ordered relative to the MAPE score in Table 12 through Table 16.

Re-calculation of all coefficients

Table 12 presents the correlation performance when re-calculating all coefficients. The results are consistently better correlation results compared to Table 9 with respect to MAPE. The results are as expected, because the PSO algorithm is allowed a higher margin of freedom with an increasing number of coefficients. However, the new coefficients must correspond to a reliable dead oil viscosity correlation model, as this parameter is implemented in all saturated correlations. Table 13 presents the empirical response of the dead oil viscosity using the new coefficients. The results are however highly inaccurate; thus, the correlation models generated by re-calculating all coefficients are discarded.

Table 12 – Empirical response of the modified saturated oil viscosity correlations by setting all coefficients unrestrained.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Beggs	0.90	0.31	16.69	0.29
M-Kartoatmodjo	0.99	0.20	17.18	0.39
M-Bergman	0.97	0.21	17.25	0.33
M-Petrosky	0.97	0.24	17.43	0.26
M-Elsharkawy	0.75	0.40	17.84	0.38
M-Standing	0.84	0.37	18.93	0.76
M-Alkhafaji	0.97	0.22	19.54	0.25
M-Labedi	0.76	0.41	22.39	0.85

Table 13 – Empirical response of dead oil viscosity using the computed coefficients corresponding to Table 12.

Correlation	R^2	MAE (cP)	MAPE (%)
M-Bergman	-1.91	1.25	31.16
M-Standing	-1.60	1.30	36.84
M-Alkhafaji	0.50	1.07	42.71
M-Petrosky	0.02	1.54	45.01
M-Kartoatmodjo	-0.75	2.45	89.00
M-Labedi	-1.35	2.84	100.00
M-Elsharkawy	-43.41	8.20	229.48
M-Beggs	-127.08	11.13	270.99
M-Bennison	-85583.23	125.43	3219.06
M-Bergman	-1.91	1.25	31.16

Fixing original dead oil correlation Coefficients

Table 14 present the empirical response of the modified correlations developed by fixing the original dead oil coefficients. The three best correlations are provided by the work of Bergman[3], Kartoatmodjo & Schmidt[34] and Elsharkawy[37], where the respective correlations presents an MAPE of 17.41%, 17.71% and 17.94%.

Table 14 – Empirical response of the modified saturated oil viscosity correlations by fixing the original coefficients

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Bergman	0.88	0.21	17.41	0.33
M-Alkhafaji	0.92	0.19	17.71	0.32
M-Elsharkawy	0.91	0.20	17.94	0.38
M-Kartoatmodjo	0.92	0.20	18.03	0.36
M-Petrosky	0.93	0.20	18.30	0.23
M-Beggs	0.85	0.24	18.93	0.20
M-Standing	0.48	0.31	21.26	0.74
M-Labedi	0.95	0.22	23.93	0.84

Fixing the original dead oil coefficients present a slightly higher prediction error compared to the base case; however, the difference is not considered significant. A

comparative analysis show that the tuned dead oil viscosity coefficients demonstrate a better MAPE, while the two models presents bordering accuracies in terms of MAE and R^2. The results indicate that the tuned model fits the measured viscosity more closely for values in the lower range, while the correlations present a somewhat similar empirical response in the higher range.

It is believed that the comparable results reflect that the dead oil saturation is not having a major impact when modifying the saturated oil coefficients. The algorithm is seemingly able to find an optimum solution independently of the initial quality of the dead oil correlations, by tuning the coefficients accordingly.

A sensitivity analysis was therefore conducted to study what impact the initial dead oil viscosity correlations have on the saturated viscosity, in terms of statistical accuracy. The investigation was conducted by using the two most prominent original dead oil correlations presented in Table 2, as input parameter for all saturated correlation models. To study if the best dead oil correlations would generate an improvement in the statistical accuracy, compared to using the original set of dead oil coefficients. The empirical response using the dead oil correlation coefficients of Petrosky & Farshad[35] and Glaso[31] in all saturated correlation models are shown in Table 15 and Table 16.

Table 15 – Empirical response of saturated viscosity by fixing and implementing the original dead oil correlation of Petrosky & Farshad.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Elsharkawy	0.93	0.19	17.90	0.38
M-Kartoatmodjo	0.94	0.18	18.00	0.36
M-Petrosky	0.93	0.20	18.31	0.23
M-Alkhafaji	0.93	0.19	18.36	0.29
M-Beggs	0.91	0.20	19.31	0.18
M-Bergman	0.87	0.22	20.87	0.19
M-Labedi	0.93	0.20	22.82	0.84
M-Standing	0.94	0.20	26.50	0.67

Table 16 - Empirical response of saturated viscosity by fixing and implementing the original dead oil correlation of Glaso.

Correlation	R^2	MAE (cP)	MAPE (%)	RI
M-Kartoatmodjo	0.93	0.19	18.00	0.36
M-Petrosky	0.93	0.19	18.07	0.24
M-Elsharkawy	0.91	0.19	18.20	0.37
M-Alkhafaji	0.93	0.18	18.40	0.29
M-Beggs	0.91	0.20	18.98	0.20
M-Bergman	0.84	0.23	20.88	0.19
M-Labedi	0.91	0.21	22.74	0.84
M-Standing	0.94	0.20	25.32	0.69

The results are fairly similar, where the two most accurate models are the modified work provided by Elsharkawy[37] and Kartoatmodjo & Schmidt[34]. The statistical accuracy between the two are also somewhat comparable. It is observed that implementing the most prominent dead oil correlations in the different correlation models for saturated viscosity, lacks the ability to provide a favorable impact on the prediction accuracy. The response is in-line with the proposed theory that the PSO is little affected by the initial dead oil correlations. A combination between published correlations is therefore discarded as an alternative to optimize the prediction models, to avoid unnecessary confusion and to maintain the correlations' functional form.

Based on the presented results and the questionable dead oil dataset, makes it difficult to confidently conclude that the tuned coefficients of dead oil present a reliable improvement in terms of predicting oil viscosity. Consequently, the base case is discarded, and the continued investigation of improving the saturated oil viscosity correlations is conducted using the original dead oil coefficients.

6.2.3 Stability of Coefficients

Based on the results and discussion presented above, new mathematical correlation models are proposed. The modified correlation models are developed by re-calculating the coefficients of published correlations. The models were trained on approximately 70% of the available data, while the remaining 30% were used to test the empirical response accuracy.

In reality, the coefficients will remain unstable until the data size becomes adequately large to present a dependable correlation model. The coefficients can be interpreted as weights to the input parameters, to predict the output. Constantly fluctuating coefficients implies that the forecast model is to some extent based on coincidence and unstructured data. The coefficients are at this stage deemed unpredictable, which makes correlations models developed on such coefficients most speculative.

A systematic analysis was therefore required to assess the coefficients stability. The analysis was conducted by observing the relative difference between the original and

the tuned coefficients when training the algorithm on increasingly data sizes. The coefficient stability is graphically demonstrated in Figure 13 through Figure 16, showing the MAPE relative to the original coefficients on a logarithmic scale.

Dead Oil Correlation Coefficients

Figure 13 demonstrates the stability between the modified dead oil correlation coefficients. An oscillating behavior is observed when using a small training data size, while most coefficients have a tendency to stabilize with increasing training sets. The most deviating correlations are displayed in an individual plot, to maintain a reasonable range on the vertical axis. Surprisingly, the dead oil coefficients display a relatively low MAPE between most of the modified and the original coefficients, the relative difference was expected to be higher, because of the problematic dead oil correlations, whose accuracy are very much dependent on the correlated reference sample. Caution should be exercised in using the adapted correlation coefficients of Bennison[36], Standing[30], and Egbogah[33], as the fluctuating behavior indicates that the coefficients are still subject to change. The coefficients of interest are the most accurate modified correlations of dead oil viscosity, based on the work of Bergman[3], Glaso[31] and Kartoatmodjo & Schmidt[34], where all demonstrates a steady plateau in the higher percentage range.

