DEVELOPMENT OF NEW PVT CORRELATIONS

FOR RESERVOIR FLUIDS FROM UNCONVENTIONAL RESERVOIRS

A Thesis

by

NELSON DAVID MEJIA MARTINEZ

Submitted to the Graduate and Professional School of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Thomas A. Blasingame
Committee Members,	Maria Barrufet
	Zachary Gagnon
Head of Department,	Jeff Spath

August 2022

Major Subject: Petroleum Engineering

ABSTRACT

The estimation of Pressure-Volume-Temperature (or PVT) properties is an extremely important aspect of oil and gas exploration and production as the parameters derived from these functions are used for reserve estimations, reservoir modeling, production and pressure analysis, and for predicting well production performance. Reservoir fluids from unconventional reservoirs tend to be significantly "lighter" (*i.e.,* contain less heavy hydrocarbon molecules) compared to the so-called "black oil" reservoir fluids, and as such, generally require specialized correlations.

The primary focus of this work is to estimate the "saturation pressure" and "oil formation volume factor" of a relatively large sampling of reservoir fluids from unconventional (*shale*) *reservoirs* where these fluids tend to be *near-critical volatile oils or retrograde gas condensate fluids*. In this work, we utilize data from 138 PVT lab reports donated by industry collaborators (on the condition that the reports and data remain anonymous/confidential). These data were obtained from various unconventional reservoirs in the United States and Latin America (where again, the origin of the data is to remain confidential).

These data are used as a test case against the most common saturation pressure correlations available in the industry since the 1940s. It is important to recognize that this is a common practice, *i.e.*, to test a "standard' model against a new data set. However; virtually all of the historical saturation pressure correlations were developed for "black oil" reservoir fluids. As such, we test each historical model with and without regression (MS Excel and Python algorithms are used to perform the regressions).

To assess a "best fit" in a statistical sense, we also utilize so-called "non-parametric" correlation methods that provide a multivariate optimization *without using a specific model*. These methods

provide guidance on the best correlation of a given multivariate data set, but these methods cannot be used as a predictor (at least not directly). Lastly, in this work, we have developed a family of exponential-polynomial and rational polynomial functions to relate the saturation pressure and oil formation volume factor with reservoir temperature, stock tank oil gravity, separator gas gravity, and solution gas-oil-ratio.

Our proposed correlations have yielded excellent functional and statistical performances for our specific database, and we believe these correlations provide a new methodology for representing saturation pressure and oil formation volume factor based on these inputs. We also note that we have used certain variations of our proposed model to predict oil viscosity, but this was more of a test-of-concept as we are not confident that the volume of oil viscosity data in our database is sufficient to provide unique and robust correlations.

DEDICATION

To my parents Nelson and Myriam who have always supported my educational and professional endeavors, and who always helped me strive to achieve new heights in my professional career. You gave me the motivation and guidance to complete this work

To my sisters Natalia and Valeria who have supported me with love and inspiration while completing this work.

ACKNOWLEDGMENTS

The author expresses his gratitude towards the following individuals for their assistance in the development of this thesis:

Dr. Thomas A. Blasingame for his constant encouragement, patience, and humor. His mentorship has been invaluable.

Dr. Maria Barrufet for her service as a member of the advisory committee.

Dr. Zachary Gagnon for his service as a member of the advisory committee.

My friends and colleagues who I've had the pleasure of meeting and working with throughout my graduate studies.

AJ Smith for her creativity, patience, and work ethic in support of my graduate career.

My family for their support throughout my educational and professional careers.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a thesis committee consisting of Dr. Thomas Blasingame and Dr. Maria Barrufet of the Department of Petroleum Engineering and Dr. Zachary Gagnon of the Department of Chemical Engineering.

Funding Sources and Responsibilities

This work was self-funded by the student. The contents of this thesis are solely the responsibility of the authors and do not necessarily represent the official views of the Department of Petroleum Engineering at Texas A&M.

NOMENCLATURE

Field Variables

API = Oil stock tank gravity, [°API]

B_o = Oil formation volume factor, [RB/STB]

B_{ob} = Oil formation volume factor at saturation pressure, [RB/STB]

P_{sat} = Saturation pressure, [Pa] or [psia]

R_s = Solution gas-oil ratio, [scf/STB]

R_{sb} = Solution gas-oil ratio at saturation pressure, [scf/STB]

T = Temperature, [°C] or [°F]

T_r = Reservoir temperature, [°C] or [°F]

Greek Variables

γ_{API} = Oil API Gravity, [°API]

γ_g = Gas specific gravity, [dimensionless] (air =1)

γ_o = Oil specific gravity, [dimensionless] (water =1)

μ = Viscosity, [Pa s] or [cP]

μ_o = Oil Viscosity, [Pa s] or [cP]

Correlation Variables

c_i = Correlation coefficients (i = 1, 2, 3, …).

C_i = Correlation coefficients (i = 1, 2, 3, …).

Statistical Variables

σ	=	Standard Deviation [psia], [RB/BBL], or [cp]
σ^2	=	Variance [$psia^2$], [RB/BBL^2], or [cp^2]
ADE	=	Absolute Difference Error [psia], [RB/BBL], or [cp]
LSE	=	Least-Squares Error [$psia^2$], [RB/BBL^2], or [cp^2]
AARE	=	Average Absolute Relative Error (percent)

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

LIST OF FIGURES

LIST OF FIGURES (Continued)

FIGURE Page

LIST OF FIGURES (Continued)

FIGURE Page

LIST OF TABLES

LIST OF TABLES(Continued)

TABLE Page

CHAPTER I

INTRODUCTION

This chapter provides an overview of the development and application of PVT correlations for reservoir fluids from unconventional reservoirs as developed in this thesis. The goal of this work is to establish practical and robust correlations for estimating the saturation pressure and the oil formation volume factor for unconventional reservoir fluids using commonly available reference properties (R_{sb}, γ_g (or SG), API, T, p_{sat} [where p_{sat} is only used for the B_{ob} correlations]).

1.1 Statement of the Problem

This thesis aims to determine the validity of historical correlations and develop a new set of equations to predict saturation pressure and oil formation volume factor *for cases of black oil, volatile oil, and retrograde gas condensate reservoir fluids from unconventional reservoirs*. Pressure-Volume-Temperature (or PVT) properties are extremely important to different facets of the oil and gas industry, from reserve estimation, reservoir modeling, production and pressure evaluation, and well performance.

As suggested above, almost all PVT correlations have been proposed for conventional "black oil" fluids; however, our work will focus on the application of these existing correlations or cases of reservoir fluids in unconventional reservoirs. In addition to "recalibrating" the existing "black oil" correlations to a database that we created from company data donations, we will also create new correlations for saturation pressure (p_{sat} for black/volatile oils, or p_{sat} for retrograde gas condensate reservoirs) and the oil formation volume factor at the bubblepoint (B_{ob}).

The increase of exploration in unconventional shale resources has accelerated the need for the accurate determination/estimation of PVT properties. Laboratory testing is the most recommended and most accurate approach for determining these properties; however, if obtaining a fluid sample is technically infeasible due to the well condition(s), and/or cost-prohibitive, then correlations must be used for the estimation of PVT properties. Since unconventional fluids tend to be significantly "lighter" (less heavy hydrocarbon molecules) compared to so-called conventional "black oil" reservoir fluids, then we should expect to see an erroneous estimation of the reservoir fluid properties when using historical correlations (as these correlations were tuned to conventional reservoir fluids).

These so-called "lighter" reservoir fluids from unconventional reservoirs are typically characterized by high stock tank oil gravity values, (>40 °API) as well as high values of solution gas-oil ratio (>1000 scf/STB), and often yield substantially different PVT behavior than a reservoir fluid from a conventional reservoir. It is not the purpose of this study to attempt to address thermodynamic causes for the behaviors of unconventional reservoir fluids, but rather to assess the validity and application of historical correlations, and to propose new correlations as necessary.

The historical correlations (for conventional reservoirs) that are investigated in this study were proposed by the following authors:

- (1947) Standing
- (1958) Lasater
- (1980) Vasquez and Beggs
- (1988) Al-Marhoun
- (1980) Glaso
- (1994) Kartoatmodjo and Schmid
- (1997) Velarde, Blasingame, and McCain
- (1998) Petrosky and Farshad
- (2004) Dindoruk and Christman

1.2 Objectives of This Research

The overall objective of this study is to provide accurate and robust correlation(s) to predict saturation pressure and oil formation volume factor at saturation pressure for reservoir fluids from unconventional reservoirs using common inputs: reservoir temperature, stock tank oil gravity, separator gas gravity, solution gas-oil-ratio (and saturation pressure for the oil formation volume correlation(s)).

The specific objectives of this study are as follows:

1. To collect high-quality PVT data from different unconventional basins — where contributions will be sought from companies (and public agencies if possible).
2. To review each data case to assess relevance and suitability — for reference, approximately 10-15 percent of the data were not suitable for correlation purposes.
3. To apply existing correlation models to this database to understand the limitations of these correlations relative to reservoir fluids from unconventional (shale) reservoirs.
4. To develop (empirical) correlations that will accurately predict saturation pressure and oil formation volume factor for reservoir fluids from unconventional (shale) reservoirs — specifically "near-critical" volatile oils and retrograde gas condensate cases.

1.3 Basic Concepts

This section briefly describes basic concepts referenced throughout this work in an effort to provide a basis for understanding the theory behind the objectives of this thesis.

Unconventional Reservoirs

Unconventional (shale) reservoirs are hydrocarbon-rich plays with typically ultra-low permeability and low to very low porosity. To produce the oil and gas trapped within these types of rock, hydraulic fracturing must be used to permit large volumes of liquids and gases to flow from the rock matrix through the created hydraulic fractures (and augmented natural fractures) into the wellbore. The liquids from these reservoirs tend to be "light" in nature and as described previously, and have high values of stock tank oil gravity and solution gas-oil ratio.

Pressure-Volume-Temperature Correlations

Pressure-Volume-Temperature (or PVT) correlations are sets of equations used to predict oil properties with typically 5 inputs. These inputs include:

Saturation pressure (p_{sat}) (for the oil formation volume factor correlations),
Reservoir temperature (T),
Stock tank oil gravity (API),
Specific gas gravity (γ_g or SG), and
Solution gas-oil ratio (R_s).

The development of correlations typically involves the development of a statistically relevant database from laboratory reports or public literature, and then creating (empirical) equations using graphical or non-linear regression methods. We note that have created a database of 138 unique data points acquired from company donations and a few publicly available datasets.

CHAPTER II

LITERATURE REVIEW

The primary objective of this chapter is to present the most commonly used PVT correlations for saturation pressure, p_{sat}, and the oil formation volume factor at saturation pressure, B_{ob}, where all of these correlations were developed for "black oils" (R_{sb} < 1000 scf/STB) from "conventional reservoirs" (as opposed to unconventional (shale) reservoirs). The comparison and "re-calibration" of these correlations to our PVT database (138 points) for fluids from unconventional reservoirs are presented in Chapter IV.

Recalling from Chapter I, our inventory of existing PVT correlations includes:

(1947)	Standing	—	Original correlations for p_{sat} and B_{ob}, generally most simple forms (105 data points).
(1958)	Lasater	—	p_{sat} only, used a completely different correlation form than Standing (158 data points).
(1980)	Vasquez and Beggs	—	Correlations for p_{sat} and B_{ob}, independently cor-related for $\leq$ or > 30 API (600 data points).
(1988)	Al-Marhoun	—	Power-law relation for p_{sat}, B_{ob} correlation uses a polynomial expansion of a power-law variable (160 Middle East data sets)
(1980)	Glaso	—	Polynomial expansion of Standing's type of correlating variables (26 North Sea data sets).
(1994)	Kartoatmodjo and Schmidt	—	Power-law relation for p_{sat} (uses separate results for $\leq$ or > 30 API) and Standing-type relation for B_{ob} (training set = 5,396 data points, validation set = 998 data points).

(1997) Velarde, Blasingame, and McCain — p_{sat} and B_{ob} correlation equations are a variation of Petrosky and Farshad's formulations (728 data points).

(1998) Petrosky and Farshad — p_{sat} and B_{ob} correlation equations are a variation of Standing forms (81 data sets).

(2004) Dindoruk and Christman — p_{sat} relation is a complex variation of Standing's correlation, and B_{ob} relation is a complex variation of Al-Marhoun's correlation (100 Gulf of Mexico data sets).

As comment, the reference date for the Petrosky and Farshad work (1998) is the date of the journal publication, the conference paper was published in 1993 and the original MS thesis by Petrosky was published in 1990. This is mentioned for clarity in that the Petrosky and Farshad work pre-dates the work of Kartoatmodjo and Schmidt (1994) and Velarde, et al (1997).

As a final note, this study does not consider correlating solution gas-oil-ratios (R_{sb}) as did several researchers who used "Standing"-type formulations, where a given relation for the saturation pressure, p_{sat}, could be "reverse-solved" for the solution gas-oil-ratio at the saturation pressure, R_{sb}.

2.1 Standing Correlations (1947)

Standing's correlations were first presented in 1947 and are considered one of the first comprehensive set of PVT correlations to be published. Standing compiled and used a database of 105 experimentally determined saturation pressures for various California crude oils and gases. Perhaps most impressively, Standing used a graphical method to develop the relationships between the different variables, correlating each variable (or group of variables) *graphically* into a common parameter, as described in the original article (Standing 1947).

Saturation Pressure:

$$p_{sat} = f(R_s, \gamma_g, API, T) \text{ (general statement of correlation variables)} \quad \text{..............................(2.1)}$$

$$p_{sat} = 18.2\left[\left[\frac{R_s}{\gamma_g}\right]^{0.83} 10^{A} - 1.4\right] \quad \text{..............................(2.2)}$$

where the "*A*" parameter is given as a function of temperature, *T*, and stock tank oil gravity, *API*.

$$A = 0.00091\, T - 0.0125\, API \quad \text{..............................(2.3)}$$

Oil Formation Volume Factor:

$$B_{ob} = 0.972 + 1.47 \times 10^{-4}\left[R_s\left[\frac{\gamma_g}{\gamma_o}\right]^{0.5} + 1.25T\right]^{1.175} \quad \text{..............................(2.4)}$$

Standing reported an average relative error of 4.8 percent for p_{sat} (Eq. 2.2) and 1.17 percent for B_{ob} (Eq. 2.4) which is remarkable given the construction of the correlations. The ranges of the variables in Standing's database used to develop these correlations are summarized as follows:

130	<	p_{sat}	<	7000	psia
20	<	R_{sb}	<	1425	scf/STB
100	<	T	<	258	°F
16.5	<	API	<	63.8	°API
0.590	<	γ_g	<	0.95	(air = 1)

2.2 Petrosky and Farshad Correlations (1998)

As noted above, the formal (journal) publication of Petrosky and Farshad's correlations were published in 1998 (prior work was in 1993 and 1990). A database of 81 laboratory PVT tests obtained from the Gulf of Mexico was used to develop these correlation models. A total of 32 different variations of the correlation formulation were assessed to determine the most accurate prediction of saturation pressure, p_{sat}. The oil formation volume factor, B_{ob}, as well as the saturation pressure, p_{sat}, correlation relations have similar forms to those proposed by Standing. Additional fitting parameters were added to increase the accuracy of correlations (Petrosky and Farshad 1998). Nonlinear regression methods were used to yield the following correlations:

Saturation Pressure:

$$p_{sat} = 112.727\left[\frac{R_{sb}^{0.5774}}{\gamma_g^{0.8439}}10^{A} - 12.340\right] \quad (2.5)$$

where the "*A*" parameter is given as a function of temperature, *T*, and stock tank oil gravity, *API*.

$$A = 4.561\times10^{-5}\ T^{4.3911} - 7.916\times10^{-4}\ API^{1.5410} \quad (2.6)$$

Oil Formation Volume Factor:

$$B_{ob} = 1.0113 + 7.2046\times10^{-5}\left[\frac{R_{sb}^{0.3738}\gamma_g^{0.2914}}{\gamma_o^{0.6265}} + 0.24626\,T^{0.5371}\right]^{3.0936} \quad (2.7)$$

Petrosky and Farshad reported an average relative error of -0.17 percent for p_{sat} (Eq. 2.5) and -0.01 percent for B_{ob} (Eq. 2.7) (note that these are <u>not</u> *absolute* relative errors, as will be used in the original portion of this work). The ranges of the variables in the database used by Petrosky and Farshad to develop these correlations are summarized as follows:

1574	<	p_{sat}	<	6523	psia
217	<	R_{sb}	<	1406	scf/STB
114	<	T	<	288	°F
16.3	<	API	<	45.0	°API
0.5781	<	γ_g	<	0.852	(air = 1)

2.3 Lasater Correlations (1958)

While most correlations follow Standing's relation for the saturation pressure (p_{sat}), Lasater took a different approach to the combination of the parameters and used Henry's law as the mathematical basis for his formulation. Lasater also chose to incorporate molecular weight and gas mole fraction derived from stock tank oil gravity and gas specific gravity into the saturation pressure correlation (Eqs. 2.11 and 2.10, respectively). Lasater used a database of 158 data points from laboratory-derived PVT reports to develop the saturation pressure correlation where these data were collected from various locations in the western hemisphere including Canada, the United States, and South America (Lasater 1958).

Saturation Pressure:

$$p_{sat} = \frac{(T + 459.67)}{y_g} p_f \quad \text{.....(2.8)}$$

Where:

$$p_f = 0.3841 - 1.2008\,\gamma_g + 9.648\,\gamma_g^{\,2} \quad \text{.....(2.9)}$$

$$y_g = \frac{\dfrac{R_{sb}}{379.5}}{\dfrac{R_{sb}}{350} + \dfrac{350}{M_o}} \quad \text{.....(2.10)}$$

$$M_o = 725.321 + 16.0333\,API + 0.09524\,API^2 \quad \text{.....(2.11)}$$

These correlations yielded an average relative error of 3.8 percent for Eq. 2.8 (Lasater did not propose an oil formation volume factor nor an oil viscosity correlation). The ranges of the variables in the database used by Lasater to develop this correlation are summarized as follows:

48	<	p_{sat}	<	5780	psia
3	<	R_{sb}	<	2905	scf/STB

$$
\begin{array}{rcccl}
81 & < & T & < & 272 \quad °F \\
17.9 & < & API & < & 51.1 \quad °API \\
0.574 & < & \gamma_g & < & 1.223 \quad (\text{air} = 1)
\end{array}
$$

2.4 Vasquez and Beggs Correlations (1980)

Vasquez and Beggs (1980) utilized a database of 600 laboratory PVT reports collected from fields all over the world. Vasquez and Beggs sought to have a wide range in fluid properties, which led to the *necessity of requiring* 2 *different sets of correlating coefficients* (for conditions $\leq$ or > 30 ^{o}API) for both the saturation pressure and oil formation volume factor correlation relations.

Saturation Pressure:

$$p_{sat} = \left[C_1 \left[\frac{R_{sb}}{\gamma_g} \right] 10^{\frac{C_2 API}{T+459.67}} \right]^{\frac{1}{C_3}} \quad \text{.....(2.12)}$$

Coefficient	$API \leq 30$	$API > 30$
C_1	27.64	56.060
C_2	1.0937	1.187
C_3	11.172	10.393

Oil Formation Volume Factor:

$$B_{ob} = 1 + C_1 R_{sb} + (T-60) \left[\frac{API}{\gamma_g} \right] (C_2 + C_3 R_{sb}) \quad \text{.....(2.13)}$$

Coefficient	$API \leq 30$	$API > 30$
C_1	4.677×10^{-4}	4.670×10^{-4}
C_2	1.751×10^{-5}	1.100×10^{-5}
C_3	-1.811×10^{-8}	1.337×10^{-9}

Vasquez and Beggs reported an average relative error of -0.7 percent for p_{sat} (Eq. 2.12) and 4.7 percent for B_{ob} (Eq. 2.13) (Vasquez and Beggs did not propose a correlation for the oil formation volume factor).

The ranges of the variables in the database used by Vasquez and Beggs to develop this correlation are summarized as follows:

15	<	p_{sat}	<	6055	psia
0	<	R_{sb}	<	2199	scf/STB
70	<	T	<	295	°F
15.3	<	API	<	59.5	°API
0.511	<	γ_g	<	1.351	(air = 1)

2.5 Glaso Correlations (1980)

Glaso (1980) used a database for 26 different crudes obtained from across the North Sea. Glaso used a combination of graphical methods and regression analysis to develop this correlation — specifically, by plotting his working correlation on a log-log plot, a parabolic curve was observed which resulted in the forms given by Eqs.2.15 and 2.17. Glaso used the basic formulations given by Standing (1947), with the addition of the quadratic correlation terms in Eqs.2.15 and 2.17.

Saturation Pressure:

$$p_{sat} = 1.7669 + 1.7447 \ \log p_{sat}{}^{*} - 0.30218 \left[\log p_{sat}{}^{*}\right]^{2} \text{ ....(2.14)}$$

$$p_{sat}{}^{*} = \left[\frac{R_s}{\gamma_g}\right]^{0.816} \frac{T^{0.172}}{API^{0.989}} \text{ ....(2.15)}$$

Oil Formation Volume Factor:

$$B_{ob} = 1 + 10^{\left[-6.58511 + 2.91329 \log B_{ob}{}^{*} - 0.27683\left[\log B_{ob}{}^{*}\right]^{2}\right]} \text{ ....(2.17)}$$

$$B_{ob}{}^{*} = R_{sb}\left[\frac{\gamma_g}{\gamma_o}\right]^{0.526} + 0.986T \text{ ....(2.16)}$$

Glaso also presented additional methods to predict saturation pressure with the presence of various impurities. Glaso reported an average relative error of 1.28 percent for p_{sat} (Eq. 2.15) and -0.43 percent for B_{ob} (Eq. 2.17). The ranges of the variables in the database used by Glaso to develop these correlations are summarized as follows:

165	<	p_{sat}	<	7142	psia
90	<	R_{sb}	<	2637	scf/STB
80	<	T	<	280	°F
22.3	<	API	<	48.1	°API
0.65	<	γ_g	<	1.273	(air = 1)

2.6 Al-Marhoun Correlations (1988)

Al-Marhoun (1988) obtained 69 bottomhole fluid samples from Middle Eastern crude oils where a total of 160 PVT reports were available. Al-Marhoun's correlations were developed using nonlinear multiple regression analysis and a trial-and-error methodology (Al-Marhoun 1988).

Saturation Pressure:

$$p_{sat} = 5.38088 \times 10^{-3} \; R_{sb}{}^{0.715082} \; \gamma_g{}^{-1.87784} \; \gamma_o{}^{3.1437} \; T^{1.32657} \quad \text{.....................................(2.18)}$$

where the oil-specific gravity is used in place of the stock tank oil gravity in ^{o}API.

$$\gamma_o = \frac{141.5}{131.5 + API} \quad \text{...(2.19)}$$

Oil Formation Volume Factor:

$$B_{ob} = 0.4971 + 8.630 \times 10^{-4} \; (T + 459.67) + 1.826 \times 10^{-3} X + 3.181 \times 10^{-6} X^2 \quad \text{....................(2.20)}$$

Where:

$$X = R_{sb}{}^{0.742390} \gamma_g{}^{0.323294} + \gamma_o{}^{-1.202040} \quad \text{..(2.21)}$$

Al-Marhoun reported an average relative error of 0.03 percent for p_{sat} (Eq. 2.18) and -0.01 percent for B_{ob} (Eq. 2.20). As comment, Al-Marhoun did not propose an oil viscosity correlation in this work. The ranges of the variables in the database used by Al-Marhoun to develop these correlations are summarized as follows:

130	<	p_{sat}	<	3573	psia
26	<	R_{sb}	<	1602	scf/STB
76	<	T	<	240	°F
19.4	<	API	<	44.6	°API
0.752	<	γ_g	<	1.367	(air = 1)

2.7 Kartoatmodjo and Schmidt Correlations (1994)

Kartoatmodjo and Schmidt (1994) used two different databases for their work, one database was used for "training" (Databank A) and another database was used for "validation" (Databank B). Databank A has 5,396 data points taken from 740 different crudes globally and Databank B derives its 998 data points from various published literature. Analogous to Vasquez and Beggs (1980), Kartoatmodjo and Schmidt used the < or > 30 °API criteria to have 2 correlations, in this case only for the saturation pressure, p_{sat}, as given below by Eqs. 2.22 and 2.22.

Saturation Pressure:

$$p_{sat} = \left[\frac{R_{sb}}{0.05958\ \gamma_g^{\ 0.7972}\ 10^{13.1405}\ API^{T+459.67}} \right]^{0.9986} \qquad (API \le 30) \ldots\ldots\ldots(2.22)$$

$$p_{sat} = \left[\frac{R_{sb}}{0.03150\ \gamma_g^{\ 0.7587}\ 10^{11.28955}\ API^{T+459.67}} \right]^{0.9143} \qquad (API > 30) \ldots\ldots\ldots(2.23)$$

Oil Formation Volume Factor:

$$B_{ob} = 0.98496 + 1\times10^{-4}\left[R_{sb}^{0.755}\left[\frac{\gamma_g^{0.25}}{\gamma_o^{1.5}}\right] + 0.45T\right]^{1.5} \quad \text{...(2.24)}$$

Kartoatmodjo and Schmidt reported an average relative error of 3.34 percent for p_{sat} (for *both* Eqs. 2.22 and 2.23, perhaps as an average) and -0.104 percent for B_{ob} (Eq. 2.24). The ranges of the variables in the database used by Kartoatmodjo and Schmidt to develop these correlations are summarized as follows:

15	<	p_{sat}	<	6054	psia
14	<	R_{sb}	<	2473	scf/STB
75	<	T	<	320	°F
14.4	<	API	<	58.9	°API
0.37	<	γ_g	<	1.71	(air = 1)

2.8 Velarde, Blasingame, and McCain Correlations (1997)

Velarde (1996) and Velarde et al (1997) used a total of 728 data points at saturation pressure to develop their correlations — 478 of these data points were collected from commercial laboratories, 160 points were collected from the Al-Marhoun (1988) study, and 90 samples were extracted from the Petrosky and Farshad (1998) study.

Velarde et al used the p_{sat} model proposed by Petrosky and Farshad (1998) as the basis of their new correlation, adding a new parameter (exponent) to increase the accuracy of their correlation. Velarde et al used a generalized form of the Standing (1947) relation for B_o.

Saturation Pressure:

$$p_{sat} = 1091.47 \left[\frac{R_{sb}^{0.081465}}{\gamma_g^{0.161488}} 10^{A} - 0.740152 \right]^{5.354891} \text{...(2.25)}$$

Where:

$$A = 0.013098\, T^{0.282372} - 8.2 \times 10^{-6}\, API^{2.176124} \text{..(2.26)}$$

Oil Formation Volume Factor:

$$B_{ob} = 1.023761 + 0.000122 \left[R_{sb}^{0.413179} \gamma_g^{0.210293} \gamma_o^{0.127123} + 0.019073T \right]^{2.465976} \text{......(2.27)}$$

Velarde et al reported an average absolute relative error of 11.7 percent for p_{sat} (Eq. 2.25) and -0.85 percent for B_{ob} (Eq. 2.27). The ranges of the variables in the database used by Velarde et al to develop these correlations are summarized as follows:

70	<	p_{sat}	<	6700	psia
10	<	R_{sb}	<	1870	scf/STB
74	<	T	<	327	°F
12	<	API	<	55.0	°API
0.556	<	γ_g	<	1.367	(air = 1)

2.9 Dindoruk and Christman Correlations

Dindoruk and Christman (2001) used a total of 100 laboratory reports derived from (US) Gulf of Mexico crude oil samples to develop their correlations. Dindoruk and Christman employed a nonlinear regression solver hosted by Microsoft Excel.

Saturation Pressure:

$$p_{sat} = 1.86998 \left[\frac{R_{sb}^{1.22145}}{\gamma_g^{1.37051}} 10^{A} + 0.011688 \right] \text{...(2.28)}$$

where the "A" parameter in Eq. 2.28 is a much more complex functional form compared to previous cases (Standing (1947), Petrosky and Farshad (1998), and Velarde et al (1997)) and is given as a function of *all* variables — reservoir temperature, stock tank oil gravity, separator gas gravity, and solution gas-oil-ratio.

$$A = \frac{1.43\times10^{-10} T^{2.84459} - 6.74\times10^{-4} API^{1.22523}}{\left[0.033833 + \dfrac{2R_{sb}^{-0.272946}}{\gamma_g^{-0.0842261}}\right]} \quad \text{...(2.29)}$$

Oil Formation Volume Factor:

$$B_{ob} = 0.98718 + 7.865\times10^{-4} A + 2.689\times10^{-6} A^2 + 1.10\times10^{-5}(T-60)\frac{API}{\gamma_g} \quad \text{...................(2.30)}$$

where the "A" parameter in Eq. 2.30 is a much more complex variation of the proposed B_{ob} correlations given by Glaso (1980) and Al-Marhoun (1988).

$$A = \frac{\left[\dfrac{R_{sb}^{2.5108}\gamma_g^{-4.8525}}{\gamma_o^{11.835}} + 1.365\times10^5 (T-60)^{2.25288} + 10.0719 R_{sb}\right]^{0.44508}}{\left[5.3526 + \dfrac{2R_{sb}^{-0.630905}}{\gamma_g^{0.9001}}(T-60)\right]^2} \quad \text{...................(2.31)}$$

Dindoruk and Christman reported an average relative error of 0.27 percent for p_{sat} (Eq. 2.28) and -5.0 percent for B_{ob} (Eq. 2.30). The ranges of the variables in the database used by Dindoruk and Christman to develop these correlations are summarized as follows:

926	<	p_{sat}	<	12230	psia
133	<	R_{sb}	<	3050	scf/STB
117	<	T	<	276	°F
14.7	<	API	<	40.0	°API
0.602	<	γ_g	<	1.027	(air = 1)

CHAPTER III

DATA DESCRIPTION

The collection of the data used in this study underwent a vigorous solicitation campaign targeting private companies and searching in public databases. The goal was to compile PVT reports from unconventional reservoirs in various global geographic areas so that the correlations proposed in this work will have a large range of applicability. In full disclosure, industry reluctance to share data (even anonymously) and internal organizational regulations were just some of the challenges that were overcome to compile the database used in this study.

Per agreements made with industry partners, *all data provided to support this work is to remain anonymous* (but can be shared with the public, and the full database for this work is presented in Appendix A). Further, *all contributors are to remain confidential*. Lastly, the vast majority of these data (approximately 90 percent) were obtained from industry donors, a minority portion of the data were obtained from public resources (state entities, public reports, technical articles, etc.).

Laboratory testing of reservoir fluids (so called "PVT sampling and testing") is standard practice throughout the oil and gas industry. These fluid samples are then used in various laboratory tests; including compositional analysis, differential liberation tests, flash liberation tests, separator tests, and viscosity measurements. These tests report various physical parameters and the molecular composition of a given reservoir fluid. It is standard practice to compile the results of these tests into a comprehensive report, most often referred to as a "PVT" or "Reservoir Fluids" report.

3.1 Data Preparation

The data that were collected for this study originate from Latin America and the United States and represent a spectrum of unconventional reservoirs within those geographical areas. For the database used in this study, the following data were extracted from donated/acquired PVT reports:

p_{sat}	= Saturation pressure (psia)	
T	= Reservoir temperature (^oF)	
R_{sb}	= Solution gas-oil-ratio at p_{sat} (scf/STB)	
API	= Stock tank oil gravity (^{o}API)	
γ_o	= Stock tank oil specific gravity (g/cc)	
γ_g	= Separator gas gravity (air=1)	
B_{ob}	= Oil formation volume factor at p_{sat} (RB/STB)	(not universally available)
μ_{ob}	= Oil viscosity at p_{sat} (cp)	(not universally available)

For a given lab report, each "data point" (saturation pressure, reservoir temperature, solution gas-oil-ratio (at p_{sat}), stock tank oil gravity, separator gas gravity, oil formation volume factor (at p_{sat}) [if available], and oil viscosity (at p_{sat}) [if available]) was selected and cross-checked against other data. Cases with obvious errors and/or systematic deviations were excluded from the final data-base (more than 20 submissions were discarded).

As practical comment, the oil formation volume factor and the oil viscosity were only reported in about 1/3 of the cases, so the data for these two properties are limited. Lastly, some of the donated data cases were given in metric units, which were converted to field units (and double-checked).

3.2 Data Description

A total of 138 PVT laboratory reports were used for the saturation pressure, p_{sat}, correlations reviewed and developed in this work. From this "master" database (of 138 "data points"), only 46 data points were available to develop correlations for the oil formation volume factor, B_{ob}. The complete database used in this work is presented in **Appendix A**.

The initial review of the database is performed by creating crossplots of solution gas-oil-ratio, stock tank oil gravity, specific gas gravity, and temperature, respectively (these variables are on the *y*-axis) versus saturation pressure (on the *x*-axis). These goal of using these crossplots is to visually assess any "simple" relationship between each variable and the saturation pressure. These crossplots are shown in **Figs 3.1-3.4**.

The ranges of data that were used to develop the p_{sat} correlations (138 data points) are as follows:

$$
\begin{array}{rcccll}
1211 & < & p_{sat} & < & 10000 & \text{psia} \\
275 & < & R_{sb} & < & 85802 & \text{scf/STB} \\
126 & < & T & < & 330 & {}^\circ\text{F} \\
30 & < & API & < & 88.98 & {}^\circ\text{API} \\
0.437 & < & \gamma_g & < & 1.3 & (\text{air} = 1)
\end{array}
$$

The ranges of data that were used to develop the B_o correlation (46 data points) are as follows:

$$
\begin{array}{rcccll}
1389 & < & p_{sat} & < & 5510 & \text{psia} \\
482.88 & < & R_{sb} & < & 4925 & \text{scf/STB} \\
126 & < & T & < & 270 & {}^\circ\text{F} \\
34.2 & < & API & < & 58.1 & {}^\circ\text{API} \\
0.694 & < & \gamma_g & < & 1.3 & (\text{air} = 1) \\
1.349 & < & B_o & < & 3.838 & \text{RB/STB} \\
0.062 & < & \mu_o & < & 0.935 & \text{cp}
\end{array}
$$

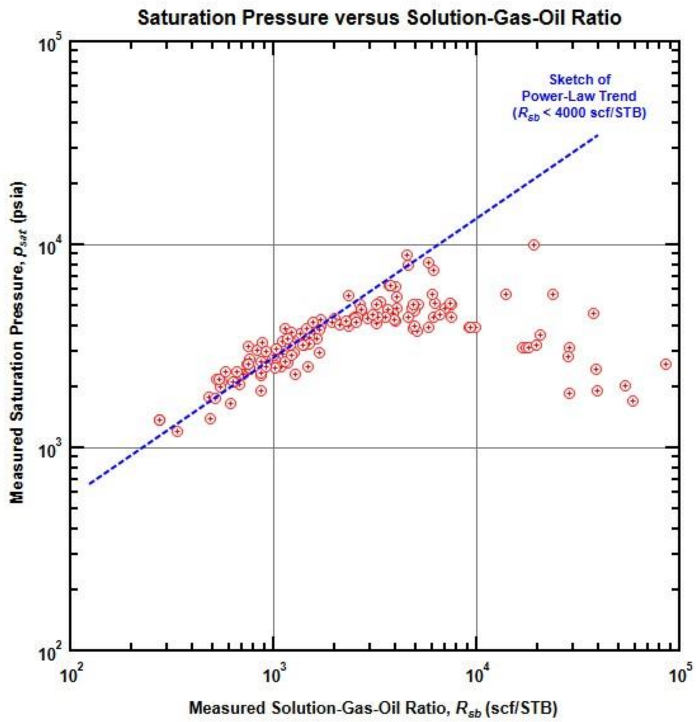

Figure 3.1 — Log-log plot for saturation pressure (p_{sat}) versus the solution-gas-oil ratio (R_s) for the unconventional reservoir fluid database (138 data points).

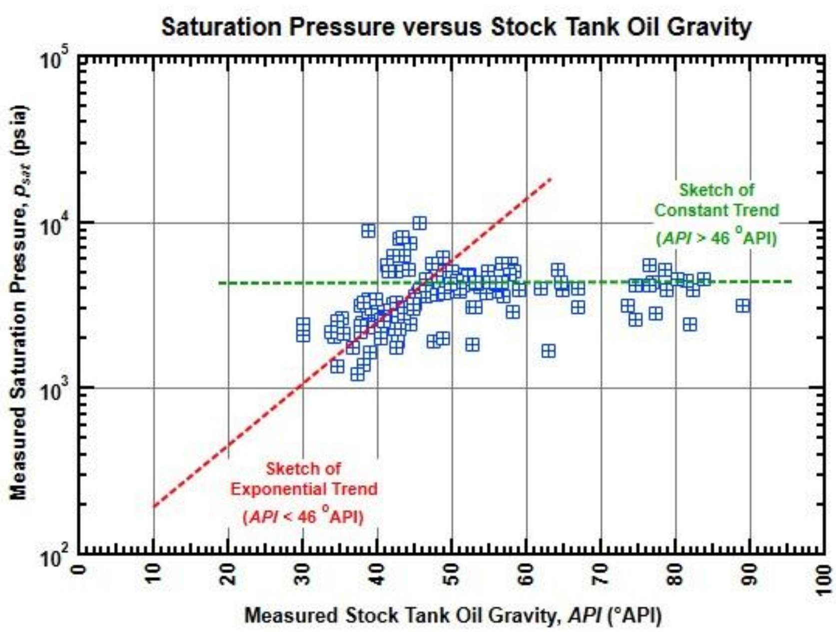

Figure 3.2 — Semi-log plot for saturation pressure (p_{sat}) versus the stock tank oil gravity (API) for the unconventional reservoir fluid database (138 data points).

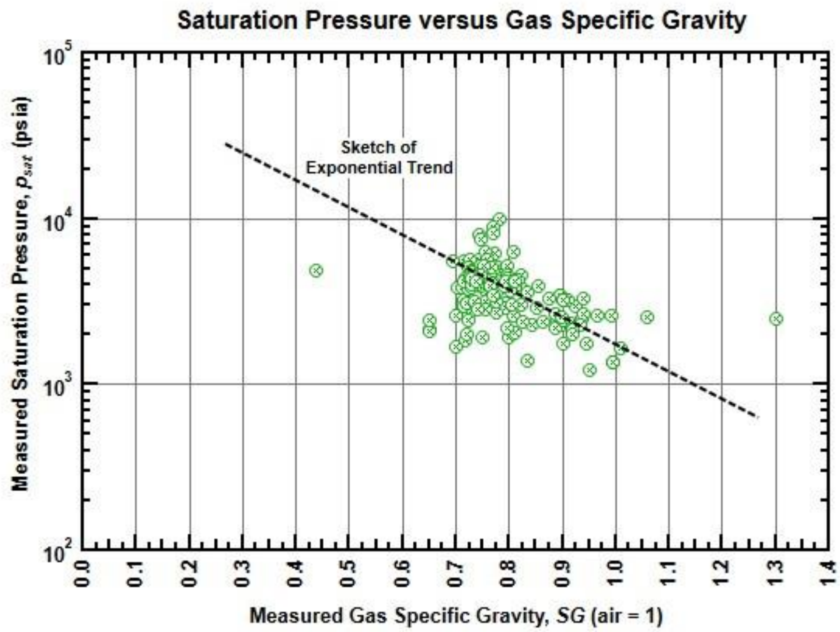

Figure 3.3 — Semi-log plot for saturation pressure (p_{sat}) versus the separator gas gravity (SG) for the unconventional reservoir fluid database (138 data points).