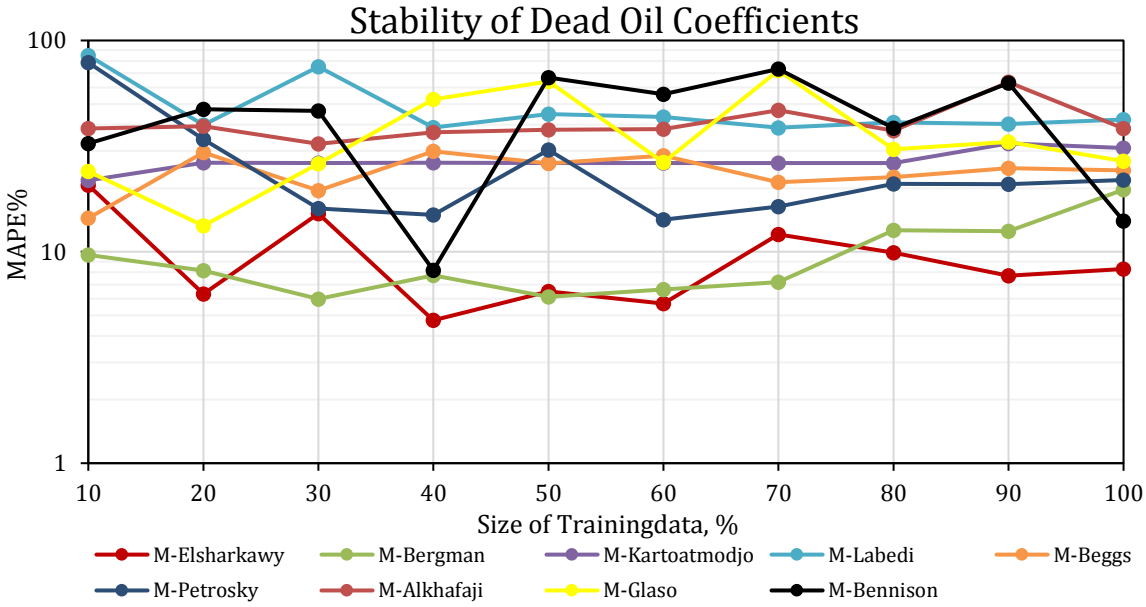

Figure 13 - MAPE development of the modified dead oil correlation coefficients with increasing training data size, relative to the original coefficients

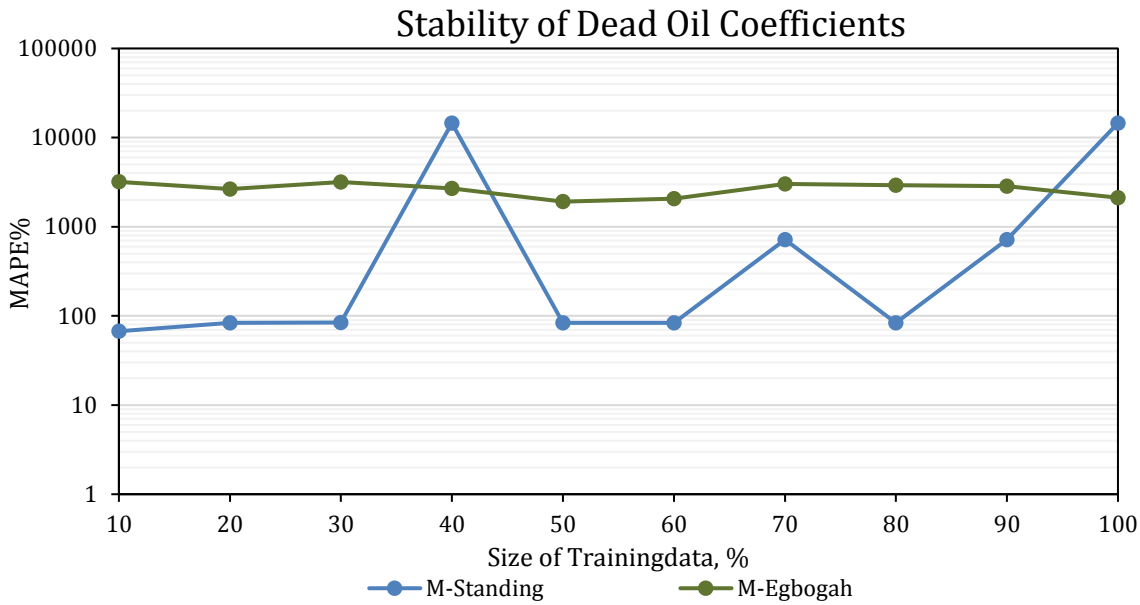

Figure 14 - Development of how the modified dead oil correlation coefficients change with increasing training data size, relative to the original coefficients

Saturated Oil Coefficients

The coefficients in Figure 15 demonstrate an oscillating trend when the number of training data points are low, while more consistent MAPE values are observed in the later stage. A stable plateau is obtained for most correlations from 70% on the abscissa, excluding the modified Standing[30] correlation. The latter correlation presents a noticeably fluctuating behavior throughout the analysis, while the other correlations remain near constant in the last third. The coefficients of interests are the modified models of Bergman[3] and Alkhafaji[32], where both correlations present reliable values in the upper third of the plot.

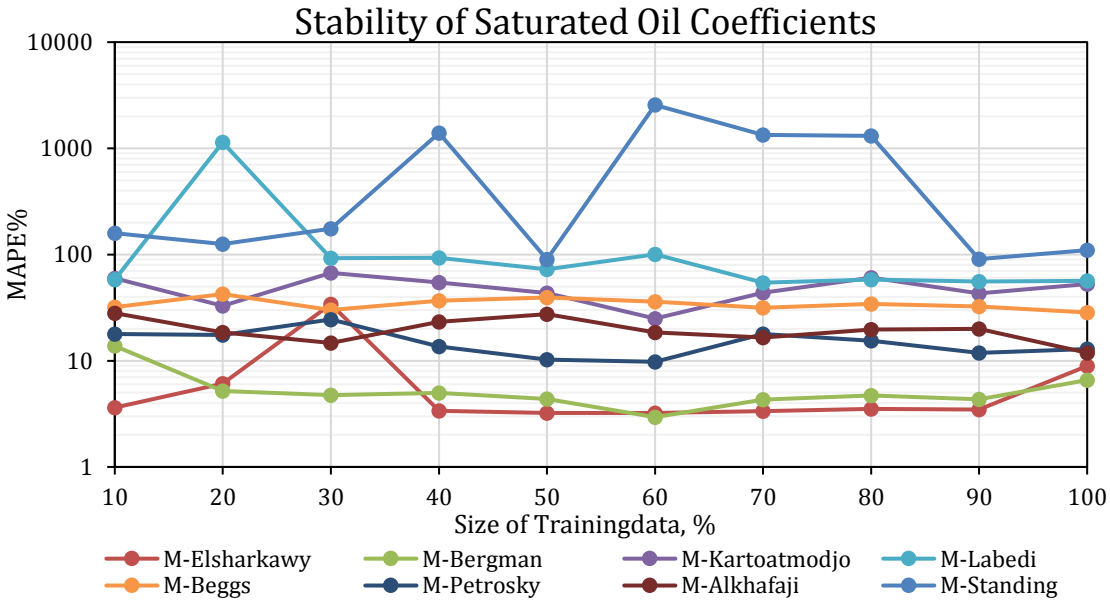

Figure 15 - The development of coefficient variation in terms of mean absolute percentage error plotted on a logarithmic scale, to investigate the stability and validity of the re-calculated correlation coefficients.

Undersaturated Oil Coefficients

Most correlations for the undersaturated oil correlations display the same trend with a stable plateau in the later stage. The deviation relative to the original coefficients is high, reflecting that the coefficients have been severely modified. The altered model of Labedi[25] and Petrosky & Farshad[35] are the two most prominent correlations, as seen from Table 11. The modified model based on the work of Petrosky & Farshad presents a deviating behavior in the later stage, which links some uncertainty to the correlation model; however, the modified correlation model of Labedi is considered to display adequately consistent coefficients values throughout the analysis.

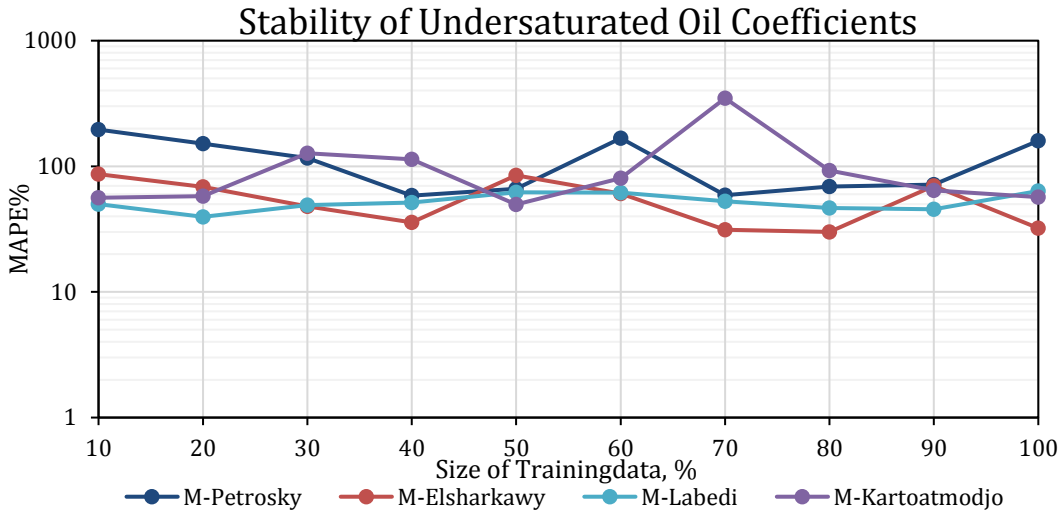

Figure 16 - Stability of undersaturated oil correlation coefficients plotted against size of training data.