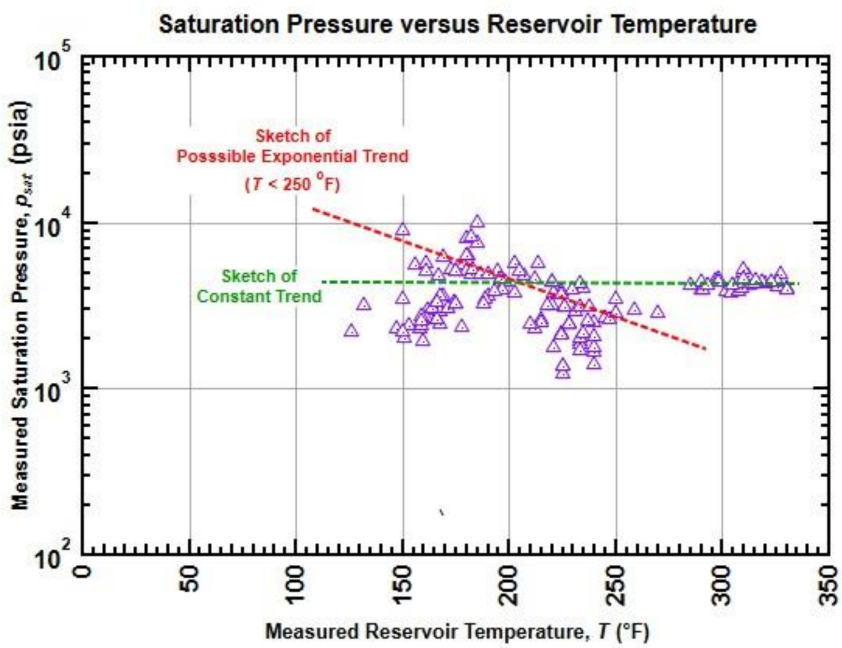

Figure 3.4 — Semi-log plot for saturation pressure (p_{sat}) versus the reservoir temperature (T) for the unconventional reservoir fluid database (138 data points).

From **Figs 3.1-3.4**, we observe the following:

- **Fig. 3.1:** The relationship of p_{sat} versus R_{sb} is probably the strongest "univariate" relationship for all of the variables considered. As comment, the initial power-law straight-line trend on **Fig. 3.1** is the same sort of observation that was made by Standing (1947). The "parabolic" behavior for R_{sb} > 4000 scf/STB seen in **Fig. 3.1** is observed for the retrograde gas condensate cases, and it will be important to watch this portion of the data in the correlation of p_{sat} as these data will deviate for all of the black oil p_{sat} correlations applied to this dataset.
- **Fig. 3.2:** There are two interpretations of the p_{sat} versus *API* (oil) gravity behavior. First, we could suggest that there is an approximate straight-line relationship for *API* < 50, and second, it appears that there is a roughly constant relationship for *API* > 50.
- **Fig. 3.3:** There is very little character in the behavior of p_{sat} versus *SG*, perhaps a slightly decreasing exponential trend.
- **Fig. 3.4:** For p_{sat} versus *T*, one could almost just state that there is no correlation, on perhaps just a slightly decreasing exponential trend. However, it is more likely that there is a slightly decreasing exponential trend for T < 250 ^{o}F, then a relatively constant trend for T > 250 ^{o}F.

The observations described above are provided not so much as suggestions of "univariate correlations," but rather as just behavioral observations. The creation of multivariate correlations is as much art as science, and understanding an apparent behavior on a variable-by-variable correlation can be useful as multivariate correlations are attempted.

CHAPTER IV

RE-FITTING OF HISTORICAL CORRELATIONS TO THE DATABASE FOR UNCONVENTIONAL RESERVOIR FLUIDS

This chapter considers the comparative application of the various historical correlations discussed in **Chapter II** for the saturation pressure(p_{sat}) and oil formation volume factor at saturation pressure (B_{ob}); and why there is a need for new correlations — particularly for hydrocarbon fluids from unconventional reservoirs. Reservoir temperature (T), stock tank oil gravity (API), solution gas-oil-ratio at saturation pressure (R_{sb}), specific gas gravity (γ_g), and the saturation pressure (p_{sat}) are all used (as required for a given correlation).

"Calculated versus Measured" log-log "validation" plots are used to compare the (calculated) results for each historical correlation to the historical (measured) data from database available for this work. The results of the non-parametric regression performed on the database (for the p_{sat} and B_{ob} data) are also shown on the validation plots to indicate the "best fit" of the dataset.

As a reminder, the following historical correlations are re-fitted to the project database:

Historical Correlation (date)	Saturation Pressure (p_{sat})	Oil Formation Volume Factor (B_{ob})
Standing (1947)	yes	yes
Lasater (1958)	yes	no
Vasquez and Beggs (1980)	yes	yes
Glaso (1980)	yes	yes
Al-Marhoun (1988)	yes	yes
Kartoatmodjo and Schmidt (1994)	yes	yes
Velarde, Blasingame, and McCain (1997)	yes	yes
Petrosky and Farshad (1998)	yes	yes
Dindoruk and Christman(2004)	yes	yes

For this work, standard non-linear regression analysis tools are used to fit the correlations to the dataset. The practice of "tunning" equations to data is now a quite common practice. In this work, Microsoft Excel and Python libraries/algorithms are used, with the Microsoft Excel "Solver" tool being the first choice (due to having the data in a standard Microsoft Excel structure). The Python libraries/algorithms are then used to check/refine the work performed in Microsoft Excel.

All regression approaches use some type of defined "error function" that is minimized (*e.g.*, using the squares of these differences yields the "least squares" approach). The more common statistical term for an "error function" is that of an "objective function," where (obviously) the goal is to minimize the objective function relative to some tolerance. For this work, the following "objective functions" are used: (note that "y" can be either p_{sat} or B_{ob}) depending on the case)

$$ADE = \sum_{i=1}^{n} \left| \ln(y_{meas})_i - \ln(y_{cal})_i \right| \qquad \text{Absolute Difference Error (units)} \quad \ldots\ldots(4.1)$$

$$LSE = \sum_{i=1}^{n} [\ln(y_{meas})_i - \ln(y_{cal})_i]^2 \qquad \text{Least-Squares Error (unit}^2\text{)} \ldots\ldots(4.2)$$

$$AARE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{(y_{meas})_i - (y_{cal})_i}{(y_{cal})_i} \right| \qquad \text{Average Absolute Relative Error (percent)} \ldots\ldots(4.3)$$

The non-parametric correlation of the data is performed by using the Xue et al (1997) implementation of Alternating Conditional Expectations (or ACE) algorithm proposed by Breiman and Friedman (1985). As a point of reference, the Xue et al implementation of the ACE algorithm is called "GRACE" as it provides a graphical rendering of the ACE algorithm outputs.

Recall that non-parametric methods do not provide a functional correlation, but rather relate the dependent and independent variables on a point-by-point basis, which should provide the minimum possible error metrics (statistics). Hence, any correlation equation which has lower error statistics that the GRACE "correlation" should be suspected of "over-fitting" the data.

4.1 Historical Correlations for the Saturation Pressure (p_{sat})

In this section, the saturation pressure (p_{sat}) correlations are applied to the project database and "calculated versus measured" plots and statistical results tables are provided for each historical correlation model. These "calculated versus measured" plots are shown in **Figs. 4.1 – 4.9** and statistical results are presented in **Tables 4.1 – 4.9**. As comment, most historical correlations tend to follow the structure of the function proposed by Standing (1947), but obviously over time, some quite complicated models have evolved (*e.g.*, the Dindoruk and Christman model [2004]).

Standing Correlation for Saturation Pressure

As background, Standing (1947) developed correlations not by using regression, but rather by a series of plots where he incorporated one independent variable at a time (*e.g.*, R_{sb}, γ_g (or SG), API, T). The re-fitted form of the Standing correlation is given as:

$$p_{sat} = 386.807 \left[\left[\frac{R_{sb}}{\gamma_g} \right]^{0.309} 10^{A} - 1.776 \right] \text{...(4.4)}$$

where the "A" parameter is given as a function of temperature, T, and stock tank oil gravity, API.

$$A = 1\times10^{-6}\ T\ - 2.6\times10^{-4}\ API \text{...(4.5)}$$

The results generated using Eqs. 4.4 and 4.5 are shown in **Fig 4.1** and the statistical results are given in **Table 4.1**. The most obvious feature in **Fig. 4.1** is the "hook" in the calculated p_{sat}

function, which is caused by the highly volatile oil and retrograde gas condensate cases. In short, the "hook" feature is present in all of the re-fitted historical correlations, and represents a limitation of these models.

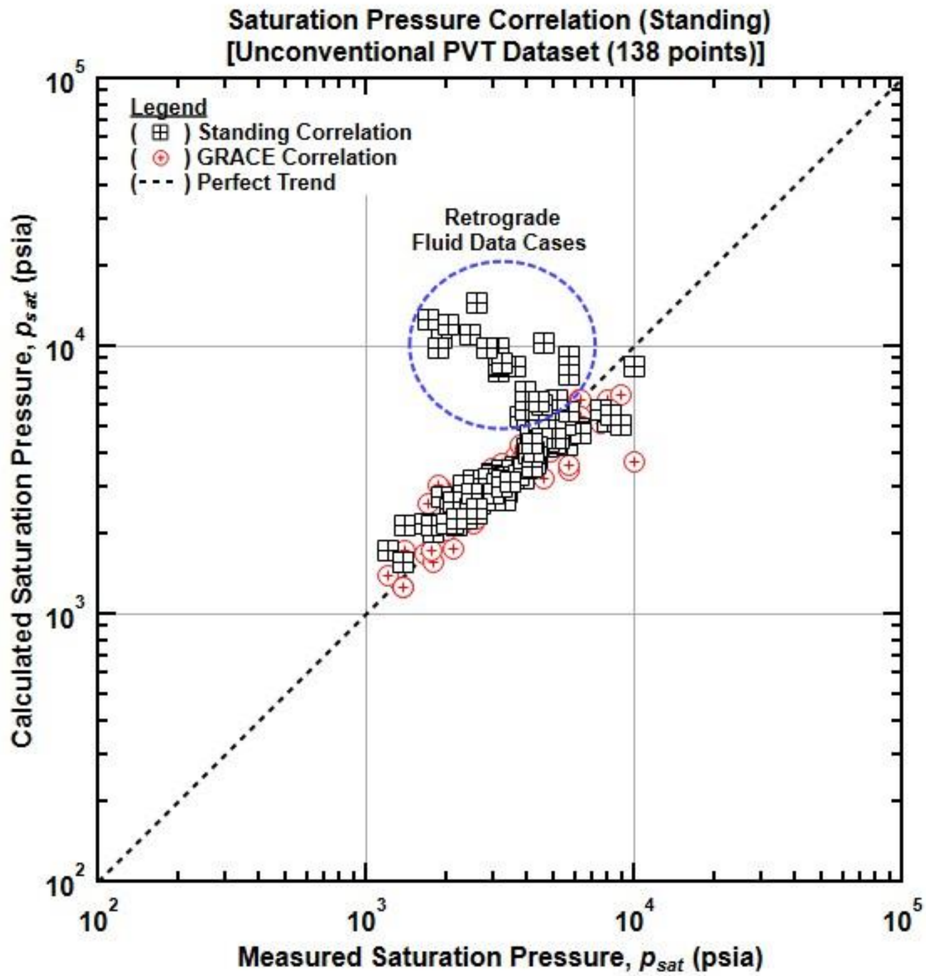

Figure 4.1 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Standing correlation and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.1 — Saturation Pressure (p_{sat}) — Standing model.

Parameter	Value
Sum of Squared Residuals, (psia)^2	31.322
Standard Deviation, psia	2,430.6
Variance, (psia)^2	5,907,833
Average Absolute Relative Error, %	20.03

Petrosky and Farshad Correlation for Saturation Pressure

Petrosky and Farshad (1998) employed a modification of the Standing correlation that added exponent parameters in the correlating function. The re-fitted form of the Petrosky and Farshad correlation is given as:

$$p_{sat} = 3530.37\left[\frac{R_{sb}^{\,0.1671}}{\gamma_g^{\,1\times10^{-6}}}10^{A} - 2.6568\right] \text{...}(4.6)$$

where the "*A*" parameter is given as a function of temperature, *T*, and stock tank oil gravity, *API*.

$$A = 0.009278\,T^{1\times10^{-6}} - 1\times10^{-5}\,API^{0.8287} \text{..}(4.7)$$

The results obtained using Eqs. 4.6 and 4.7 are presented in **Fig 4.2** and the statistical results are given in **Table 4.2**. As expected, most obvious feature in **Fig. 4.2** is the "hook" in the calculated p_{sat} function, and while this feature is not as significant as for the Standing correlation (**Fig. 4.1**), the Petrosky and Farshad formulation is also insufficient for practical purposes when highly volatile and retrograde gas condensate fluids are present in the dataset. Counterintuitively, the *AARE* for the Petrosky and Farshad correlation (21.1 percent) is actually higher than that of the Standing correlation (20.0 percent). This could be due to the relative simplicity of the Standing model.

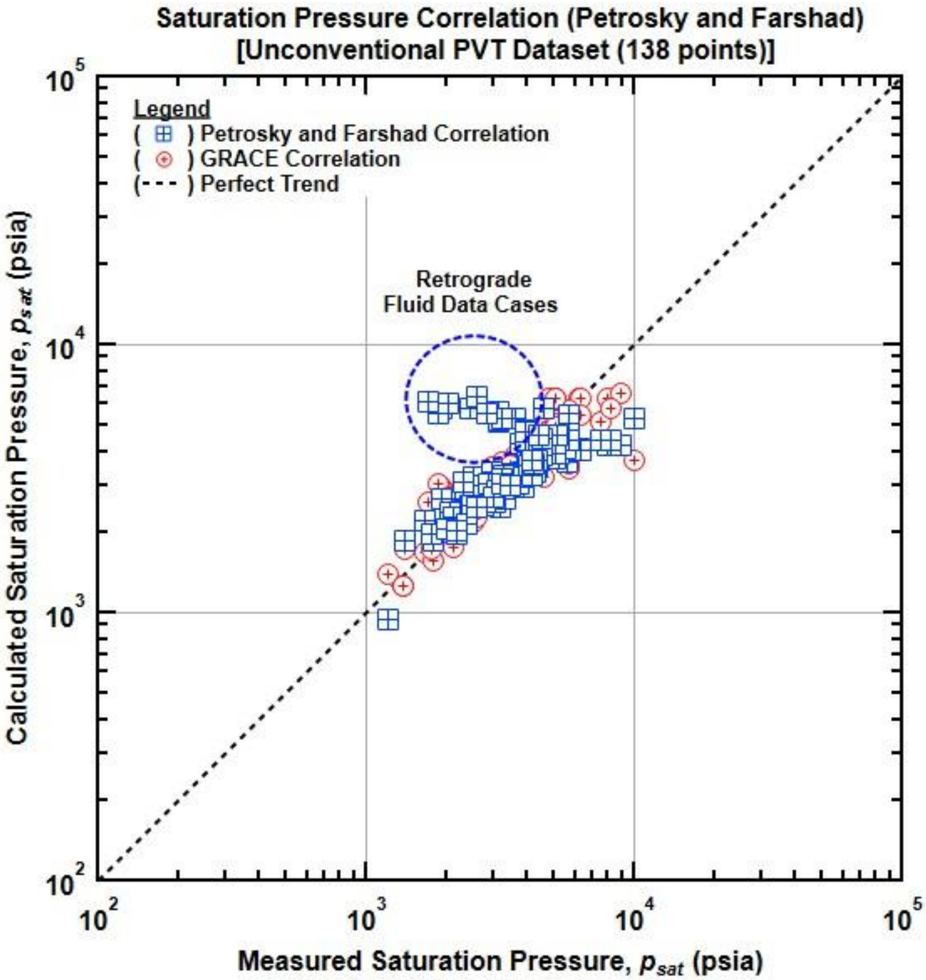

Figure 4.2 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Petrosky and Farshad correlation and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.2 — Saturation Pressure (p_{sat}) — Petrosky and Farshad model.

Parameter	Value
Sum of Squared Residuals, (psia)^2	14.300
Standard Deviation, psia	1,057.24
Variance, (psia)^2	1,117,756
Average Absolute Relative Error, %	21.06

Lasater Correlation for Saturation Pressure

Lasater's correlation (1958) for the saturation pressure (p_{sat}) is not of the "Standing" formulation and should likely have a different characteristic performance compared to the Standing-type relations. The re-fitted form of Lasater's correlation is given by:

$$p_{sat} = \frac{p_f \left[T + 459.67\right]}{y_g} \quad \text{(4.8)}$$

Where the following terms are defined:

$$p_f = 3.04536 - 1.9270\ \gamma_g + 0.407\ \gamma_g^{\ 2} \quad \text{(4.9)}$$

$$y_g = \frac{\dfrac{R_{sb}}{385.527}}{\dfrac{R_{sb}}{96.186} + \dfrac{-2305.32}{M_o}} \quad \text{(4.10)}$$

$$M_o = 3664.867 + \left[-101.484\right]\ API + 0.77884\ API^2 \quad \text{(4.11)}$$

The results using Eqs. 4.8-4.11 are presented in **Fig 4.3** and the statistical results are given in **Table 4.3**. As noted above, the Lasater formulation is substantially different than the Standing-type relations, and as such, the behavior of the calculated p_{sat} function shown in **Fig. 4.3** suggests that the Lasater model is not appropriate for this dataset, or has not been effectively fitted to the data, or both. In particular, the "horizontal spread" of the calculated p_{sat} function shown in **Fig. 4.3** is generally indicative of a failed regression, but as this cannot be uniquely verified, it could simply be that the character of the Lasater form is inadequate for this specific type of data.

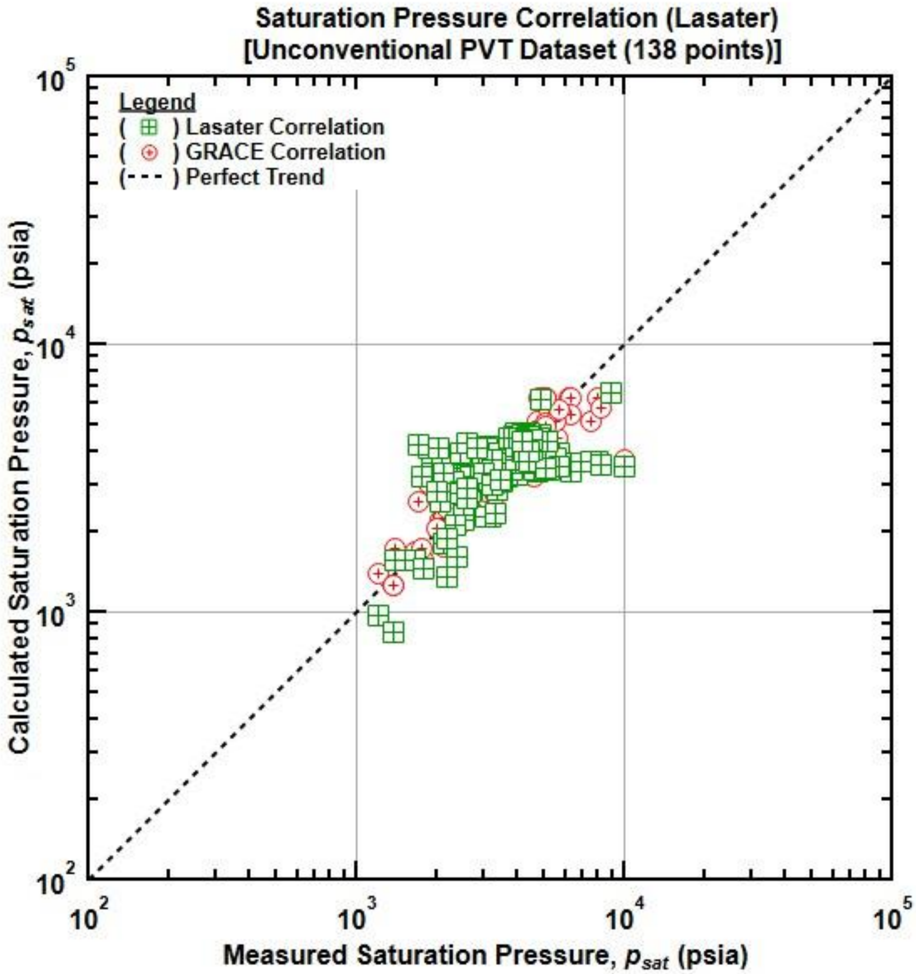

Figure 4.3 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Lasater correlation and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.3 — Saturation Pressure (p_{sat}) — Lasater model.

Parameter	Value
Sum of Squared Residuals, $(\text{psia})^2$	14.415
Standard Deviation, psia	956.66
Variance, $(\text{psia})^2$	915,192.56
Average Absolute Relative Error, %	21.26

Vasquez and Beggs Correlation for Saturation Pressure

The saturation pressure formulation developed by Vasquez and Beggs (1980) is another "non-Standing"-type formulation (although one could argue that its power-law structure is quite similar), and is actually a single relation but uses 2 sets of coefficients (for oil gravity $\leq 30\ ^{o}$API and for oil gravity > 30 ^{o}API). The re-fitted form of the Vasquez and Beggs correlation is given by:

$$p_{sat} = \left[C_1 \left[\frac{R_{sb}}{\gamma_g} \right] 10^{\frac{C_2 API}{T+459.67}} \right]^{\frac{1}{C_3}} \quad \text{.....(4.12)}$$

Coefficient	$API \leq 30$	$API > 30$
C_1	1.091×10^5	3.405×10^6
C_2	2.3913	2.7754
C_3	1×10^{-6}	1×10^{-6}

The results obtained using Eq. 4.12 are presented in **Fig 4.4** and the statistical results are given in **Table 4.4**, perhaps due to the similarity of the Vasquez and Beggs relation to the Standing-type formulation, a very strong "hook" feature on **Fig 4.4** is noted. In fact, the performance of Eq. 4.12 (**Fig. 4.4**) is very similar in a visual sense to the re-fitted Standing correlation (*i.e.*, Eqs. 4.4 and 4.5 plotted on **Fig. 4.1**). Comparing, the re-fit of the Standing correlation has an *AARE* of 20.0 and the re-fit of the Vasquez and Beggs correlation has an *AARE* of 20.3 percent, which is virtually indistinguishable in both a statistical and a visual sense.

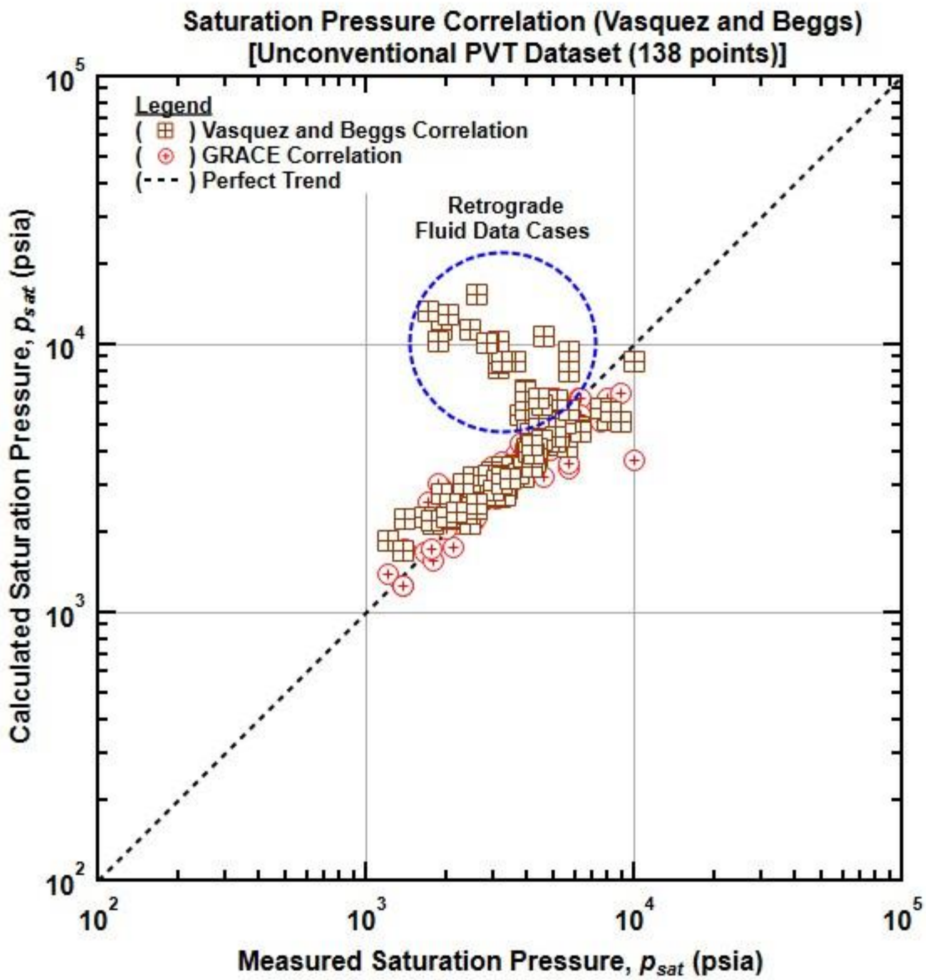

Figure 4.4 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Vasquez and Beggs correlation and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.4 — Saturation Pressure (p_{sat}) — Vasquez and Beggs model.

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	32.915
Standard Deviation, psia	2,542.70
Variance, $(psia)^2$	6,465,340
Average Absolute Relative Error, %	20.33

Glaso Correlation for Saturation Pressure

The saturation pressure formulation developed by Glaso (1980) is quadratic polynomial formulation with a power-law term that captures of all of the independent variables (R_{sb}, γ_g (or *SG*), *API*, *T*). Obviously this relation is not of the "Standing" form (*i.e.*, the quadratic polynomial form), but the power-law term may yield a similar behavior as the "Standing"-type relations. The re-fitted form of the Glaso correlation is given by:

$$p_{sat} = 2.3094 + 2.1375 \ \log p_{sat}^{\ *} - 0.3472 \ \left[\log p_{sat}^{\ *}\right]^2 \quad \text{.........(4.13)}$$

$$p_{sat}^{\ *} = \left[\frac{R_{sb}}{\gamma_g}\right]^{4.807} \frac{T^{1\times10^{-16}}}{API^{2.388\times10^{-4}}} \quad \text{.........(4.14)}$$

The results obtained using Eqs. 4.13 and 4.14 are presented in **Fig 4.5** and the statistical results are given in **Table 4.5**. As with the Standing-type relations, a very strong "hook" feature is observed in the calculated model results shown in **Fig 4.5**. As with other cases, this behavior is very similar to the re-fitted Standing correlation (*i.e.*, Eqs. 4.4 and 4.5 plotted on **Fig. 4.1**). Comparing, the re-fit of the Standing correlation has an *AARE* of 20.0 and the re-fit of the Glaso correlation has an *AARE* of 20.0 percent, which is identical to the results of the Standing correlation.

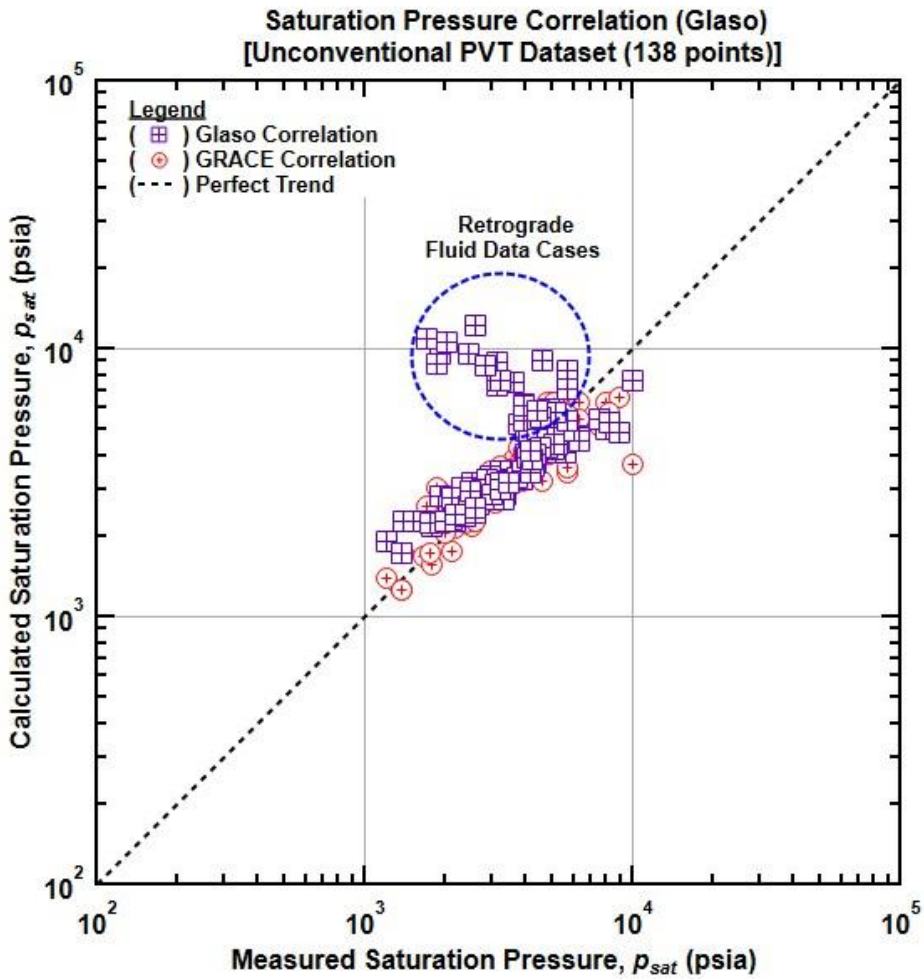

Figure 4.5 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Glaso correlation and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.5 — Saturation Pressure (p_{sat}) — Glaso model.

Parameter	Value
Sum of Squared Residuals, (psia)^2	26.726
Standard Deviation, psia	2,022.99
Variance, (psia)^2	4,092,484
Average Absolute Relative Error, %	20.04

Al-Marhoun Correlation for Saturation Pressure

The saturation pressure formulation developed by Al-Marhoun (1988) is a multivariate power-law formulation which is not a "Standing"-type form. The re-fitted form of the Al-Marhoun correlation is given by:

$$p_{sat} = 1.332\times10^{-2}\ R_{sb}^{\ 0.38396}\ \gamma_g^{\ 1\times10^{-6}}\ \gamma_o^{\ 1.612146}\ T^{1.542573} \qquad \text{.......................................(4.15)}$$

Where the oil specific gravity is defined as:

$$\gamma_o = \frac{141.5}{131.5 + API} \qquad \text{...(4.16)}$$

The results obtained using Eqs. 4.15 and 4.16 are presented in **Fig 4.6** and the statistical results are given in **Table 4.6**. The performance of the Al-Marhoun formulation is actually much less well-behaved than the Standing-type relations as the calculated results are somewhat more dispersed and the "hook" feature is larger than in other cases (see **Fig 4.6**). The re-fit of the Al-Marhoun correlation has an *AARE* of 24.2 percent, which is the second highest error metric for all of the correlations tested.

If permitted to speculate, it could be suggested that the simple power-law inter-relation proposed by Eq. 4.15 is not capable of representing more complex behavior which connects the independent variables with the saturation pressure. As a summary statement, the Al-Marhoun relation assumes that all variables are related by a simple power-law expression, where each variable has only a single parameter relationship — this appears to be inadequate for the correlation of the saturation pressure dataset for this project.

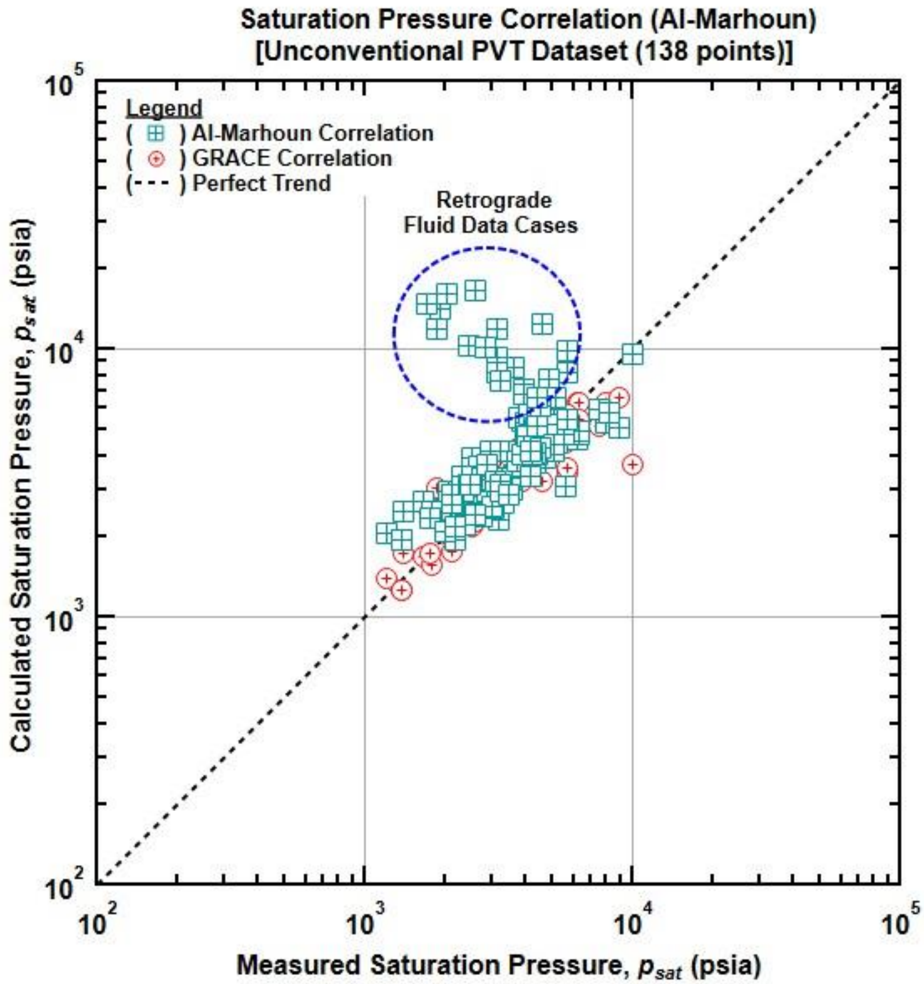

Figure 4.6 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Al-Marhoun correlation and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.6 — Saturation Pressure (p_{sat}) — Al-Marhoun model.

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	35.021
Standard Deviation, psia	2,432.13
Variance, $(psia)^2$	5,915,237
Average Absolute Relative Error, %	24.20

Kartoatmodjo and Schmidt Correlation for Saturation Pressure

The formulation proposed by Kartoatmodjo and Schmidt (1994) is quite similar in a mathematical sense to the form proposed by Al-Marhoun (1988), with the exception that Kartoatmodjo and Schmidt consider a different set of correlation coefficients for oil gravity ≤30 °API and for oil gravity > 30 °API (analogous to the Vasquez and Beggs approach). The re-fitted form of the Kartoatmodjo and Schmidt correlation is given by:

$$p_{sat} = \left[\frac{R_{sb}}{0.04476\ \gamma_g^{\ 8.5\times10^{-11}}\ 10^{8.78383}\ API^{(T+459.67)}} \right]^{0.88234} \qquad (API \leq 30) \text{ ....(4.17)}$$

$$p_{sat} = \left[\frac{R_{sb}}{1.3\text{E-}5\ \gamma_g^{\ 8.7\times10^{-11}}\ 10^{3.2\times10^{-8}}\ API^{(T+459.67)}} \right]^{0.43818} \qquad (API > 30) \text{ ....(4.18)}$$

The results obtained using Eqs. 4.17 and 4.18 are presented in **Fig 4.7** and the statistical results are given in **Table 4.7**. The Kartoatmodjo and Schmidt relations yield a response very similar to the Standing-type relations, and a very strong "hook" feature is observed in **Fig 4.7**. Comparing, the re-fit of the Standing correlation has an *AARE* of 20.0 and the re-fit of the Kartoatmodjo and Schmidt correlation has an *AARE* of 21.1 percent, so the performance of these relations can be considered "comparable," but on balance, it is probably best to have a correlation with only a single set of results parameters, as opposed to the Kartoatmodjo and Schmidt approach which considers different set of correlation coefficients for oil gravity ≤30 °API and for oil gravity > 30 °API.

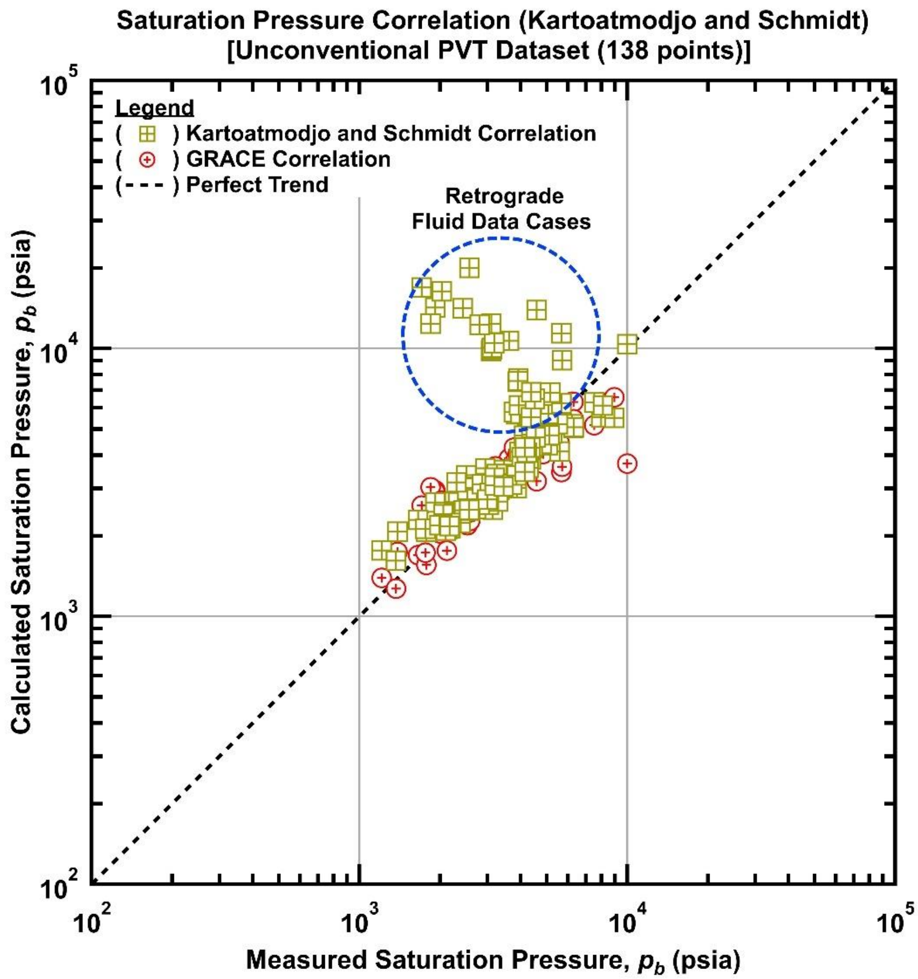

Figure 4.7 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Kartoatmodjo and Schmidt model and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.7 — Saturation Pressure (p_{sat}) — Kartoatmodjo and Schmidt model.