6.2.4 Proposed Modified Correlation Models

The previous section strongly indicates that the most accurate and pertinent correlations are valid and robust in terms of coefficient stability, i.e. that the coefficients are preserved even at more excessive datasets. At this point, the PSO algorithm is able to find suitable weights to the input parameters to fit a correlation model. Based on these results, one can confidently conclude that modified correlations are consistent. The following presents the proposed correlation models for dead, saturated and undersaturated oils:

Dead Oil Viscosity

The performance of the published dead oil correlations displays a wide span in statistical accuracy. The erroneous behavior could be linked to the questionable dataset, as discussed in section 6.1, or originate from the correlation method of dead oil viscosity, which is solely a function of two parameters. The two-parameter approach is considered problematic, as two dead oil samples may present a large difference in viscosity, while having the same API and temperature. The dead oil viscosity correlations are further based on the assumption that the fluid is characterized as Newtonian[45], which is not always the case, considering that viscosity is not a state property. Some viscosity correlations may therefore contradict the assumption related to the fluid behavior, especially at higher viscosities. The dead oil viscosity is a function of asphaltic and paraffinic content, which makes it extremely

complex to develop an empirical correlation model without loss of generality, as the reference samples from one region to the next could potentially present massive differences, which attribute in highly inaccurate empirical prediction models.

Nevertheless, the proposed correlation model is superior to all established models, improving the correlation results by 39.7%, relative to the best original correlation expression. The model is based on the work of Bergman[3], and demonstrates a statistical accuracy of 15.08 MAPE and 0.87 in R^2. The correlation performance provides consistent viscosity values to the 45° straight line in the lower range, while a more deviant correlation response is displayed in the three uppermost viscosity values, as shown in Figure 17. The proposed correlation is related to equation (53, where the corresponding coefficients of both the modified and the original correlation model is presented in Table 17.

$$\mu_{od} = -1 + e^{e^{a_1 - a_2 * \gamma_{API} + a_3 * \gamma_{API}^2 - (a_4 - a_5 * \gamma_{API}^2) \ln(T+310)}} \quad (53)$$

Table 17 - Coefficients of proposed and original dead oil correlation coefficients

Coefficients (μ_{od} Correlation)	Value of Modified Correlation	Value of Original Correlation
a_1	17.864000	22.330000
a_2	0.157680	0.194000
a_3	0.000264	0.000330
a_4	2.560000	3.200000
a_5	0.014800	0.018500

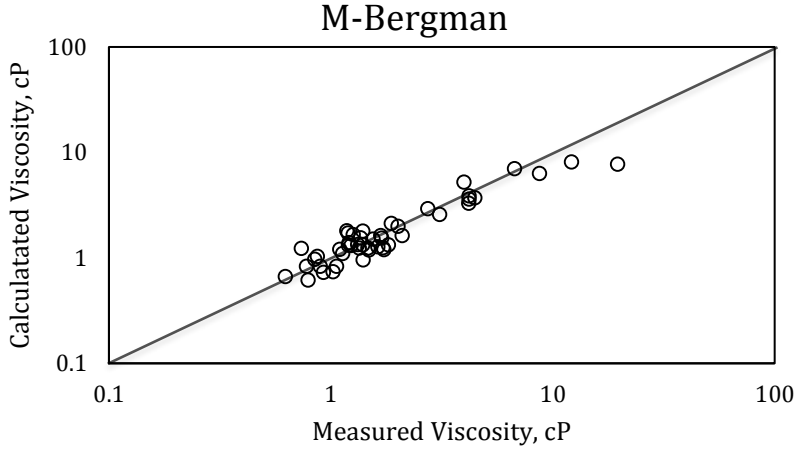

Figure 17 – Proposed correlation model to predict dead oil viscosity, compared to measured viscosity values.

As discussed, the model should be used with caution, considering the uncertainty related to the questionable PVT data used as reference. However, Figure 18 show that the physical trends of the correlated dead oil viscosity are corresponding nicely with the general effects of temperature and API, where a decrease in temperature or API results in an increasingly viscosity response. Furthermore, Table 2 shows that the original models do not present a bias towards the dataset, which indicates that the

dataset does not reflect a consistent deviating trend from the reference data, used to establish the correlation models from literature. Based on these results, more confidence is added to the proposed dead oil correlation; however, it is not possible to conclude with conviction that the model is valid and reliable without testing it on an unbiased data-set.

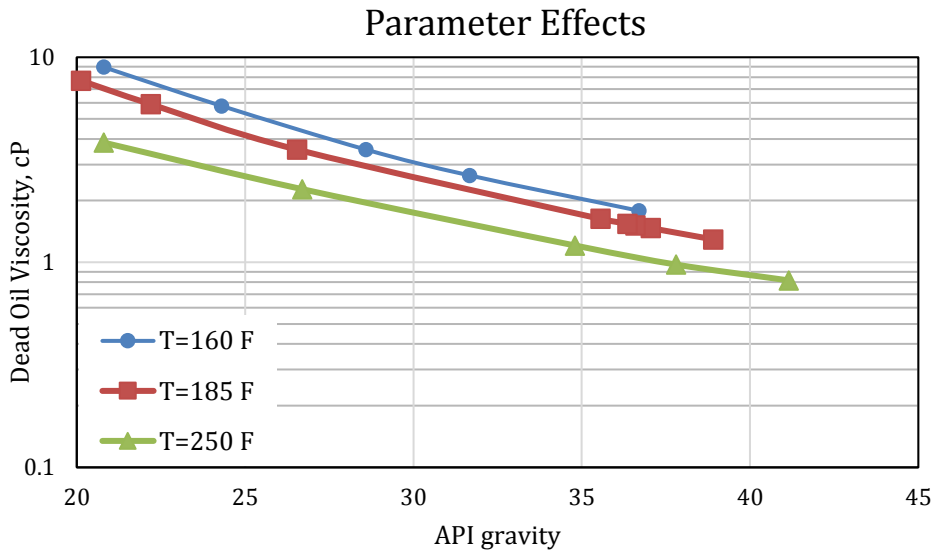

Figure 18 – Proposed dead oil viscosity correlation plotted as a function of API at various temperatures.

Saturated Viscosity

Eight published correlations were investigated as basis to create a new and improved estimation model for saturated oil viscosity. The proposed correlation model is modified by keeping the coefficients of the original dead oil correlation fixed, rather than using the modified dead oil coefficients, even though the latter correlation proves slightly more accurate results. The correlation is discarded because of the debatable data set used as reference to develop the model, which makes it uncertain if this correlation truly is an improved estimation model.

The proposed new correlation model is based on the original work of Bergman[3], presenting an improved accuracy of 26.4% relative to the best original correlation model. The performance is graphically shown in Figure 19. A low dispersion relative to the reference line is observed throughout the whole viscosity range, with only a few deviant points. The results indicate that the prediction model is satisfactorily in agreement with the true viscosity. The final re-calculated coefficients are based on all data; thus, it is reasonable to assume that the modified model presents either the same, or an improved empirical response, compared to the stated statistical accuracy. The new correlation model is related to the following expression, where the correlation coefficients of both the original and proposed model is presented in Table 18:

$$\mu = A\mu_{OD}^{B} \tag{54}$$

Where,

$$A = \exp(a_1 - a_2 \ln(R_s + 300))$$

$$B = a_3 + \frac{a_4}{R_s + 300}$$

Table 18 – Coefficients for the proposed and original correlation model to predict saturated oil viscosity.

Coefficients (μ_{ob} Correlation)	Values of Modified Correlation	Values of Original Correlation
a_1	4.678758	4.768000
a_2	0.835810	0.835900
a_3	0.469241	0.555000
a_4	133.500000	133.500000

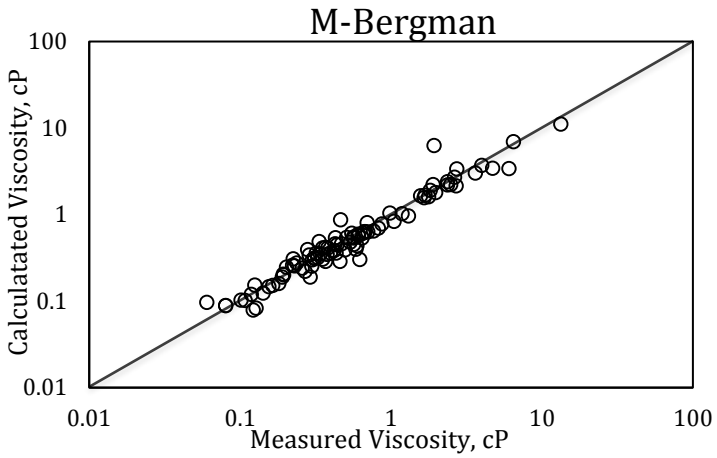

Figure 19 - Proposed correlation model for saturated oil viscosity, compared to measured viscosity values

Undersaturated Oil Viscosity.

The overall performance of the published viscosity correlations for undersaturated oil presented highly accurate results, where the best correlation provides an empirical response of 4.34% MAPE, which is believed to be well within the statistical uncertainty related to laboratory measurements. The low error is to some extent expected as the undersaturated oil viscosity is strongly correlated to pressure and saturated oil viscosity.

Nevertheless, the proposed modified correlation model for undersaturated oil viscosity presents superior results by tuning the model of Labedi[25], compared to the original correlations. The model provides an accuracy of 3.35% MAPE, 0.03cP MAE and 0.99 in R^2, using only 70% of the total data set. The highly accurate empirical performance is emphasized graphically in Figure 20, where the response is observed to be virtually in perfect agreement with the diagonal reference line. The modified correlation model is related to expression **Error! Reference source not found.**, with c orresponding coefficients of proposed and original correlation model presented in Table 19.

$$\mu_o = \mu_{ob} - \left[(1 - \frac{p}{p_b})(\frac{10^{-a_3} * \mu_{od}^{a_1} * p_b^{a_2}}{10^{a_4 * \gamma_{API}}}\right] \quad (55)$$

Table 19 – Coefficients for the proposed and original undersaturated oil viscosity correlation model.

Coefficients (μ_o Correlation)	Value of Modified Correlation	Value of Original Correlation
a_1	1.241755	2.488000
a_2	0.894156	0.903600
a_3	0.114034	0.615100
a_4	9.952798E-03	1.976000E-02

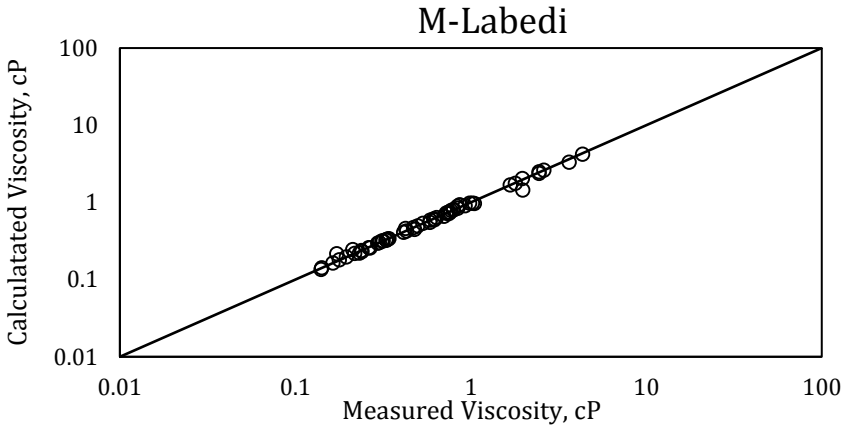

Figure 20 - Proposed correlation model to predict undersaturated oil viscosity, compared to measured viscosity values.

No further sensitivity analyses were conducted to improve the model, as the statistical error is considered to be well within a satisfactory range.

6.3 Performance of Surrogate Models

This section provides information about the statistical accuracy of the different surrogate models presented in the theory section. The methodology of RBFN is discussed in Chapter 5, while the algorithm of Kriging and Neural Network is provided by the work of Arief et al[14]. The following is a study related to the performance of RBFN in comparison to Kriging, Neural Network and the most accurate modified correlation models.

Dead Oil Viscosity

The statistical accuracies to predict dead oil viscosity are presented in Table 20. The correlations were run using only temperature and oil gravity as input parameters, as all discussed empirical correlations are based on these two reservoir properties. The response implies that the two mentioned parameters are able to create a sufficient correlation model for dead oil viscosity, which is consistent with the suggested concept of Beal[27].

Table 20 – Response accuracy of surrogate models correlating for dead oil viscosity.

Correlation	R^2	MAE (cP)	MAPE (%)
M-Bergman	0.87	0.48	15.08
RBFN	0.88	0.44	16.89
Kriging	0.89	0.46	17.14
Neural Network	0.13	0.71	28.66

The RBFN algorithm provides the most precise surrogate model. The response is regarded as adequately accurate, considering the nature of the dead oil. The Gaussian RBF was selected as activation function, because of its response to outliers; however, the different RBFs provide practically the same empirical response accuracies. The numerical stability is provided in Figure 21, demonstrating the behavior of MAPE relative to a changing shape parameter.

A small shape parameter in the correlation algorithm of RBFN is recommended, to create a narrow basis function, i.e. to generate a smooth fit to a wide span of output values. The shape parameter provides numerically stable values from ε=1e-8, values below this point neutralizes the basis function to act as a constant. The best accuracy is obtained by using a shape parameter equal to 1E-5, the response in this area is further considered to be numerically stable.

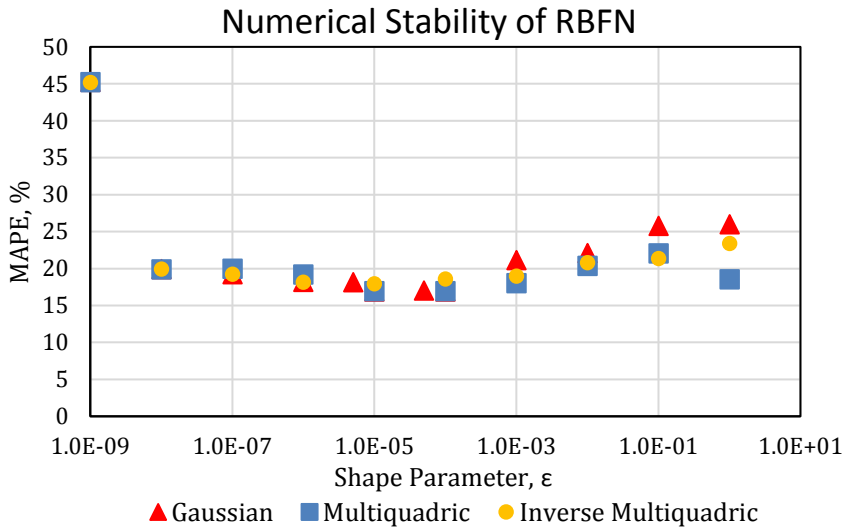

Figure 21 – Numerical stability of different RBF, relative to different shape parameters.

The correlated viscosity values are presented below as a cross-plot related to the measured values on a log-log scale. Some deviation is observed for higher viscosities relative to the 45^0 horizontal line, but the prediction model is still considered to be satisfactorily accurate. However, some uncertainty is linked to the model, as the quality of the dead oil data set is not guaranteed. Despite the satisfactory results of RBFN, the model is inferior to the proposed modified correlation, which presents higher accuracy in all estimation criteria.

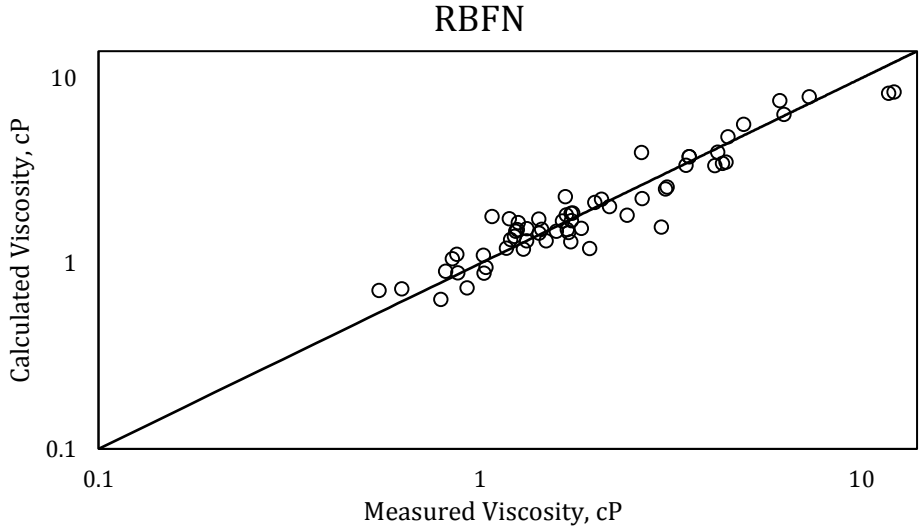

Figure 22 – Correlated dead oil viscosity as a function of API and temperature, plotted against measured viscosity.

Saturated Oil viscosity

Table 21 presents the correlation results using the surrogate models to predict the saturated oil viscosity. The models are trained on four input variables; temperature, API, GOR and saturation pressure, equivalent to the input of the reviewed empirical correlation models. Kriging demonstrates the most accurate response amongst the surrogate models, the response is, however, not an improvement compared to the proposed correlation model.

Table 21 - Results of surrogate models using four input variables.

Correlation	R^2	MAE (cP)	MAPE (%)
M-Bergman	0.88	0.21	17.41
Kriging	0.88	0.23	18.99
RBFN	0.92	0.21	21.7
Neural Network	0.81	0.24	22.68

Although the surrogate models are superior to the original correlation models, the results are still not as accurate as expected. An uncertainty study was therefore conducted to investigate the saturated oil viscosity relation to saturated density, the results are presented in Table 22.

Table 22 - Results of surrogate models using five inputs, including saturation density.

Correlation	R^2	MAE (cP)	MAPE (%)
Kriging	0.81	0.21	17.13
M-Bergman	0.88	0.21	17.41
RBFN	0.84	0.2	19.75
Neural Network	0.44	0.33	26.05

The Kriging provides the most accurate model, where the results are an improvement of 1.86%, in terms of MAPE, making the surrogate model superior to the proposed

correlation model. The improved response is justified by looking at the comparable behavior between viscosity and density. An increment in pressure creates a corresponding compressibility effect for both parameters. A rise in pressure induces an increment in both the viscosity and density, whereas an increase in gas content induces a decline in viscosity and density. The results corresponds to the findings of Arief et al.[14], that the prediction performance of Kriging exceeds the discussed correlations from literature. The authors reported an estimation error of 20.7 MARE%, while this study presents a higher accuracy by implementing the saturated density.