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	42.435
Standard Deviation, psia	3,336.57
Variance, $(psia)^2$	11,132,681
Average Absolute Relative Error, %	21.13

Velarde Correlation for Saturation Pressure

Velarde et al (1997) employed a modification of the Petrosky and Farshad (1998) relation, which is actually a modification of the original Standing correlation. The Velarde et al does has additional exponents, and depending on the robustness of the regression approach, these exponents may yield problematic (*i.e.*, less statistically optimal) results. The re-fitted form of the Velarde et al correlation is given as:

$$p_{sat} = 1500.675 \left[\frac{R_{sb}^{0.052729}}{\gamma_g^{1.216\times10^{-11}}} 10^A - 0.229257 \right]^{2.35544} \quad \text{..........(4.19)}$$

where the "*A*" parameter is given as a function of temperature, *T*, and stock tank oil gravity, *API*.

$$A = 0.01271\, \mathrm{T}^{0.21736} - 2.02\times10^{-6}\, API^{1.601} \quad \text{..........(4.20)}$$

The results obtained using Eqs. 4.19 and 4.20 are presented in **Fig 4.8** and the statistical results are given in **Table 4.8**. From the character of the calculated results in **Fig. 4.8**, one can only conclude that the regression has failed — most likely due to the additional exponents in the Velarde et al relationship. The results given by Eqs. 4.19 and 4.20 do yield the highest errors for any correlation, but more concerning is the actual character of these results as shown in **Fig. 4.8**, clearly something is critically wrong with this correlation attempt.

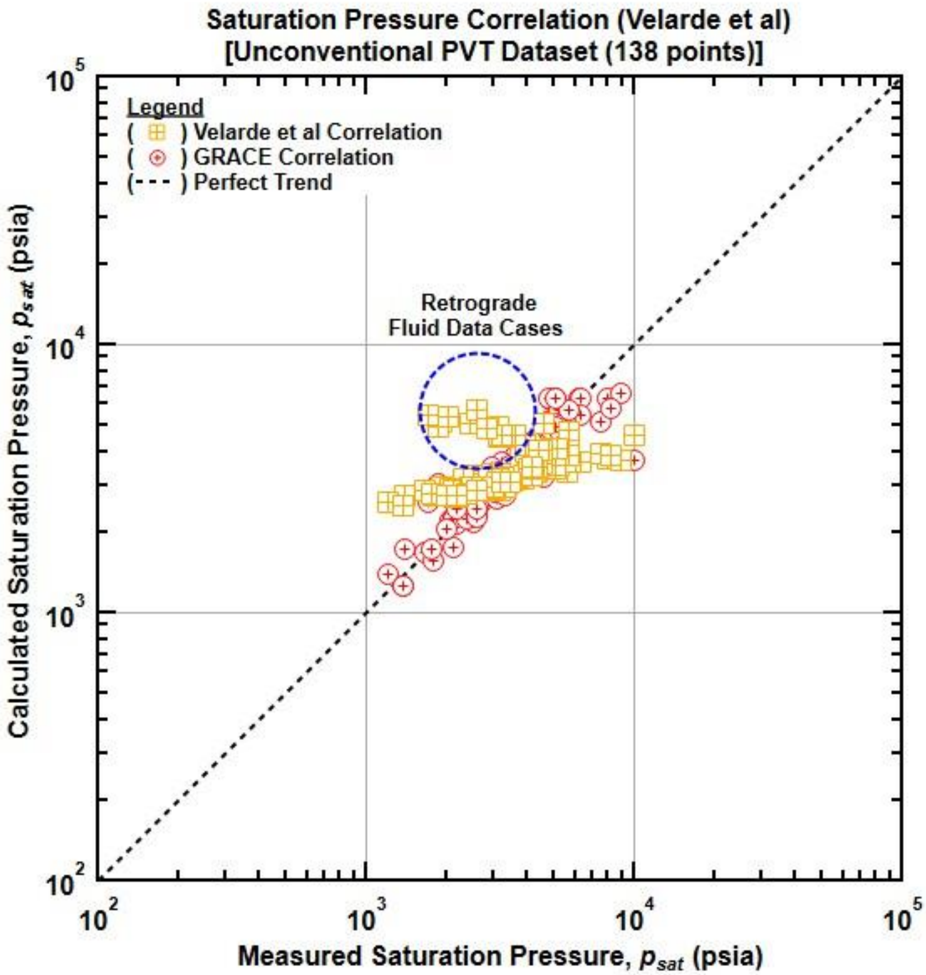

Figure 4.8 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Velarde et al model and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.8 — Saturation Pressure (p_{sat}) — Velarde et al model.

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	27.258
Standard Deviation, psia	663.54
Variance, $(psia)^2$	440,298
Average Absolute Relative Error, %	27.26

Dindoruk and Christman Correlation for Saturation Pressure

Dindoruk and Christman (2004) proposed a form similar to Standing (1947) and Petrosky and Farshad (1998), with a very complex relationship for an internal exponent parameter. Their goal was to achieve a generalized best fit model. The re-fitted form of the Velarde et al correlation is given as:

$$p_{sat} = 1.86998\left[\frac{R_{sb}^{\ 1.22145}}{\gamma_g^{\ 1.37051}}\ 10^{A} + 0.011688\right] \quad \text{...(4.21)}$$

where the "*A*" parameter is given as a function of all of the variables (R_{sb}, γ_g (or *SG*), *API, T*).

$$A = \frac{0.86429T^{8.5421} + 0.009925API^{1.77618}}{\left[44.25 + \frac{2R_{sb}^{\ 5.78962}}{\gamma_g^{\ -0.084226}}\right]^2} \quad \text{...(4.22)}$$

The results obtained using Eqs. 4.21 and 4.22 are presented in **Fig 4.9** and the statistical results are given in **Table 4.9**. The calculated results shown in **Fig. 4.9** are, in general, very tight around both the perfect correlation trend and the GRACE correlation, but this model still demonstrates the "hook" feature, indicating that it does not (or cannot) capture the behavior of the highly volatile oil samples and the retrograde gas condensate samples. The *AARE* for the Dindoruk and Christman correlation is 20.4 percent, but that is actually a bit higher than that of the Standing correlation which has an *AARE* of 20.0 percent. As was stated for other comparisons, it is likely that the Standing correlation performs relatively better because it is the simplest of all the correlation relations (with the possible exception of the Al-Marhoun correlation).

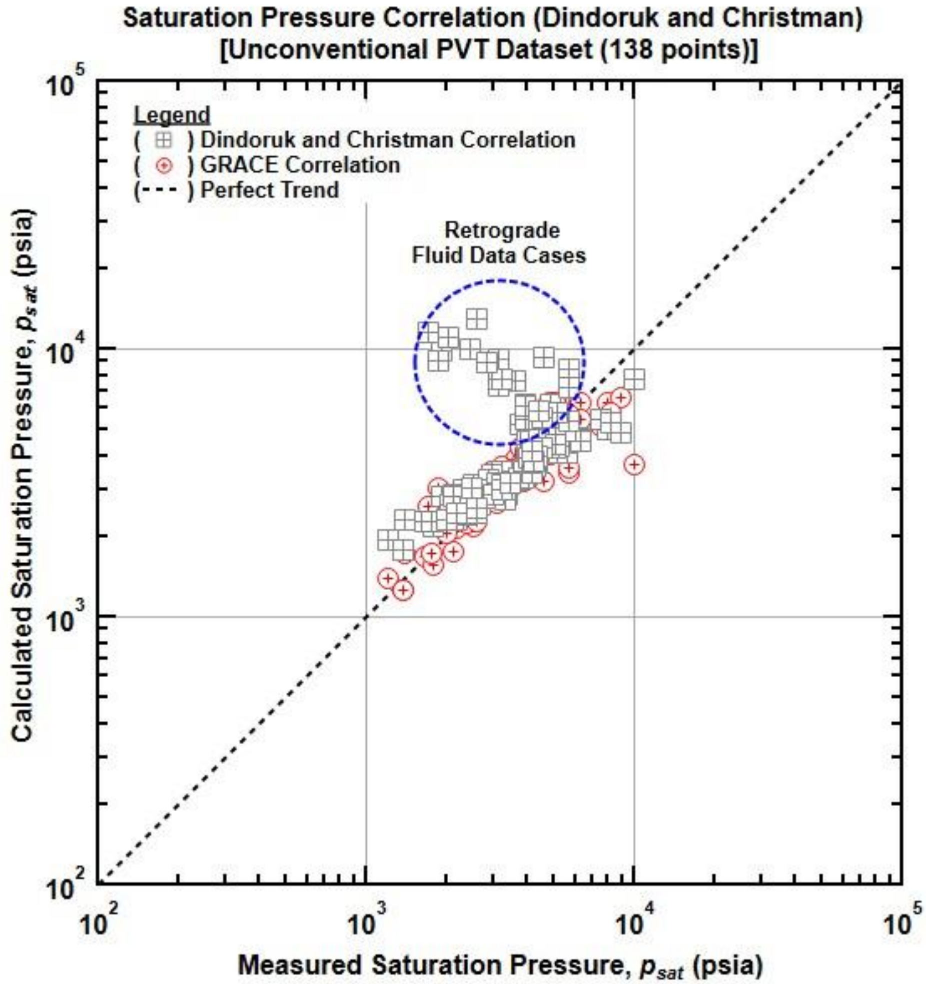

Figure 4.9 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); Dindoruk and Christman model and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 4.9 — Saturation Pressure (p_{sat}) — Dindoruk and Christman model.

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	27.985
Standard Deviation, psia	2,106.62
Variance, $(psia)^2$	4,437,844
Average Absolute Relative Error, %	20.35

As noted in earlier discussions in this section, one overriding concern for this work is that there are only 138 datasets for use in the correlation of saturation pressure p_{sat}; of which probably 30 percent are highly volatile oils and retrograde gas condensate cases (which certainly challenge, if not bias the effort to correlate these data). A summary of the performances for all of the saturation pressure (p_{sat}) correlations is provided in Table 4.10, and the most significant observation is that no single correlation presents an *AARE* of less than 20 percent, except for the GRACE non-parametric correlation which has an *AARE* of 13.2 percent. In short, these correlations were traditionally used for black oil reservoir fluids and the nature of the database (black oils, highly volatile oils, and retrograde gas condensates) precludes a strong performance in a statistical sense for these historical p_{sat} correlations.

Table 4.10 — Statistics: Saturation Pressure, p_{sat} (Various Correlations).

Source	Standard Deviation, psia	Variance, (psia)^2	Average Absolute Relative Error, % (this work)
Standing (1947)	2,431	5,907,833	20.0
Lasater (1958)	957	915,193	21.3
Vasquez and Beggs (1980)	2,543	6,465,340	20.3
Glaso (1980)	2,023	4,092,484	20.0
Al-Marhoun (1988)	2,432	5,915,237	24.2
Kartoatmodjo and Schmidt (1994)	3,337	11,132,681	21.1
Velarde, Blasingame, and McCain (1997)	664	440,298	27.3
Petrosky and Farshad (1998)	1,057	1,117,756	21.1
Dindoruk and Christman (2004)	2,107	4,437,844	20.3
Non-Parametric Solution (GRACE)	1,197	1,432,965	13.2

4.2 Application of Oil Formation Volume Factor Correlations

In this section, the oil formation volume factor (B_{ob}) correlations presented in **Chapter II** are applied to the project database and "calculated versus measured" plots and statistical results tables are provided for each historical correlation model. These "calculated versus measured" plots are shown in **Figs. 4.10 – 4.17** and statistical results are presented in **Tables 4.11 – 4.18**. In this section, the oil formation volume factor (B_{ob}) is correlated with reservoir temperature (T), solution gas-oil ratio (R_{sb}), stock tank oil (specific) gravity (γ_o), and specific gas gravity (γ_g) — or more compactly, $B_{ob} = f(T, R_{sb}, \gamma_o, \gamma_g)$. It is important to note that there are only 46 datasets available for the correlation of the oil formation volume factor (B_{ob}), and that this relatively low quantity of data will likely bias data correlations.

Standing Correlation for Oil Formation Volume Factor

In 1947, Standing proposed a power-law-based formulation for the oil formation volume factor (B_{ob}) — the re-fitted form of the Standing correlation for B_{ob} is given as:

$$B_{ob} = 0.843681 + 8.0 \times 10^{-5} \left[R_{sb} \left[\frac{\gamma_g}{\gamma_o} \right]^{0.0788} + 5.27268T \right]^{1.2107} \quad \text{.....................................(4.23)}$$

The results generated using Eq. 4.23 are presented in **Fig 4.10** and the statistical results are given in **Table 4.11**. The most obvious features in **Fig. 4.10** are spurious points in the calculated B_{ob} function (in fact, these spurious calculated points occur for all of the correlation models). The *AARE* for the Standing model is 5.29 percent, and the *AARE* for the GRACE non-parametric correlation is 3.41 percent.

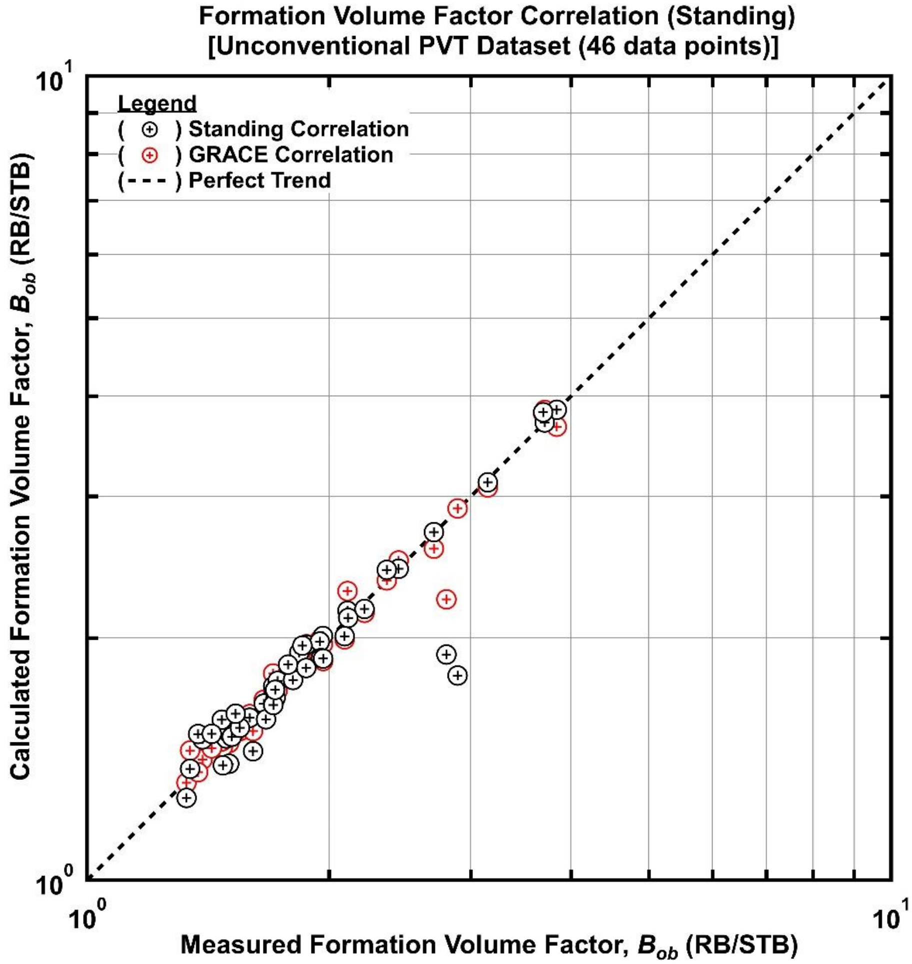

Figure 4.10 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Standing model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.11 — Statistics: Oil FVF (B_{ob}) — Standing model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.445
Standard Deviation, RB/STB	0.610
Variance, $(RB/STB)^2$	0.372
Average Absolute Relative Error, %	5.288

Petrosky and Farshad Correlation for Oil Formation Volume Factor

Petrosky and Farshad (1990, 1998) developed a modification of the Standing (1947) model (adding an exponent to the temperature term) and this model should be expected to produce a slightly better correlation to the original Standing work. The re-fitted form of the Petrosky and Farshad correlation for B_{ob} is given as:

$$B_{ob} = 0.927938 + 7.6145\times10^{-5}\left[\frac{R_{sb}^{0.34199}\gamma_g^{\;0.00025}}{\gamma_o^{\;0.74221}} + 0.2659T^{0.5832}\right]^{3.1677} \quad \text{................(4.24)}$$

The results generated using Eq. 4.24 are presented in **Fig 4.11** and the statistical results are given in **Table 4.12**. As Eq. 4.24 (Petrosky and Farshad) is a (more complex) modification of Eq. 4.23 (Standing), Eq. 4.24 may be expected to have better statistical accuracy, which is true — the *AARE* for the Petrosky and Farshad correlation is 4.97 percent and the *AARE* for the Standing correlation is 5.29 percent. However, comparing the results of these 2 relations in **Figs. 4.10 and 4.11** (respectively), there is very little difference in these correlations, both provide an excellent visual match of the data.

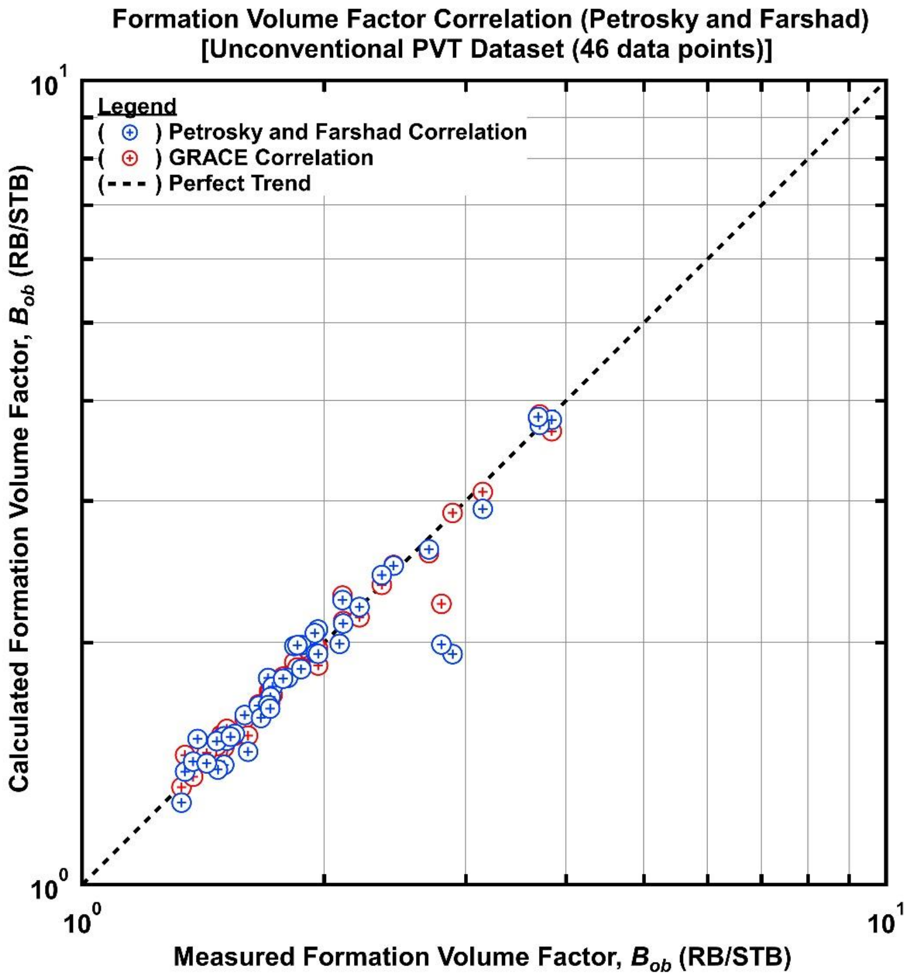

Figure 4.11 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Petrosky and Farshad model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.12 — Statistics: Oil FVF (B_{ob}) — Petrosky and Farshad model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.351
Standard Deviation, RB/STB	0.602
Variance, $(RB/STB)^2$	0.363
Average Absolute Relative Error, %	4.970

Vasquez and Beggs Correlation for Oil Formation Volume Factor

Vasquez and Beggs (1980) developed a relatively simple expansion for the oil formation volume factor that, like their correlation for saturation pressure, requires 2 *different sets of correlating coefficients* (for conditions $\leq$ or > 30 °API). In concept, this approach should yield a better correlation at the local level (*i.e.*, $\leq$ or > 30 °API)., but the mathematical simplicity of the formulation may yield less accuracy in general. The re-fitted form of the Vasquez and Beggs correlation for B_{ob} is given as:

$$B_{ob} = 1 + C_1 R_{sb} + (T - 60)\left[\frac{API}{\gamma_g}\right](C_2 + C_3 R_{sb}) \quad \text{.......}(4.25)$$

Coefficient	$API \leq 30$	$API > 30$
C_1	1.092×10^{5}	4.898×10^{-4}
C_2	2.391297	2.682×10^{-5}
C_3	1×10^{-6}	1.337×10^{-9}

The results generated using Eq. 4.25 are presented in **Fig 4.12** and the statistical results are given in **Table 4.13**. As caution, the simplicity of Eq. 4.25 may lead to less statistical accuracy, but the visual match of the results computed with the Vasquez and Beggs correlation shown in **Fig 4.12** is excellent. However, the statistical performance of the Vasquez and Beggs correlation is weak (*AARE* = 5.52 percent), which is the second worst performing correlation after the Velarde et al result (*AARE* = 5.54 percent).