The RBFN improved the accuracy by 1.95%, compared to the base case, emphasizing the relation between the viscosity and density at the saturation point. The empirical response using the different radial basis function is coinciding in the area of 20 MAPE%, while a more distinct difference is observed in the response with increasing shape parameters. The Gaussian and inverse quadratic function quickly presents inadequate values to fit the test data, as the function response increases, while the Multiquadric show less deviation with increasing shape parameter, as the function response declines. The different behaviors are presented in Figure 23. A distinct outlier is observed in the Gaussian RBF. The behavior is potentially linked to overfitting. The phenomenon occurs in machine learning when a model perfectly fits the training data so that the correlation loses its generality. The neurons are overreacting to small deviations in the test data relatively to the training input. The model is in this case considered to describe random noise instead of a true connection between the input and output data.

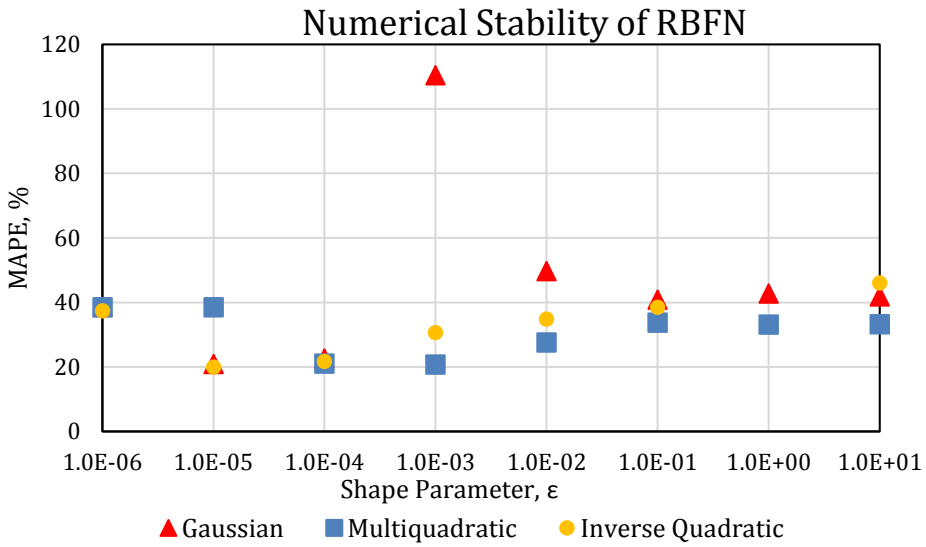

Figure 23 - The different radial basis functions plotted against increasing shape parameters. The Multiquadric behaves steady in the lower MAPE compared to the others, because of the decreasing function response.

Undersaturated Oil Viscosity

All published empirical correlations discussed in this thesis include reservoir pressure, saturation pressure and saturation viscosity as independent variables to estimate the

undersaturated oil viscosity. The surrogate model base case analysis was therefore carried out using the same variables, the results are displayed in Table 23.

Table 23 - Statistical accuracy of the three surrogate models using reservoir pressure, bubble point pressure and saturation viscosity. The table further includes the empirical response of the proposed modified correlation.

Correlation	R^2	MAE (cP)	MAPE (%)
M-Labedi	0.99	0.03	3.35
Kriging	0.98	0.06	8.36
RBFN	0.9	0.15	9.45
Neural Network	0.98	0.11	10.89

The surrogate models provide statistically accurate results, the performance is, however, secondary compared to the modified correlation model which presents extremely accurate results in all estimation criteria. Kriging and the RBFN presents a distinct difference in statistical accuracy to modified correlations model, compared to the two previous oil correlations. The response is believed to be linked to the cost functions, which squares the residuals. Most of the undersaturated oil viscosities are in the lower viscosity range; consequently, the respective models prioritize only a small share of the viscosities, as these present a higher cost. The behavior is more evidently displayed in Figure 24.

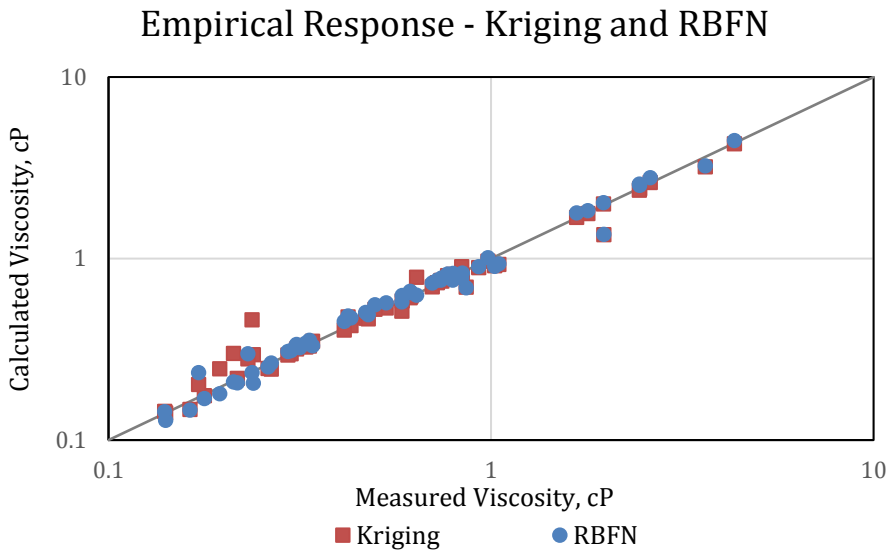

Figure 24 – Empirical correlation response for undersaturated oil viscosity of Kriging and RBFN

The RBFN model was developed using the Multiquadric radial basis function, as the accuracy and numerical stability of the model is consistently improving with a more sensitive function response; thus, a high shape parameter in the area of 10,000 is recommended, to ensure consistent correlation results. The behavior of the Gaussian and the inverse Multiquadric, is justifiable by looking at the local search pattern, where the search algorithm is potentially trapped in local extrema, which consequently completes the search prematurely.

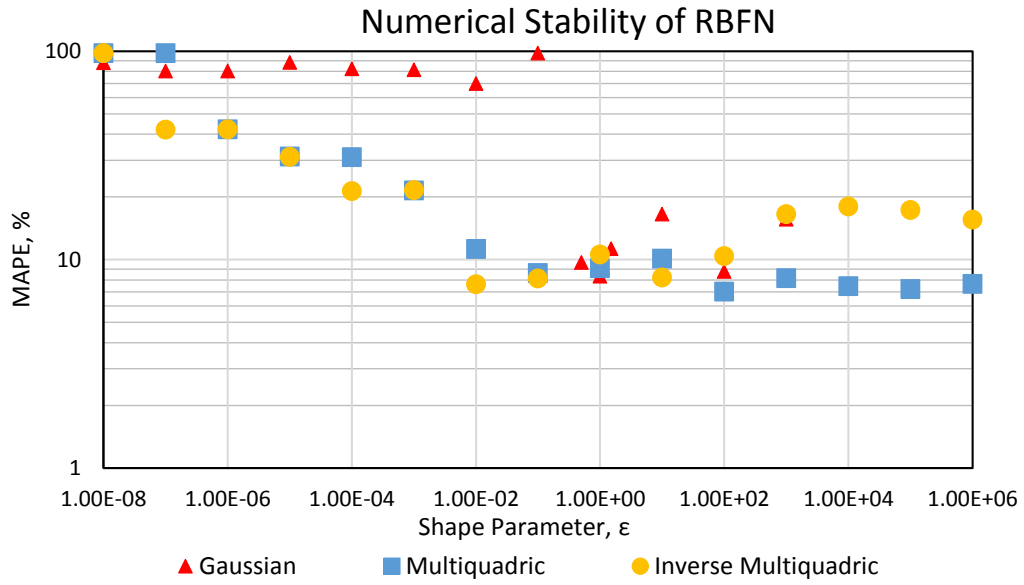

Figure 25 - The different radial basis functions plotted against increasing shape parameters. The Multiquadric provides the most stable numerical stability at high shape parameters, because of a more sensitive function response.

The surrogate models were further used to investigate if other parameters could explain the dependent output variable more competently. Oil viscosity increases proportionally with increasing pressure above the saturation pressure. The pressure dependency relates to a single-phase state; consequently, the solution GOR remains constant. Pressure is therefore considered to be the single most important independent variable to predict the viscosity of undersaturated oil. Numerous studies have been made to express the correlation of undersaturated oil viscosity to saturated oil viscosity, and to the pressure increment above bubble point. The most suitable function form was found by plotting $(\mu_o - \mu_{ob})$ vs. $(p - p_b)$[23]. The relation between the two provides a series of straight lines through the origin for various oils, where the linear equations were found to be a function of dead oil viscosity. The undersaturated viscosity may therefore be considered as a function of dead oil viscosity, saturated oil viscosity and saturation pressure. A sensitivity study was therefore carried out to see if implementing the dead oil viscosity would generate an improved accuracy on the RBFN, the results are shown in Table 24.

Table 24 - Accuracy of undersaturated oil correlations to predict viscosity using surrogate models with saturation pressure, saturated viscosity and dead oil viscosity as input data.

Correlation	R^2	MAE (cP)	MAPE (%)
M-Labedi	0.99	0.03	3.35
RBFN	0.98	0.06	7.22
Kriging	0.98	0.05	8.06
Neural Network	0.98	0.05	10.06

The model is more in agreement with the test data than the base case, where all surrogate models show an improvement; consequently, the dead oil viscosity

parameter is considered to present a favorable relation to the undersaturated viscosity, compared to only using the base case parameters. However, the proposed correlation model is still exceeding all correlations in terms of statistical accuracy.

6.4 Error Analysis

The following section presents a comparative analysis between the most prominent correlation models proposed in this thesis, and the corresponding original correlation model. The aim of the study is to verify the validity of the dead and saturated oil viscosity correlations, with respect to specified input ranges. The undersaturated viscosity correlations are not included in this analysis, as the correlation performance is believed to be well within a satisfactory accuracy, regardless of the input parameter range. The error analysis is illustrated in Figure 26 through Figure 28.

The most accurate viscosity correlations are the modified work of Bergman, RBFN and Kriging. The different correlations are functions of temperature, oil API gravity and GOR. The error analysis was conducted by categorizing temperature and API gravity into high and low subsets, while the GOR and targeted viscosity values are sectioned into specified ranges. Based on the temperature sensitivity analysis, both the RBFN and the modified correlation are little affected by either low or high temperature, while the original model of Bergman presents a highly inaccurate response at temperatures below 190°F. As expected, a more erroneous response is observed in the lower range in oil API gravity. In general, viscosity increases with heavier oil samples, the behavior is linked to an increase in heavier molecular components, such as asphaltenes. The increase in heavy molecules is troublesome, with respect to empirical viscosity correlations, as it may ultimately generate two different viscosity behaviors at the same API and temperature values. The RBFN is seemingly handling the higher viscosity values better than the other correlations, which is believed to be linked to the objective function of the model. The RBFN aims to minimize the squared distance between a reference vector and the calculated data, where the higher viscosity values presents a higher cost; consequently, these values have priority. The findings are further emphasized in the error analysis, with respect to GOR, where the RBFN is the most accurate prediction tool at low gas-oil ratios, i.e. at higher viscosities.

Based on the graphical interpretations, the modified correlation model of Bergman presents the overall most consistent empirical performance, as the modified correlations models do not demonstrate any distinctly inaccurate responses at different reservoir conditions. The consistent accuracy adds more confidence that the model is a reliable and a true improvement, compared to the existing correlation models.

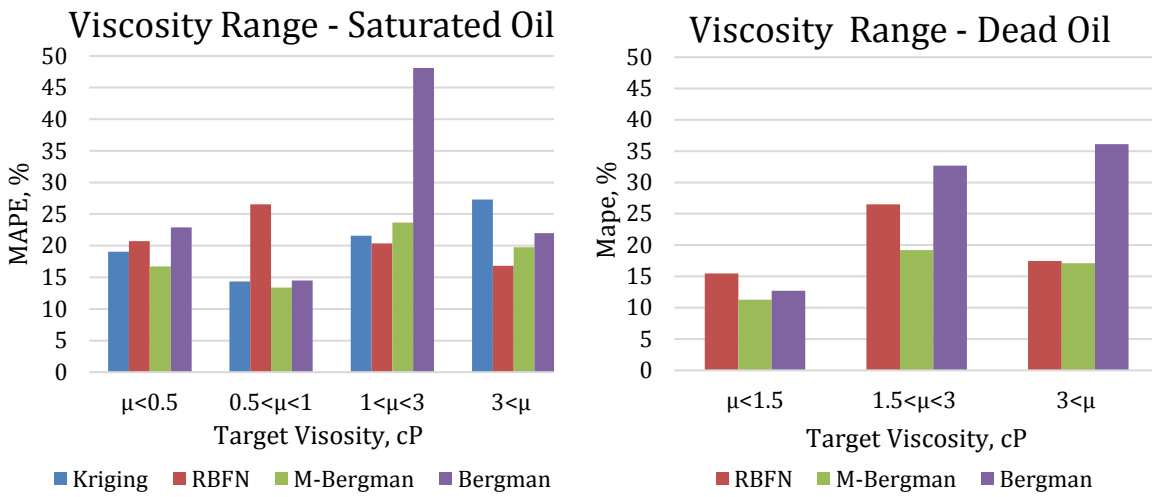

Figure 26 – Error Analysis of dead oil and saturated oil viscosity to investigate the reliability of the presented models, with respect to different range of viscosity.

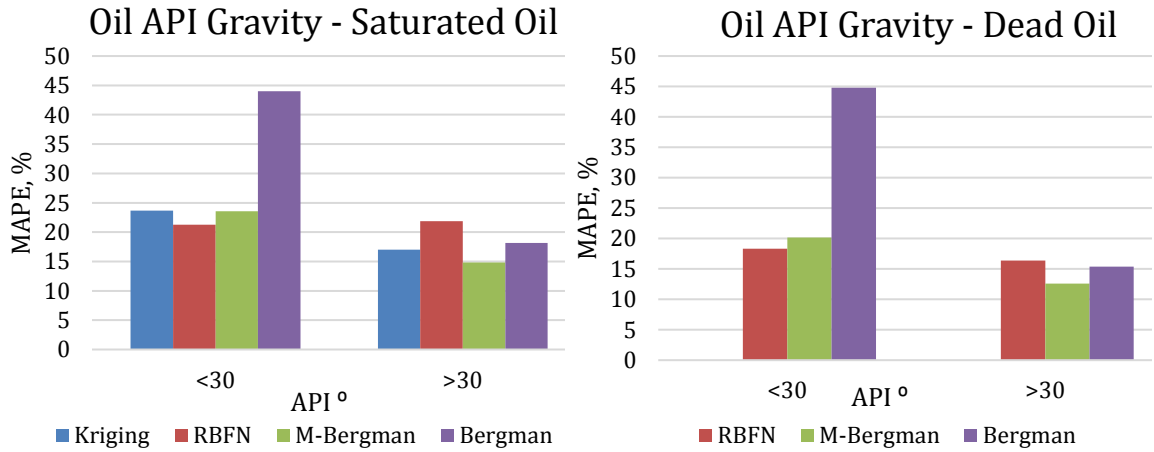

Figure 27 – Error Analysis of dead oil and saturated oil viscosity to investigate the reliability of the presented models, with respect to different range of oil API gravity.

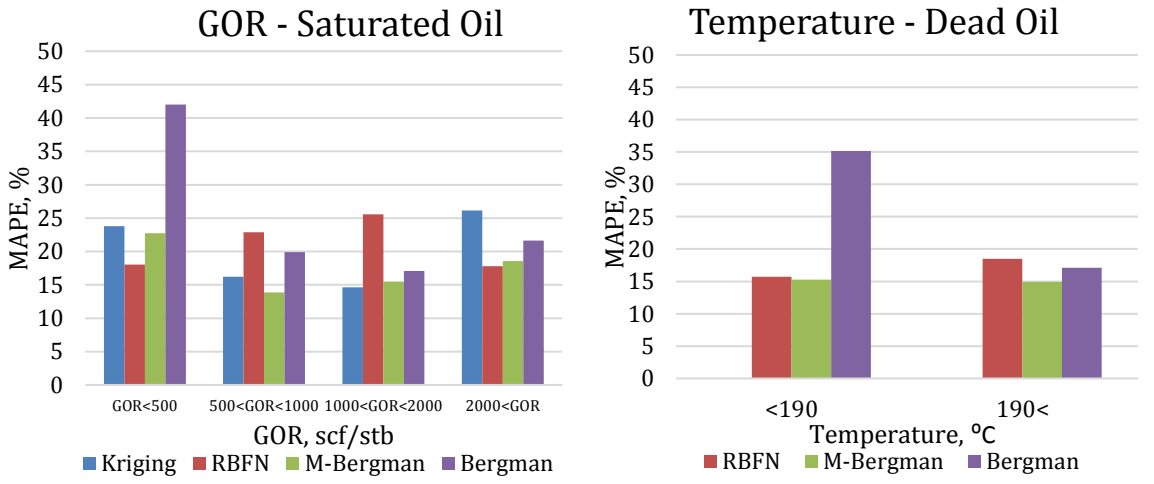

Figure 28 – Error Analysis of dead oil and saturated oil viscosity to investigate the reliability of the presented models, with respect to different range of GOR and temperature.

Conclusions

The primary objective of this thesis was to develop a new and improved viscosity correlation model for dead, saturated and undersaturated oils on the NCS. The concluding remarks of this study are summarized in the following bullet points:

- The study is based on a comprehensive fluid database, representing the range of fluid properties on the NCS. The dataset has been thoroughly quality assured, in order to develop accurate and reliable prediction models. However, uncertainty is linked to the dead oil data, because of the experimental measurement conditions.
- The surrogate models require a large dataset, as the algorithms orient around statistical patterns with respect to the input variation, rather than the traditional mathematical approach, ultimately no mathematical correlation equations are created; thus, the models are considered less functional, compared to the explicit correlation models.
- Based on statistical accuracy, the proposed prediction model for dead oil viscosity, eq.53, demonstrates a significant improvement, compared to the original correlations. However, discretion is recommended, considering the questionable PVT-data used as reference.
- The modified correlation model of saturated oil viscosity, eq. 54, demonstrates a higher accuracy than all the presented correlations, where the re-calculated model provides an improved MAPE of 26.4%, relative to the best original correlation model. Based on the performance one can confidently conclude that the proposed model is a significant improvement to the established correlations. However, the model is secondary compared to the response of Kriging, which exceeds all correlations when including the saturated density as input parameter. It is therefore recommended to use Kriging when correlating for saturated oil viscosity, provided that the model is practically available.
- From the correlation results of undersaturated oil viscosity, the empirical response of the original correlations was highly accurate; nevertheless, the new correlation model, eq.55, demonstrates a significantly improved accuracy well within the measurement uncertainty, compared to the discussed correlation models from literature.
- MAPE demonstrates considerably better correlation results as objective function in the PSO algorithm, compared to the use of R^2.
- It is observed that the proposed modified correlation models are valid and robust in terms of numerical stability, i.e. the coefficients are preserved even at more excessive datasets.
- Regarding the RBFN, a global search orientation proves superior compared to a localized function, as the local search is believed to get trapped in local extrema. The multiquadric function demonstrates a higher accuracy and an extended plateau with respect to an optimum function response, compared to the other RBF functions, when correlating for saturated and undersaturated oil viscosity.