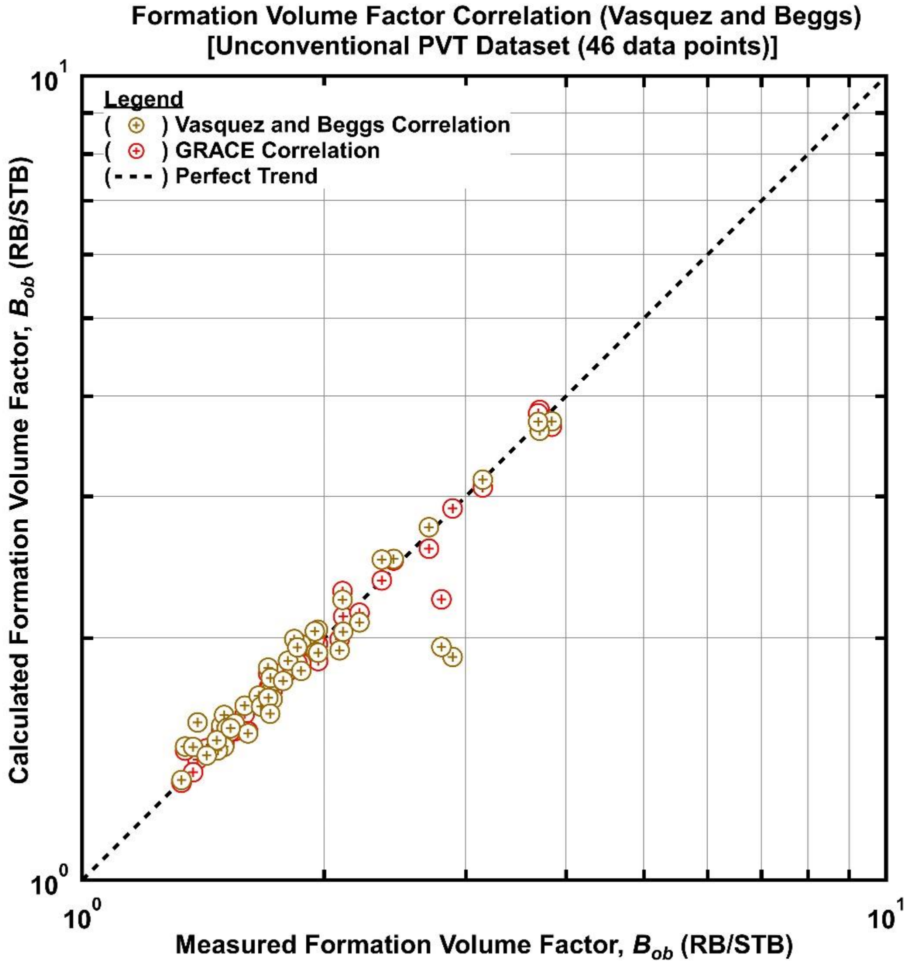

Figure 4.12 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Vasquez and Beggs model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.13 — Statistics: Oil FVF (B_{ob}) — Vasquez and Beggs model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.398
Standard Deviation, RB/STB	0.585
Variance, $(RB/STB)^2$	0.343
Average Absolute Relative Error, %	5.520

Glaso Correlation for Oil Formation Volume Factor

The Glaso (1980) correlation is a somewhat unusual formulation in mathematical structure, using a power-law internal variable with a linear temperature correction defined by R_{sb}, γ_g, γ_o, and T; where this internal variable is then used in another expansion. Given this relatively complex formulation, the Glaso correlation is, in concept, expected to yield better-than-average statistical results. The re-fitted form of the Glaso correlation for B_{ob} is given as:

$$B_{ob} = 0.94748 + 10^{\left[-4.47225 + 0.83529 \log B_{ob}^{*} - 0.20826 \left[\log B_{ob}^{*}\right]^{1.11297}\right]} \quad \text{.............................(4.26)}$$

Where the " B_{ob}*" parameter contains R_{sb}, γ_g, γ_o, and T; and is given by:

$$B_{ob}^{*} = R_{sb}\left[\frac{\gamma_g}{\gamma_o}\right]^{0.19847} + 4.27078T \quad \text{..(4.27)}$$

The results generated using Eqs. 4.26 and 4.27 are presented in **Fig 4.13** and the statistical results are given in **Table 4.14**. In **Fig 4.14** the performance of the Glaso correlation appears to be excellent, and while the usual outlier points are significant, the Glaso results compare very well (in a visual sense) with the perfect correlation trend and the results of GRACE non-parametric correlation. The statistical performance of the Glaso correlation is excellent (*AARE* = 5.12 percent) compared to the Dindoruk and Christman correlation (*AARE* of 4.30 percent) and the GRACE non-parametric correlation (*AARE* of 3.41 percent).

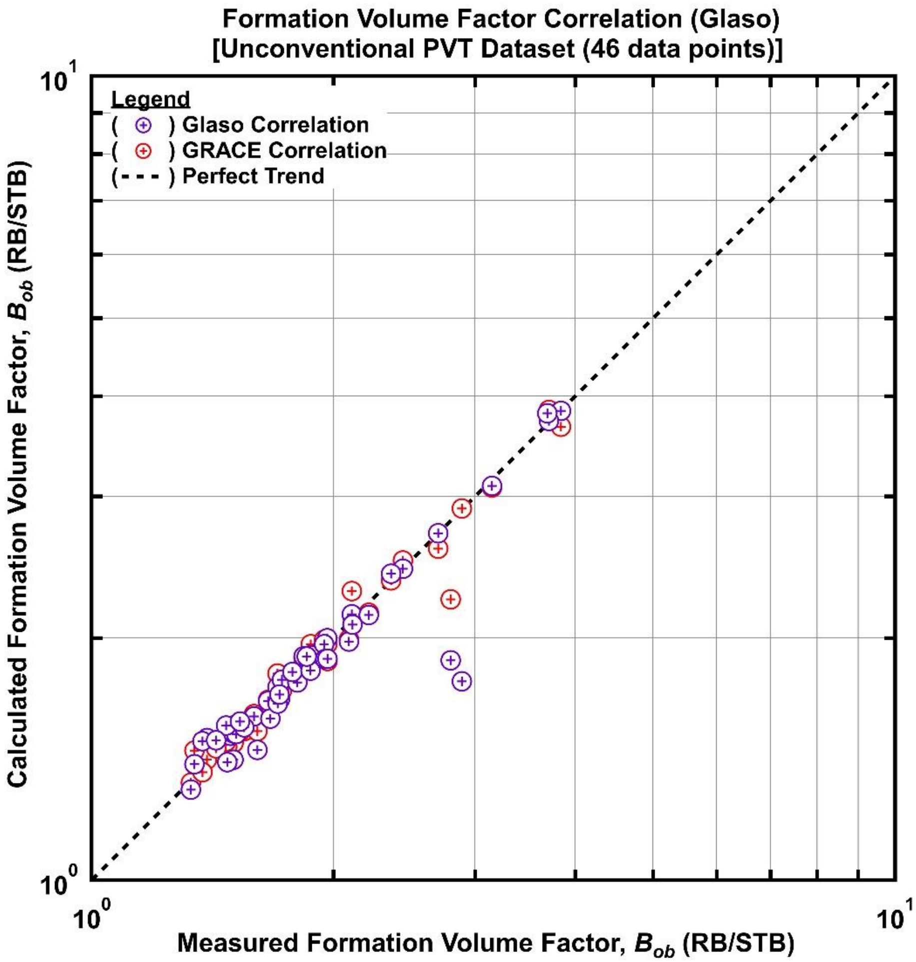

Figure 4.13 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Glaso model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.14 — Statistics: Oil FVF (B_{ob}) — Glaso model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.458
Standard Deviation, RB/STB	0.607
Variance, $(RB/STB)^2$	0.368
Average Absolute Relative Error, %	5.121

Al-Marhoun Correlation for Oil Formation Volume Factor

The correlation proposed by Al-Marhoun (1988) uses a power-law internal variable defined by R_{sb}, γ_g, and γ_o; where this internal variable is then used in a polynomial expansion that contains a single-term component to represent the influence of temperature (T). In concept, the Al-Marhoun correlation should exhibit better than average statistical performance due to its polynomial expansion. The re-fitted form of the Al-Marhoun correlation for B_{ob} is given as:

$$B_{ob} = 0.509 + 8.915\times10^{-4}\ (T + 459.67) + 1.849\times10^{-3} X + 1.449\times10^{-6} X^2 \text{ ....(4.28)}$$

Where the "X" parameter contains R_{sb}, γ_g, and γ_o; and is given in power-law form as:

$$X = R_{sb}{}^{0.806752} \gamma_g{}^{0.32296} + \gamma_o{}^{-1.202101} \text{ ....(4.29)}$$

The results generated using Eqs. 4.28 and 4.29 are presented in **Fig 4.14** and the statistical results are given in **Table 4.15**. The character of the Al-Marhoun calculated results shown in **Fig 4.14** are very good (with outliers noted), and are particularly well-correlated with the results of GRACE non-parametric correlation (at least visually).

What is surprising is that the polynomial expansion of terms in the Al-Marhoun correlation did not yield best-in-class statistical results. In fact, with an *AARE* = 5.33 percent, the Al-Marhoun correlation is one of the worst performing correlations (at least statistically). This performance suggests that while the polynomial expansion in the Al-Marhoun correlation may provide more "flexibility" in concept, this feature did not yield significant statistical benefit for this particular case.

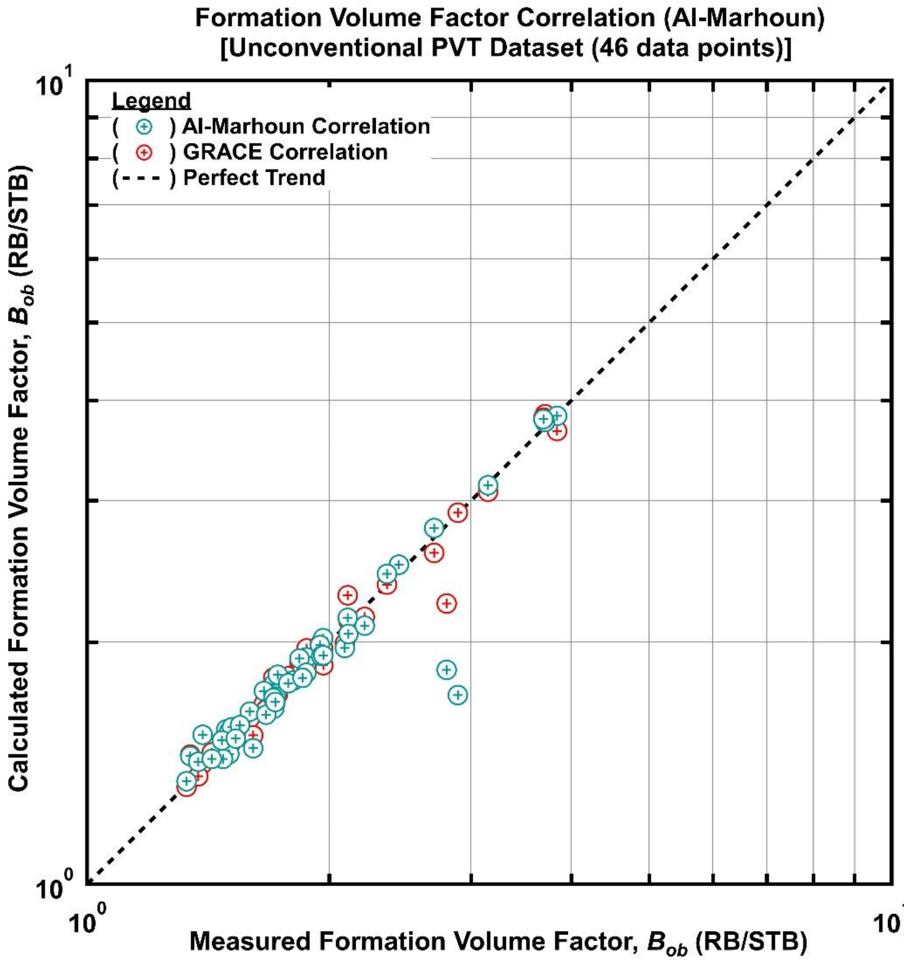

Figure 4.14 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Al-Marhoun model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.15 — Statistics: Oil FVF (B_{ob}) — Al-Marhoun model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.498
Standard Deviation, RB/STB	0.608
Variance, $(RB/STB)^2$	0.370
Average Absolute Relative Error, %	5.330

Kartoatmodjo and Schmidt Correlation for Oil Formation Volume Factor

Kartoatmodjo and Schmidt (1994) also developed a modification of the Petrosky and Farshad (1990,1998) model and is essentially the same form (with minor mathematical arrangements) as the Velarde et al correlation (1997). The re-fitted form of the Kartoatmodjo and Schmidt correlation for B_{ob} is given as:

$$B_{ob} = 0.83161 + 2.887\times10^{-4}\left[R_{sb}^{0.62788}\left[\frac{\gamma_g^{-0.00932}}{\gamma_o^{0.98995}}\right] + 0.23709T\right]^{1.60977} \quad \text{...................(4.30)}$$

The results generated using Eq. 4.30 are presented in **Fig 4.15** and the statistical results are given in **Table 4.16**. As observed in **Fig. 4.15**, the Kartoatmodjo and Schmidt calculated results correspond very well with the perfect correlation trend and the results of GRACE non-parametric correlation (with the usual outlier points noted).

What is remarkable is the significantly better statistical performance of the Kartoatmodjo and Schmidt correlation (*AARE* = 5.10 percent) compared to the Velarde et al correlation (*AARE* = 5.54 percent), where it is noted that *these correlations have essentially the same mathematical formulations*. Lastly, for comparison, the GRACE non-parametric correlation has an *AARE* of 3.41 percent and the Petrosky and Farshad model has an *AARE* of 4.97 percent (the Kartoatmodjo and Schmidt and Petrosky and Farshad models also have very similar mathematical constructions).

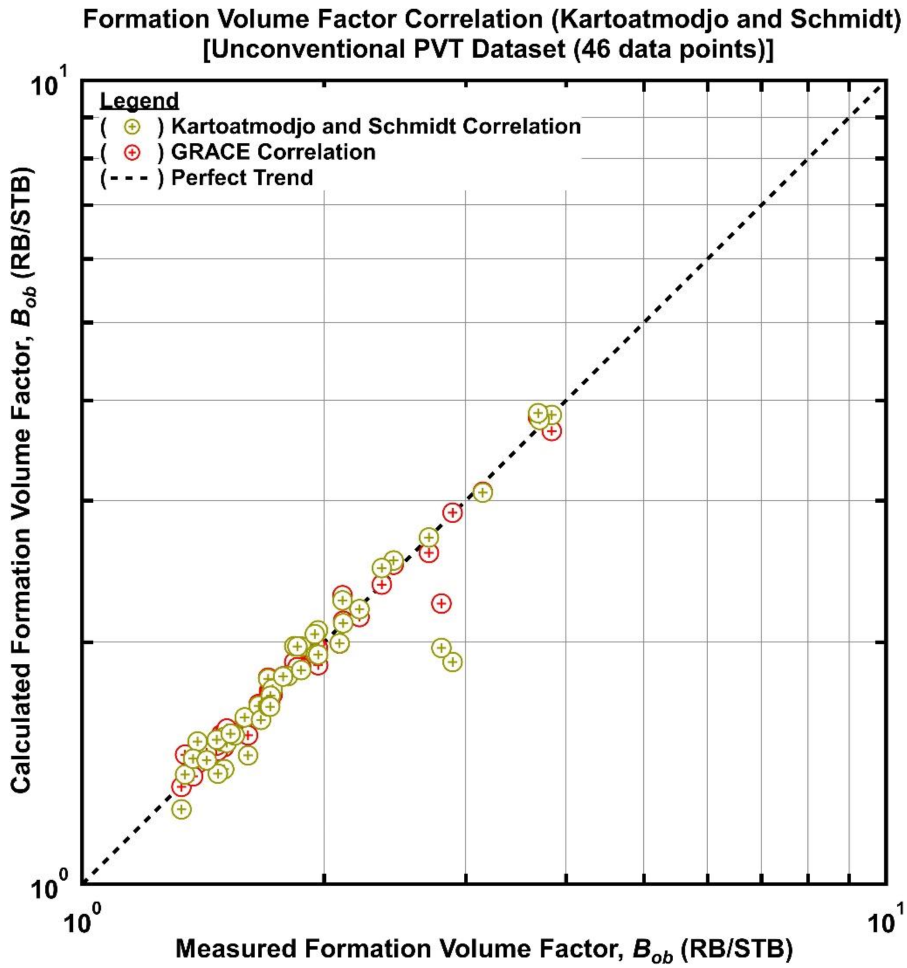

Figure 4.15 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Kartoatmodjo and Schmidt model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.16 — Statistics: Oil FVF (B_{ob}) — Kartoatmodjo and Schmidt model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.380
Standard Deviation, RB/STB	0.624
Variance, $(RB/STB)^2$	0.389
Average Absolute Relative Error, %	5.095

Velarde et al Correlation for Oil Formation Volume Factor

Velarde et al (1997) developed a modification of the Petrosky and Farshad (1990,1998) model (actually reduced by one parameter) and expressed the final result in a slightly more simplified mathematical form. The re-fitted form of the Velarde et al correlation for B_{ob} is given as:

$$B_{ob} = 0.889 + 4.662 \times 10^{-6} \left[R_{sb}{}^{0.498} \gamma_g{}^{-0.011} \gamma_o{}^{0.768} + 1.245 T \right]^{1.804} \text{..............................} (4.31)$$

The results generated using Eq. 4.31 are presented in **Fig 4.16** and the statistical results are given in **Table 4.17**. By inspection in Fig. **4.16**, the Velarde et al calculated results yield a very good performance, essentially overlaying the trend of GRACE non-parametric correlation, except for the main outlier points (which is an issue for all of the correlations in this work). For reference, the *AARE* for the Velarde et al model is 5.54 percent (which is the highest for all of the correlations tested in this work), compared to the GRACE non-parametric correlation which has an *AARE* of 3.41 percent. In addition, for comparison, the *AARE* for the Petrosky and Farshad model is 4.97 percent.