7 Future Work

The following outlines the proposed course for future research related to this dissertation:

- The proposed correlations are based on field-measurements from the NCS. Further investigations on the proposed correlations should be carried out to test the correlations against an unbiased databank, in order to study the general validity of the models outside the reference region.
- If the proposed models fail provide superior results outside the NCS, the same correlation approach should be conducted elsewhere. To investigate if the PSO algorithm provides equally prosperous results in other reference regions, as to this study.
- As mentioned, the most accurate model to predict saturated oil viscosity is the modified correlation model based on the re-calculated coefficients from the modified dead oil correlations, this model was neglected based on the uncertainty linked to the PVT-data. A study should therefore be conducted to investigate the validity of the proposed dead oil correlation. If proven valid, a further study should be conducted to investigate the consistency and performance of the saturated oil viscosity correlation models using the re-calculated dead oil coefficients.
- With respect to the objective function, it was argued that the MAPE was the best alternative. However, it would be interesting to study the impact using the MAPE with the forecast value in the denominator, as objective function in the PSO.
- The extension of predicting other fluid properties using the RBFN should be addressed, to explore if the surrogate model could improve existing correlations, with respect to bubble point pressure, oil formation volume factor and oil density.
- Seeing as the reference samples in this study are regarded as medium to light crude oils, a study should be carried out to test the surrogate models and the PSO on heavy oil samples at API<22.3^0. The heavier oil samples characteristically present higher viscosity values; consequently, a sensitivity study should be administered, with respect to the objective function in the PSO, as higher viscosity values might induce more favorable results using R^2 more than MAPE.

Nomenclature

α	– Iteration number
ε	– Shape Parameter
v_x	– Velocity in Applied Stress Direction (m/s)
ν	– Kinematic Viscosity (cSt)
ξ	– Viscosity Reducing Parameter
ρ	– Density (g/cm^3)
ρ_c	– Critical Density (g/cm^3)
ρ_r	– Reduced Density (g/cm^3)
ρ_S	– Sphere Density (g/cm^3)
τ	– Shear Stress per Unit Area (Pa)
Υ_{API}	– Stock Tank Oil Gravity (API^0) ($°_{API} = \frac{141.5}{SG} - 131.5$)
η	– Dynamic Viscosity, cP
μ	– Absolute Viscosity (cP)
μ^*	– Low Pressure Gas Mixture
μ^*	– Low Pressure Gas Mixture
$\mu_{estimated}$	– Calculated Viscosity
$\mu_{measured}$	– True Viscosity
$\bar{\mu}_{measured}$	– Average Measured Viscosity (cP)
μ_{od}	– Dead Oil Viscosity (cP)
μ_{ob}	– Viscosity at Bubble Point (cP)
μ_o	– Undersaturated Viscosity (cP)
$\vec{x}$	– Decision Vector
$\frac{\partial v_x}{\partial y}$	– v_x gradient perpendicular to the stress direction (m/s)
_atm	– At Atmospheric Conditions
c1	– Cognitive Acceleration Component
c2	– Social Acceleration Component
g(t)	– Global Best
gbest	– Global Best
k	– Number of Cluster Centres
k_i	– No of Data Points in the i^{th} Cluster
K_p	– Proportionality Constant
L	– Mean Free Path (m)

LBC	– Lohrenz, Bray and Clark
M-	– Modified
MAE	– Mean Absolute Error (cP)
MAPE	– Mean Absolute Percentage Error (%)
MW	– Molecular Weight
N	– Number of Components
n	– No. of Molecules per Unit Volume
p	– Reservoir Pressure (psia)
P_b	– Pressure at Bubble Point (psia)
pbest	– Personal Best
p_i	– Personal Best of Particle *i*
PSO	– Particle Swarm Optimization
R^2	– R squared
RI	– Relative Improvement between modified and original correlation
RBF	– Radial Basis Function
RBFN	– Radial Basis Function Network
R_s	– Solution Gas Oil Ratio (scf/stb)
SG	– Specific Gravity
Standard Conditions	– Temperature=68 °F, Pressure=14.696 psi
STO	– Stock Tank Oil
T	– Time (s)
T	– Temperature (^{0}F)
T_c	– Critical Temperature (^{0}F)
T_p	– Pour Point Temperature (^{0}C)
T_r	– Reduced Temperature (^{0}F)
v	– Average Molecular Speed (m/s)
v_i	– Velocity of Particle *i* (m/s)
w(α)	– Inertia Weight
w	– Intertia Value
W_i	– Weight of Corresponding Regionalized Value, Z_i
x_i	– Mole Fraction of Component *i*
x_i	– Position of Particle *i*
$Z(s_o)$	– Estimated Value of Unsampled Region
Z_i	– Regionalized Variable

References

1. Finnemore, E. J., Franzini, J. B., & Daugherty, R. L. (2002). *Fluid mechanics with engineering applications* (10th ed. ed.). Boston: McGraw-Hill.
2. Pedersen, K. S., Christensen, P. L., & Shaikh, J. A. (2006). *Phase Behavior of Petroleum Reservoir Fluids*. Taylor and Francis Group: CRC Press.
3. Whitson, C. H., and Brule, M.R. (2000). Gas and Oil Properties and Correlations *Phase Behaviour* (Vol. 20, pp. 18-44). Texas: Monograph.
4. Poling, B. E., Prausnitz, J. M., & O'connell, J. P. (2001). *The properties of gases and liquids* (Vol. 5): Mcgraw-hill New York.
5. Pedersen, K. S., Fredenslund, A., & Thomassen, P. (1989). *Properties of oils and natural gases*: Gulf Pub. Co., Book Division.
6. Xiang, H. W. (2005). Theoretical Basis of the Corresponding-State Principle *The Corresponding-States Principle and its Practice* (pp. 17-36). Amsterdam: Elsevier.
7. Birkett, G. (2013). Corresponding States. Retrieved 30/05/2017, 2017, from https://www.youtube.com/watch?v=Cy-gMUYrUhA
8. Hornnes, H. K. (2015). Swelling of Johan Sverdrup oil with rich gas, lean gas and CO2 (pp. 156). Statoil PVT-database: Statoil Research Centre.
9. Measuring Principle. (2017). Retrieved 07/06/2017, 2017, from http://www.viscopedia.com/methods/measuring-principles/
10. Lohrenz, J., Bray, B. G., & Clark, C. R. (1964). Calculating Viscosities of Reservoir Fluids From Their Compositions. doi: 10.2118/915-PA
11. Jossi, J. A. (1962). The viscosity of pure substances in the dense gaseous and liquid phases. *AIChE Journal, 8*, 59-63. doi: 0.1002/aic.690080116
12. Pedersen, K. S., & Fredenslund, A. (1987). An improved corresponding states model for the prediction of oil and gas viscosities and thermal conductivities. *Chemical Engineering Science, 42*(1), 182-186. doi: http://dx.doi.org/10.1016/0009-2509(87)80225-7
13. Ljung, L. (2001). *Black-box Models from Input-output Measurements*. Paper presented at the 18th IEEE Instrumentation and Measurement Technology Conference, Budapest.
14. Arief, I. H., Forest, T., & Meisingset, K. K. (2017). *Estimating Fluid Properties Using Surrogate Models and Fluid Database*. Paper presented at the SPE Bergen One Day Seminar, Bergen, Norway.
15. Al-Mashagbah, A., Al-Adamat, R., & Salameh, E. (2012). The use of Kriging Techniques with in GIS Environment to Investigate Groundwater Quality in the Amman-Zarqa Basin/Jordan. *Journal of Environmental and Earth Sciences, 4*(2), 177-185.
16. Farrokh, N. (2015). Accounting for Uncertainty and Variability in Geotechnical Characterization of Offshore Sites. *Geotechnical Safety and risk*(V), 23-34.
17. Puri, M., Solanki, A., Padawer, T., Tipparaju, S. M., Moreno, W. A., & Pathak, Y. (1991). Neural Networks and Other Information Processing Approaches *Introduction to Neural Networks* (Vol. 2, pp. 7-14). Amsterdam: Elsevier.
18. Shiffman, D. (2012). Neural Networks *The Nature of Code* (illustrated ed.): D. Shiffman.
19. McCormick, C. (2013). Radial Basis Function Network (RBFN) Tutorial. Retrieved 30/03/2017, 2017, from

http://mccormickml.com/2013/08/15/radial-basis-function-network-rbfn-tutorial/
20. Afiatdoust, F., & Esmaeilbeigi, M. (2015). Optimal variable shape parameters using genetic algorithm for radial basis function approximation. *Ain Shams Engineering Journal, 6*(2), 639-647. doi: https://doi.org/10.1016/j.asej.2014.10.019
21. Cui, X. (2016). *Poststack impedance inversion using improved particle swarm optimization*. Paper presented at the 2016 SEG International Exposition and Annual Meeting, Dallas, Texas.
22. Evers, G. (2009). *An automatic regrouping mechanism to deal with stagnation in particle swarm optimization.* (Master of Science), University of Texas-Pan American.
23. Dandekar, A. Y. (2013). *Petroleum Reservoir Rock and Fluid Properties, Second Edition*: Taylor & Francis.
24. Naseri, A., Nikazar, M., & Mousavi Dehghani, S. A. (2005). A correlation approach for prediction of crude oil viscosities. *Journal of Petroleum Science and Engineering, 47*(3–4), 163-174. doi: https://doi.org/10.1016/j.petrol.2005.03.008
25. Labedi, R. (1992). Improved correlations for predicting the viscosity of light crudes. *Journal of Petroleum Science and Engineering, 8*(3), 221-234. doi: http://dx.doi.org/10.1016/0920-4105(92)90035-Y
26. Abu-Khamsin, S. A., & Al-Marhoun, M. A. (1991). Development of a New Correlation for Bubblepoint Oil Viscosity. *Arabian Journal for Science and Engineering, 16*.
27. Beal, C. (1946). The Viscosity of Air, Water, Natural Gas, Crude Oil and Its Associated Gases at Oil Field Temperatures and Pressures. *Transaction of AIME, 165*(01), 22. doi: 10.2118/946094-G
28. Chew, J.-N., & Connally, C. A., Jr. (1959). A Viscosity Correlation for Gas-Saturated Crude Oils (Vol. 216, pp. 23-25). AIME: Society of Petroleum Engineers.
29. Beggs, H. D., & Robinson, J. R. (1975). Estimating the Viscosity of Crude Oil Systems. *Journal of petroleum science & engineering, 27*(09), 2. doi: 10.2118/5434-PA
30. Standing, M. B. (1977). *Volumetric and Phase Behaviour of Oil Field Hyrdrocarbon Systems*. Dallas: SPE of AIME.
31. Glaso, O. (1980). Generalized Pressure-Volume-Temperature Correlations. *Journal of petroleum science & engineering, 32*(05), 11. doi: 10.2118/8016-PA
32. Al-khafaji, A. H., Abdul-Majeed, G.H., and Hasoon, S.F. (1987). Viscosity Correlation for Dead, Live and Undersaturated Crude Oil. *J.Pet.Res., 6*, 1-16.
33. Ng, J. T. H., & Egbogah, E. O. (1983). *An Improved Temperature-Viscosity Correlation For Crude Oil Systems*. Paper presented at the Annual Technical Meeting, Banff, Alberta.
34. Kartoatmodjo, T. R. S., & Schmidt, Z. (1991). New Correlations For Crude Oil Physical Properties (pp. 39): Society of Petroleum Engineers.
35. Petrosky, G. E., Jr., & Farshad, F. F. (1995). *Viscosity Correlations for Gulf of Mexico Crude Oils*. Paper presented at the Production Operations Sympsium, Oklahoma City.

36. Bennison, T. (1998). *Prediction of Heavy Oil Viscosity*. Paper presented at the IBC Heavy Oil Field Development Conference, London.
37. Elsharkwy, A. M., & Gharbi, R. B. C. (2001). Comparing classical and neural regression techniques in modeling crude oil viscosity. *Advances in Engineering Software, 32*(3), 215-224. doi: https://doi.org/10.1016/S0965-9978(00)00083-1
38. Foundation, P. S. (Version Python Language Reference, version 2.7). Retrieved from http://www.python.org
39. Balaprakash, P., Wild, S. M., & Hovland, P. D. (2013). *An Experimental Study of Global and Local Search Algorithms in Empirical Perfomance Tuning*. Paper presented at the 10th International Meeting on High-Performance Computing for Computational Science (VECPAR 2012), Kobe, Japan.
40. Mongillo, M. (2011). Choosing Basis Functions and Shape Parameters for Radial Basis Function Methods. *SIAM*, 209. doi: 10.1137/11S010840
41. Wu, Y., Wang, H., Zhang, B., & Du, K.-L. (2012). Using Radial Basis Function Networks for Function Approximation and Classification. *ISRN Applied Mathematics, 2012*, 34. doi: 10.5402/2012/324194
42. Naik, A. (2010). K-means Clustering Algorithm. Retrieved 14/02/2017, 2017, from https://sites.google.com/site/dataclusteringalgorithms/k-means-clustering-algorithm
43. Feng, C. S., Cong, S., & Feng, X. Y. (2007, 25-28 Sept. 2007). *A new adaptive inertia weight strategy in particle swarm optimization.* Paper presented at the 2007 IEEE Congress on Evolutionary Computation.
44. de Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016). Mean Absolute Percentage Error for regression models. *Neurocomputing, 192*, 38-48. doi: https://doi.org/10.1016/j.neucom.2015.12.114
45. De Ghetto, G., Paone, F., & Villa, M. (1995). *Pressure-Volume-Temperature Correlations for Heavy and Extra Heavy Oils*. Paper presented at the SPE International Heavy Oil Symposium, Calgary, Alberta.
46. Alam, N. (2016). *Codes in Matlab for Particle Swarm Optimization*.doi: 10.13140/RG.2.1.1078.7608.

Appendix A. - Statistical Correlation Data Reported From Literature:

Dead Oil:

Correlation	Year	No. Of data	Region	Temperature [°F]	Gravity [°API]	Viscosity [cp]	AAE* %
Beal	1946	655	US	100-220	15-53	0.9-1550	24.2
Beggs & Robinson	1975	460	n/a	70-295	16-58	n/a	-0.64
Standing	1977	n/a	n/a	n/a	n/a	n/a	n/a
Glaso	1980	26	North Sea	50-300	20-46	50-300	n/a
Al-khafaji et al.	1987	1270	Middle East	60-300	15-51	0.09-7.14	-2.4
Egbogah & Ng	1990	394	n/a	59-144	5-58	N/A	-5.13
Labedi	1992	100	Libya	100-306	32-48	0.66-4.79	-2.61
Bergman	1992	460	n/a	>295	16-58	n/a	n/a
Kartoatmodjo & Schmidt	1994	661	World Wide	75-320	14.4-58.9	0.5-682	-13.16
Petrosky & Farshad	1995	118	Gulf of Mexico	114-288	25.4-46.1	0.725-10.249	-3.38
Bennison	1998	16	North Sea	39-300	11.1-19.7	6.4-8,398	n/a
Elsharkawy	1999	254	Middle East	100-300	19.9-48	0.6-33.7	19.3

*Author Average Percentage Error

Saturated Oil:

Correlation	Year	No. Of Data	Region	GOR [scf/STB]	Temperature [°F]	P_b [psia]	μ_{ob} [cp]	AAE* %
Chew & Connally	1959	457	n/a	51-3544	72-292	132-5645	n/a	n/a
Beggs & Robinson	1975	3143	n/a	20-2070	70-295	n/a	n/a	-1.83
Standing	1977	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Al-Khafaji et al.	1987	1270	Middle East	0-2100	60-300	n/a	n/a	1.80
Labedi	1992	91	Libya	n/a	100-306	60-6358	0.12-3.7	-2.30
Bergman	1992	n/a	n/a	n/a	n/a	n/a	n/a	n/a
Kartoatmodjo and Schmidt	1994	5321	World Wide	2.3-572	80-320	0.5-582	0.096-586.0	0.08
Petrosky & Farshad	1994	864	Gulf of Mexico	21-1885	n/a	n/a	0.21-7.4	-3.12
Elsharkawy	1999	254	Middle East	10-3600	n/a	100-3700	0.05-20.89	18.7

*Average percentage error

Undersaturated Oil:

Correlation	Year	No. Of Data	Oil viscosity, μ_o, [cp]	Oil Viscosity @P_b, μ_{ob}, [cp]	P [psia]	P_b [psia]	APE* %
Beal	1946	26	0.16 – 315	0.142 - 127	n/a	n/a	2.7
Al-Khafaji	1987	210	0.096 – 28.5	n/a	n/a	n/a	0.0578
Labedi	1992	31	n/a	0.098 – 10.9	n/a	715 – 4,794	-3.1
Kartoatmodjo & Schmidt	1994	3,588	0.168 – 517.03	0.168 – 184.86	25 – 6,015	25 – 4,775	-4.29
De Ghetto	1995	195	0.13 – 354.6	n/a	n/a	n/a	n/a
Petrosky and Farshad	1995	404	0.22 – 4.09	0.211 – 3.546	1,600 – 10,250	1,574 – 9,552	-0.19
Elsharkawy	1999	254	Middle East	n/a	n/a	1287-10,000	4.9

*Average percentage error