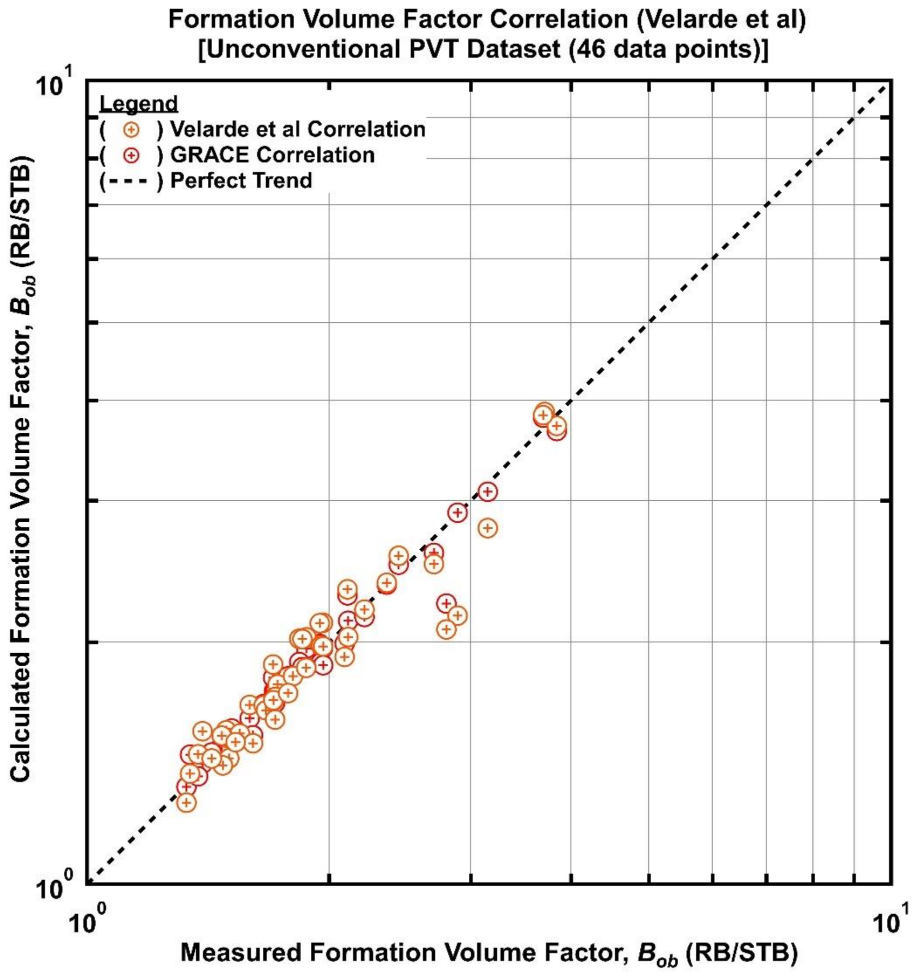

Figure 4.16 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Velarde et al model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.17 — Statistics: Oil FVF (B_{ob}) — Velarde et al model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.306
Standard Deviation, RB/STB	0.601
Variance, $(RB/STB)^2$	0.361
Average Absolute Relative Error, %	5.543

Dindoruk and Christman Correlation for Oil Formation Volume Factor

Dindoruk and Christman (2004) provided the most "complex" correlation to date for the oil formation volume factor, complete with a power-law function as an internal variable. The authors proposed this formulation as a "best fit" empirical model. The re-fitted form of the Dindoruk and Christman correlation for B_{ob} is given as:

$$B_{ob} = -5.97\times10^{-2} - 9.75\times10^{-4} A + 8.43\times10^{-7} A^2 + 1.14\times10^{-4}(T-60)\frac{API}{\gamma_g} \quad \text{.................(4.32)}$$

where the "A" parameter is given as $f(T, R_{sb}, \gamma_o, \gamma_g)$ in the following form:

$$A = \frac{\left[\dfrac{R_{sb}^{1.686}\gamma_g^{1.35262}}{\gamma_o^{-4.0761}} + 8.145\times10^{5}(T-60)^{-0.1735} - 2.7079R_{sb}\right]^{0.4695}}{\left[0.47716 + \dfrac{2R_{sb}^{-1.69497}}{\gamma_g^{4.9247}}(T-60)\right]^{2}} \quad \text{.........................(4.33)}$$

The results generated using Eqs. 4.32 and 4.33 are presented in **Fig 4.17** and the statistical results are given in **Table 4.18**. The calculated data show an excellent clustering around the perfect correlation trend in **Fig. 4.10** and the spurious calculated points are in better agreement with the dominant trend than for any other correlation — the *AARE* for the Dindoruk and Christman model is 4.30 percent, which is the best performance against the GRACE non-parametric correlation which has an *AARE* of 3.41 percent.

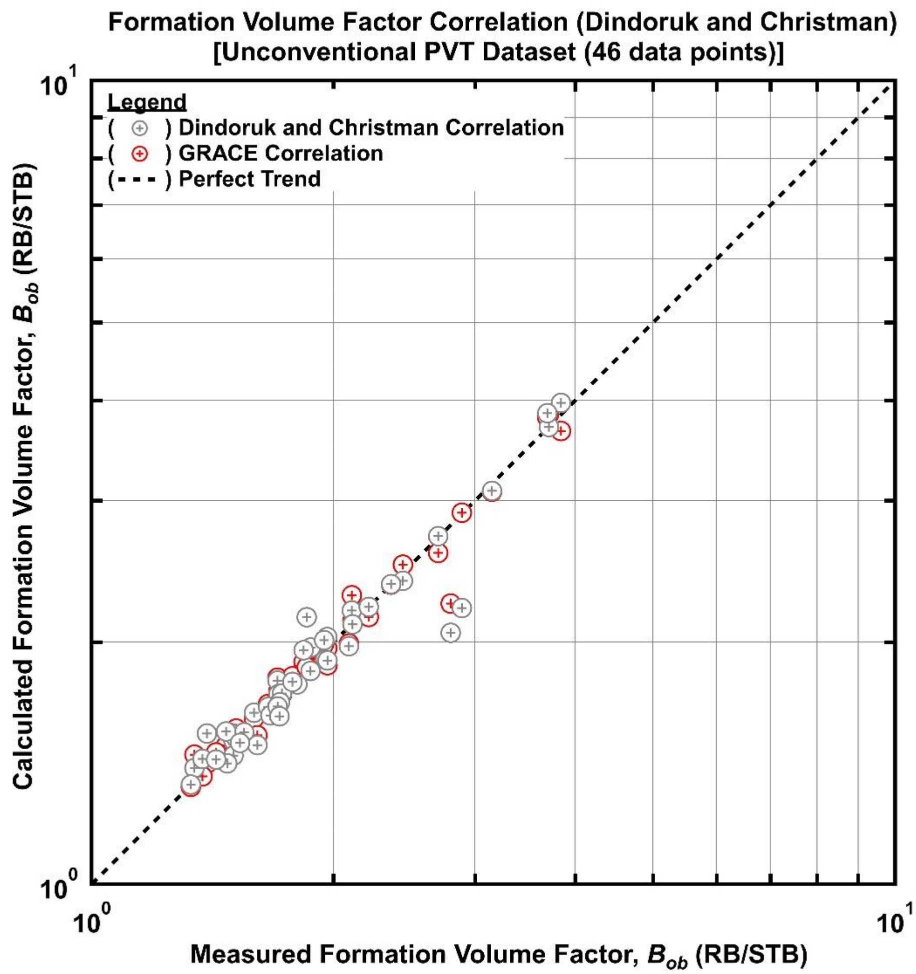

Figure 4.17 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); Dindoruk and Christman model and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 4.18 — Statistics: Oil FVF (B_{ob}) — Dindoruk and Christman model.

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.245
Standard Deviation, RB/STB	0.620
Variance, $(RB/STB)^2$	0.384
Average Absolute Relative Error, %	4.295

For this exercise, it is again noted that there are only 46 datasets available for the correlation of the oil formation volume factor, B_o — and again, the limited dataset is a recognized constraint for this correlation work. A complete summary of the statistical results for all of the B_o correlations is provided in **Table 4.19**, where it is noted that the averaged AARE for all correlations is 5.15 percent, and the *AARE* for the GRACE non-parametric correlation is 3.41 percent. The best performance for all of the correlations was for the Dindoruk and Christman model, which has an *AARE* of 4.30 percent. As final comment, the oil formation volume factor (B_o) is well-correlated in this exercise — all of the historical models we re-fitted well and, despite the small dataset size, the engineer should have reasonable confidence in using these relations for practical applications in reservoir and production engineering.

Table 4.19 — Statistics: Oil Formation Volume Factor, B_{ob} (Various Correlations).

Source	Standard Deviation, RB/STB	Variance, $(RB/STB)^2$	Average Absolute Relative Error, % (this work)
Standing (1947)	0.610	0.363	5.29
Vasquez and Beggs (1980)	0.585	0.343	5.52
Glaso (1980)	0.607	0.368	5.12
Al-Marhoun (1988)	0.608	0.370	5.33
Kartoatmodjo and Schmidt (1994)	0.624	0.389	5.09
Velarde, Blasingame, and McCain (1997)	0.601	0.361	5.54
Petrosky and Farshad (1998)	0.603	0.372	4.97
Dindoruk and Christman (2004)	0.620	0.384	4.30
Non-Parametric Solution (GRACE)	0.622	0.387	3.41

CHAPTER V

DEVELOPMENT OF NEW CORRELATIONS

The purpose of this chapter is to provide the methodology used to develop new correlations for the prediction of saturation pressure (p_{sat}) and the oil formation volume factor (B_o). Given the relatively good performance of the historical B_o correlations used in **Chapter IV**, there is little need for a "new" correlation (1 new correlation is proposed). Therefore, most of the effort in this chapter will be devoted to the development of a strong family of very flexible multivariate models for correlating the saturation pressure (p_{sat}) (6 new correlation models are proposed).

These proposed models are logarithmic in both the dependent and independent variables, and use linear, quadratic, and rational polynomial expansions of the logarithms of the independent variables. Generic expressions of these formulations are given below:

Linear Form:

$$\ln(y) = [c_1 + c_2 \ln(x_1)] \times [c_3 + c_4 \ln(x_2)] \times [c_5 + c_6 \ln(x_3)] \times \ldots$$

Quadratic Form:

$$\ln(y) = [c_1 + c_2 \ln(x_1) + c_3 \ln(x_1)^2] \times [c_4 + c_5 \ln(x_2) + c_6 \ln(x_2)^2] \times \ldots$$

Rational Polynomial Form:

$$\ln(y) = \frac{[c_1 + c_2 \ln(x_1) + c_3 \ln(x_1)^2] \times [c_4 + c_5 \ln(x_2) + c_6 \ln(x_2)^2] \times \ldots}{[d_1 + d_2 \ln(x_1) + d_3 \ln(x_1)^2] \times [d_4 + d_5 \ln(x_2) + d_6 \ln(x_2)^2] \times \ldots}$$

Obviously the "*y*-variable" will be either the saturation pressure (p_{sat}) and the oil formation volume factor (B_o). The "*x*-variables" are reservoir temperature (T), stock tank oil gravity (API), separator gas gravity (γ_g [or SG]), and solution gas-oil-ratio (R_{sb}).

In this work, these new models are referred to as "exponential" because the ln(*y*) function must be exponentiated by the type of polynomial (linear, quadratic, or rational). The correlation model will also be identified by its number of coefficients, and it is acknowledged that relations with more coefficients generally provide more complex relationships between the independent variables, but the complexity of some relationships may actually "over-fit" the data.

The mechanism for fitting these equations to data is much the same as the process used in Chapter IV — non-linear regression methods implemented using modules in Microsoft Excel and/or Python algorithms. For Microsoft Excel, the "Solver" tool was used exclusively, and in particular, the "Generalized Reduced Gradient" option. Microsoft Excel was used as the "first pass" tool, after which a working correlation programed into Python and various algorithms/libraries were deployed — in particular, the "numpy", "matplotlib", and "scipy. optimize" libraries were used to validate and/or enhance the results obtained using Microsoft Excel.

As was also performed in **Chapter IV**, the Xue et al (1997) implementation of Alternating Conditional Expectations (or ACE) algorithm proposed by Breiman and Friedman (1985) provides a non-parametric correlation of the data, Recall that the Xue et al implementation is known as "GRACE" and that any proposed correlation equation which has lower error statistics than the GRACE "correlation" should be suspected of "over-fitting."

5.2 Correlations of the Saturation Pressure (p_{sat})

As noted in the section above, the correlation of the saturation pressure (p_{sat}) will use the newly defined models. The models are summarized in **Table 5.1**.

Table 5.1 — Models: Saturation Pressure (p_{sat}) — $p_{sat} = f(R_s, \gamma_g$ (or SG), API, T).

Figure	Variable	Model	Number of Coefficients
Fig. 5.1	p_{sat}	Exponential-Linear Polynomial	8
Fig. 5.2	p_{sat}	Exponential-Linear Polynomial	16
Fig. 5.3	p_{sat}	Exponential-Quadratic Polynomial	12
Fig. 5.4	p_{sat}	Exponential-Rational Polynomial	8
Fig. 5.5	p_{sat}	Exponential-Rational Polynomial	16
Fig. 5.6	p_{sat}	Exponential-Rational Polynomial	10

The results of each proposed correlation are presented in **Figs. 5.1-5.6** and statistical results are presented in **Tables 5.2-5.7.** For the database used in this work, the GRACE algorithm method yielded an average absolute relative error of 13.17 percent, which will serve as the basis for comparison with each of the p_{sat} model correlations.

Case 1 — Saturation Pressure — Exponential-Linear Polynomial (8-coefficients)

The first model deployed is the Exponential-Linear Polynomial (8-coefficient) model. The general form of the model is given as:

$$\ln p_{sat} = [c_1 + c_2 \ln T] \times [c_3 + c_4 \ln API] \times [c_5 + c_6 \ln R_s] \times [c_7 + c_8 \ln \gamma_g] \quad \text{....................(5.1a)}$$

Fitting Eq. 5.1a to the data in the project database yields:

$$\begin{aligned} \ln p_{sat} = \; & [-0.1522 + 7.584\times10^{-2} \ln T] \times [-5.719 + 17.371 \ln API] \\ & \times [3.508\times10^{2} + 1.216 \ln R_{sb}] \times [-56.095 - 2.931 \ln \gamma_g] \end{aligned} \quad \text{...............................(5.1b)}$$

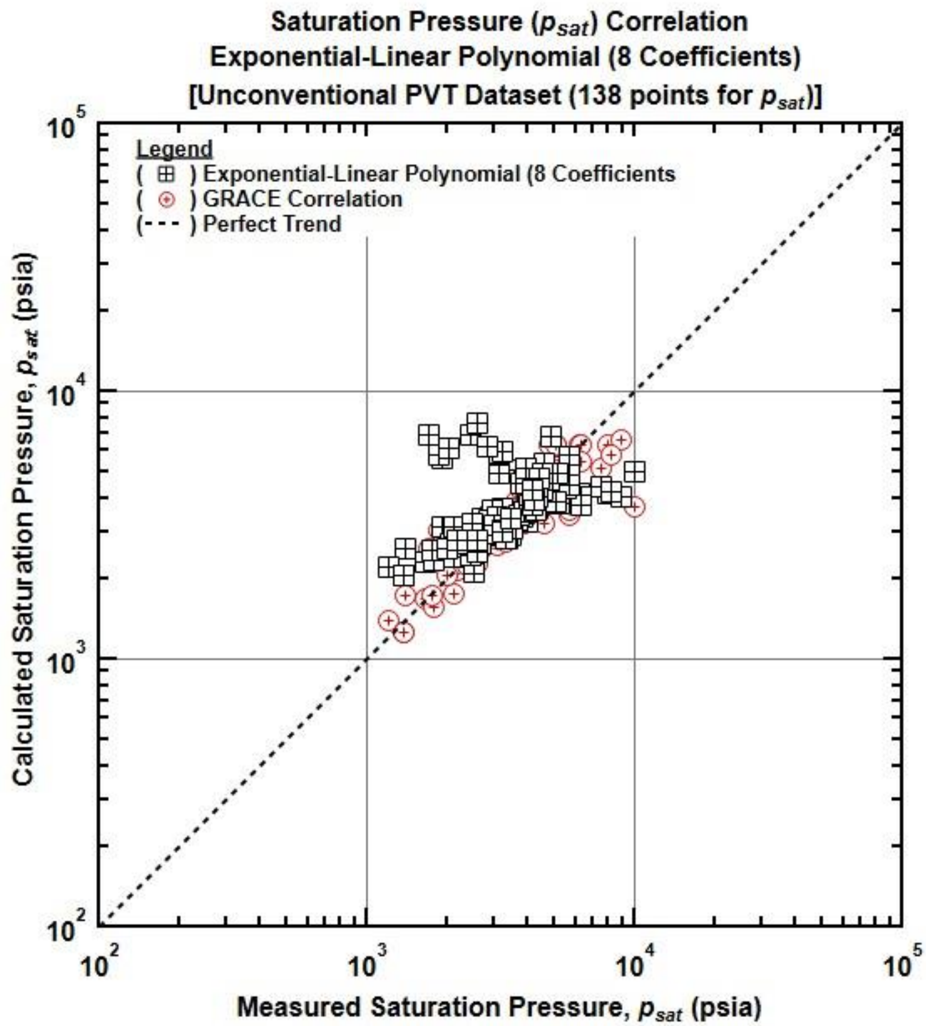

Figure 5.1 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); exponential-linear polynomial model (8 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 5.2 — Saturation Pressure (p_{sat}) — Exponential-Linear Model (8 coefficients).

Parameter	Value
Sum of Squared Residuals, (psia)^2	17.633
Standard Deviation, psia	1,109.06
Variance, (psia)^2	1,230,013
Average Absolute Relative Error, %	21.96

This model is the weakest of all the general models, having an *AARE* of 22.96 percent (from **Table 5.2**). The "hook" feature observed in **Fig. 5.1** is due to the retrograde gas condensate data samples, and this performance is very similar to the historical correlation models for p_{sat}.

Case 2 — Saturation Pressure — Exponential-Linear Polynomial (16-coefficients)

The second model deployed is the Exponential-Linear Polynomial (16-coefficient) model. The general form of the model is given as:

$$\begin{aligned}
\ln p_{sat} = \; & c_1 \\
& + c_2 \ln T + \\
& + c_3 \ln API + \\
& + c_4 \ln R_{sb} + c_5 \ln \gamma_g \\
& + c_6 \ln T \ln API \\
& + c_7 \ln T \ln R_{sb} \\
& + c_8 \ln T \ln \gamma_g \\
& + c_9 \ln API \ln R_{sb} \\
& + c_{10} \ln API \ln \gamma_g \\
& + c_{11} \ln \gamma_g \ \ln R_{sb} \\
& + c_{12} \ln T \ln API \ln R_{sb} \\
& + c_{13} \ln T \ln API \ln \gamma_g \\
& + c_{14} \ln T \ln \gamma_g \ \ln R_{sb} \\
& + c_{15} \ln API \ln \gamma_g \ \ln R_{sb} \\
& + c_{16} \ln T \ln API \ln R_{sb} \ln \gamma_g
\end{aligned} \quad \text{(5.2a)}$$

Eq. 5.2 is written in this fully expanded form to highlight the nature of the inter-relations of the individual dependent variables (R_s, γ_g (or *SG*), *API*, *T*). This is the most complete "simple" expansion (no higher order terms) and should be thought of as "less likely" to over-fit the data.

Fitting Eq. 5.2a to the data in the project database yields:

$$\begin{aligned}
\ln p_{sat} = &- 0.158 \\
&+ 7.129\times10^{-2} \ln T \\
&- 6.371 \ln API \\
&+ 18.431 \ln R_{sb} \\
&+ 3.250\times10^{2} \ln \gamma_g \\
&+ 1.332 \ln T \ln API \\
&- 53.095 \ln T \ln R_{sb} \\
&- 3.084 \ln T \ln \gamma_g \\
&- 92.747 \ln API \ln R_{sb} \\
&- 3.573 \ln API \ln \gamma_g \\
&+ 1.038 \ln \gamma_g \ln R_{sb} \\
&+ 14.728 \ln T \ln API \ln R_{sb} \\
&+ 0.605 \ln T \ln API \ln \gamma_g \\
&- 0.597 \ln T \ln \gamma_g \ln R_{sb} \\
&+ 1.016 \ln API \ln \gamma_g \ln R_{sb} \\
&- 8.488\times10^{-3} \ln T \ln API \ln R_{sb} \ln \gamma_g
\end{aligned} \qquad \text{(5.2a)}$$

This model is (perhaps) the strongest of all the general models, having an *AARE* of 12.67 percent (from **Table 5.3**), which is the lowest of all models in this work. There is a possibility that Eq. 5.2a has slightly over-fitted the database — however; the behavior shown in **Fig. 5.2** is very "tight" and well-centered about the perfect correlation trend. There are a few minor outlying points observed in **Fig. 5.2**, but these should not be seen as significant relative the excellent overall performance of Eq. 5.2a.

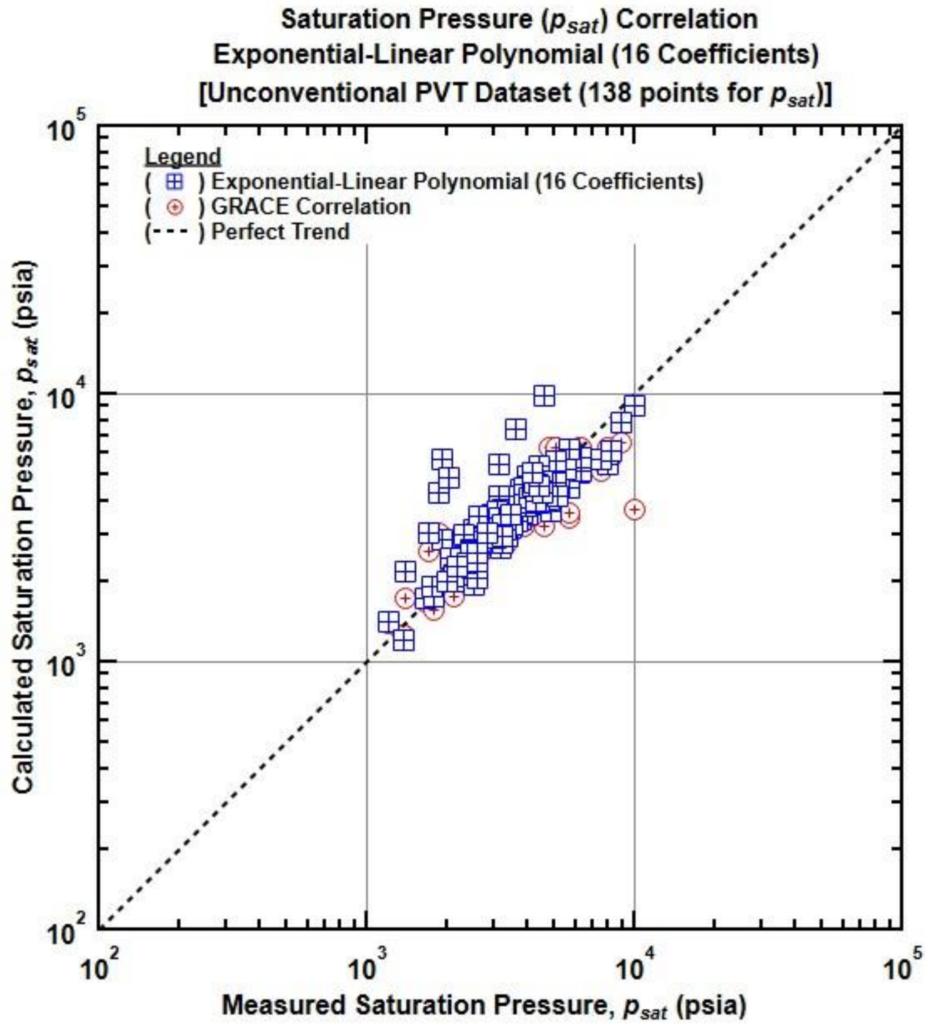

Figure 5.2 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); exponential-linear polynomial model (16 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 5.3 — Saturation Pressure (p_{sat}) — Exponential-Linear Model (16 coefficients).

Parameter	Value
Sum of Squared Residuals, $(\text{psia})^2$	6.825
Standard Deviation, psia	1,363.74
Variance, $(\text{psia})^2$	1,859,775
Average Absolute Relative Error, %	12.67

Case 3 — Saturation Pressure — Exponential-Quadratic Polynomial (12 coefficients)

The third model deployed is the Exponential- Quadratic Polynomial (12-coefficient) model. This model utilizes "squared" terms for each independent variable and the general form of the model is given by:

$$\begin{aligned} \ln p_{sat} = & \left[c_1 + c_2 \ln T + c_3 [\ln T]^2\right] \\ & \times \left[c_4 + c_5 \ln API + c_6 [\ln API]^2\right] \\ & \times \left[c_7 + c_8 \ln R_{sb} + c_9 [\ln R_{sb}]^2\right] \\ & \times \left[c_{10} + c_{11} \ln \gamma_g + c_{12} [\ln \gamma_g]^2\right] \end{aligned} \quad \text{..........(5.3a)}$$

Fitting Eq. 5.3a to the data in the project database yields:

$$\begin{aligned} \ln p_{sat} = & \left[6.021 - 1.398 \ln T + 0.125 [\ln T]^2\right] \\ & \times \left[3.434 - 0.136 \ln API + 1.325 \times 10^{-2} [\ln API]^2\right] \\ & \times \left[-1.021 + 0.725 \ln R_{sb} - 4.189 \times 10^{-2} [\ln R_{sb}]^2\right] \\ & \times \left[0.602 - 3.130 \times 10^{-2} \ln \gamma_g - 2.994 \times 10^{-3} [\ln \gamma_g]^2\right] \end{aligned} \quad \text{..........(5.3b)}$$

This model has an exceptional visual correlation as seen in **Fig. 5.3**, and with an *AARE* of 13.41 percent (from **Table 5.4**), it is very consistent with the GRACE correlation which has an *AARE* of 13.17 percent. While not "statistical," a visual comparison of the results of Eq. 5.3b and the GRACE correlation is the best of all cases considered in this work — particularly in the sense that almost no "outliers" are observed in **Fig. 5.3.**

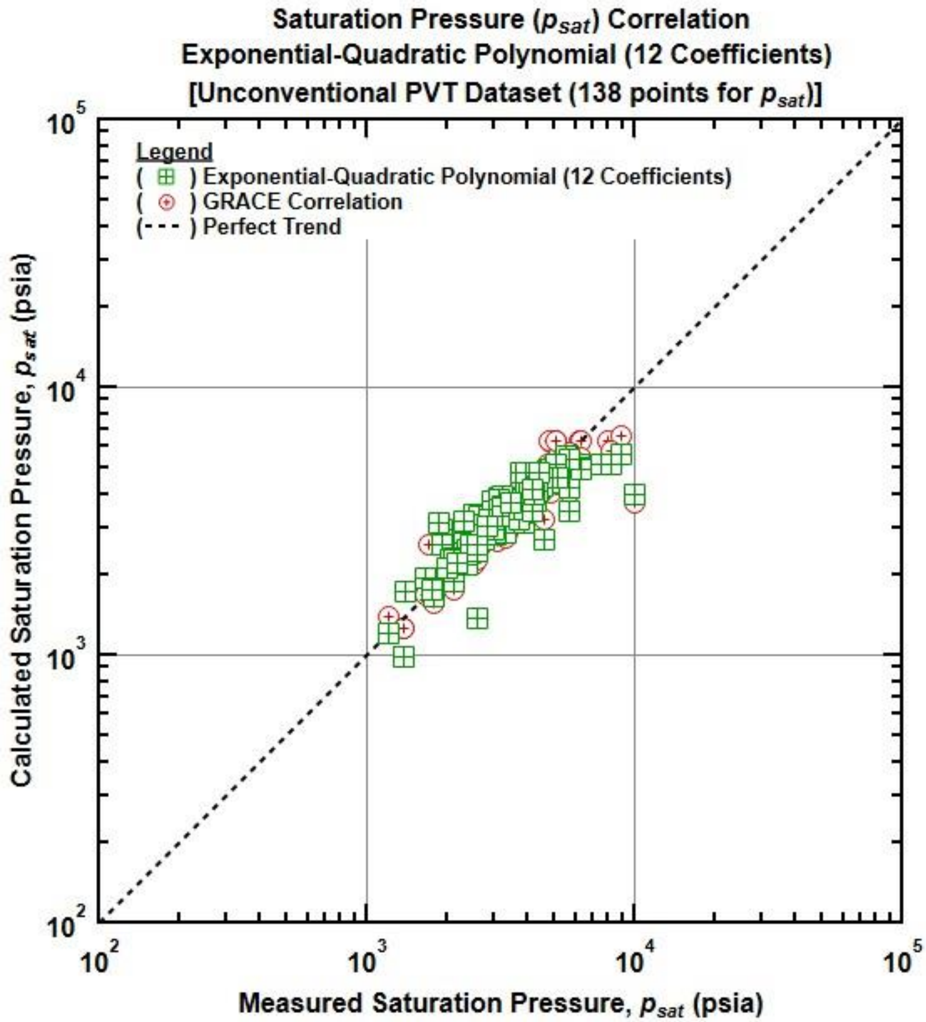

Figure 5.3 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); exponential-quadratic polynomial model (12 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 5.4 — Saturation Pressure (p_{sat}) — Exponential-Quadratic (12 coefficients).

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	5.031
Standard Deviation, psia	1,073.78
Variance, $(psia)^2$	1,153,004
Average Absolute Relative Error, %	13.41

Case 4 — Saturation Pressure — Exponential-Rational Polynomial (8 coefficients)

The fourth model deployed is the Exponential-Rational Polynomial (8-coefficient) model. This is the most simple of the Rational Polynomial models and is used to set a sort of basis for the other more complex Exponential-Rational Polynomial models. This form is given as:

$$\ln p_{sat} = \frac{[c_1 + c_2 \ln T]}{1 + [c_3 + c_4 \ln API] \times [c_5 + c_6 \ln R_{sb}] \times [c_7 + c_8 \ln \gamma_g]} \quad \text{.....................................(5.4a)}$$

Fitting Eq. 5.4a to the data in the project database yields:

$$\ln p_{sat} = \frac{[9.021 - 0.119 \ln T]}{1 + [2.221 - 0.531 \ln API] \times [0.144 - 1.842 \times 10^{-2} \ln R_{sb}] \times [12.802 + 8.309 \ln \gamma_g]} \quad (5.4b)$$

Eq. 5.4b has an *AARE* of 14.24 percent (from **Table 5.5**) and the results shown in **Fig. 5.4** suggest that despite having the lowest complexity (and number of coefficients) of any of the rational polynomial models, this formulation is quite credible and may be valued because of its relative simplicity compared to the (much) more complex models. In **Fig. 5.2** the observed correlation is very good, if not excellent, and there are but a few minor outlying points

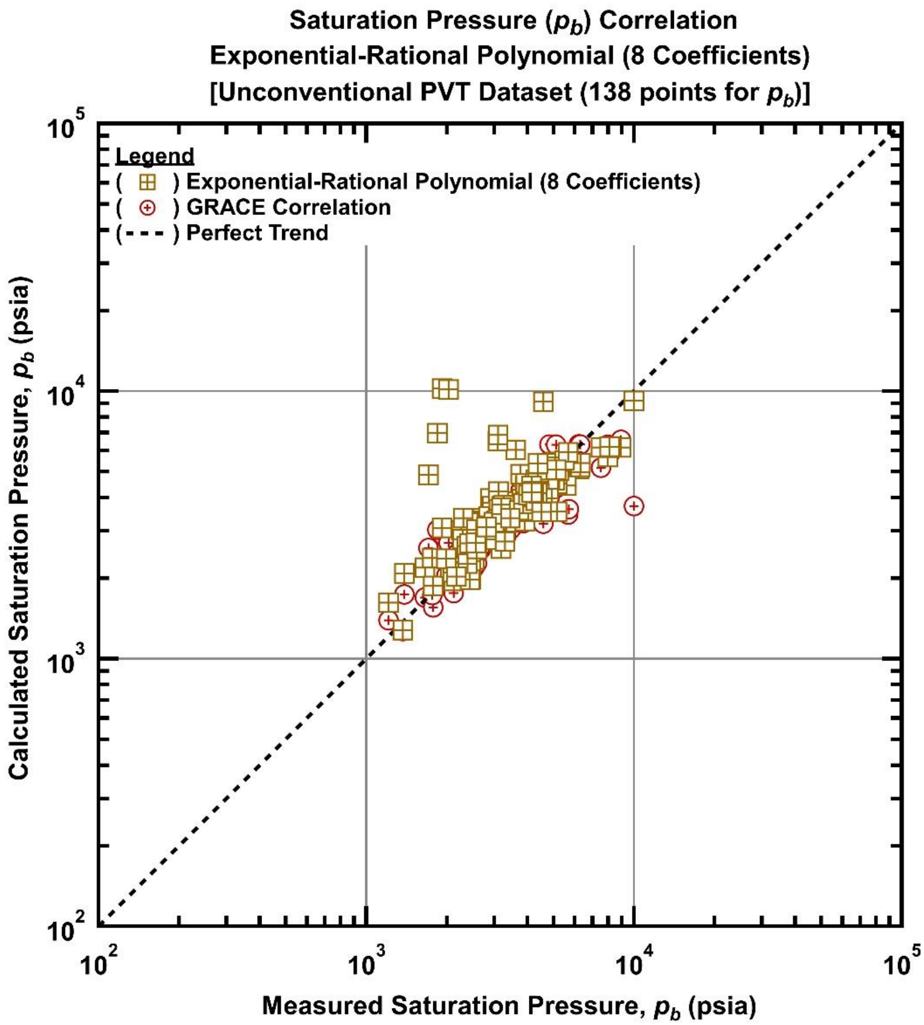

Figure 5.4 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); exponential rational polynomial model (8 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 5.5 — Saturation Pressure (p_{sat}) — Exponential-Rational Polynomial (8 coefficients).

Parameter	Value
Sum of Squared Residuals, (psia)^2	13.25
Standard Deviation, psia	1,534.63
Variance, (psia)^2	2,355,080
Average Absolute Relative Error, %	14.24

Case 5 — Saturation Pressure — Exponential-Rational Polynomial (16 coefficients)

The fifth model is the Exponential-Rational Polynomial (16-coefficient) model, which is arguably the most complex of all the models conceived in this work (the "rational" formulation gives extraordinary flexibility to this model). This form is given as:

$$\ln p_{sat} = \frac{[c_1 + c_2 \ln T] \times [c_3 + c_4 \ln API] \times [c_5 + c_6 \ln R_{sb}] \times [c_7 + c_8 \ln \gamma_g]}{[c_9 + c_{10} \ln T] \times [c_{11} + c_{12} \ln API] \times [c_{13} + c_{14} \ln R_{sb}] \times [c_{15} + c_{16} \ln \gamma_g]} \quad \text{........(5.5a)}$$

Fitting Eq. 5.5a to the data in the project database yields:

$$\ln p_{sat} = \frac{\begin{array}{l} [0.858 - 7.881\times10^{-2} \ln T] \\ \times\ [3.198 - 0.457 \ln API] \\ \times\ [0.146 + 0.322 \ln R_{sb}] \\ \times\ [3.172 + 1.015 \ln \gamma_g] \end{array}}{\begin{array}{l} [-0.340 + 0.540 \ln T] \\ \times\ [-0.665 + 0.458 \ln API] \\ \times\ [-0.545 + 3.343\times10^{-2} \ln R_{sb}] \\ \times\ [0.454 - 0.281 \ln \gamma_g] \end{array}} \quad \text{........(5.5b)}$$

Eq. 5.5b has an *AARE* of 12.75 percent (from **Table 5.6**), which is very close to the result from the Exponential-Linear model (16 coefficients) (Eq. 5.3b) which has an *AARE* of 12.67. While these models have a different mathematical basis, the similarity in performance may be due to the high number of coefficients (16) for both models. In addition, the visual comparison of results in results shown in **Fig. 5.5** are very well correlated with the non-parametric results (GRACE) as well as the perfect correlation trend (*i.e.*, the 45 degree line). Finally, the "tightness" of the fit of Eq. 5.5b is excellent, with only 8 points having significant "outlier" deviation.

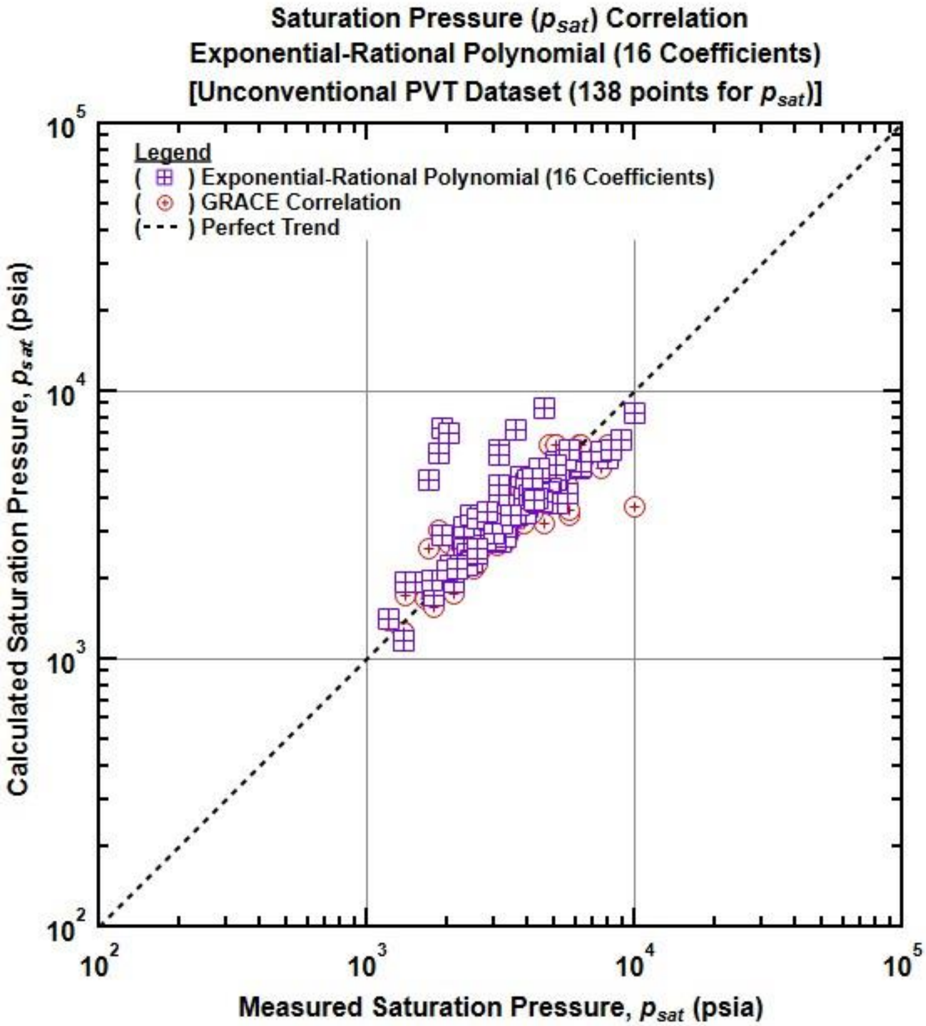

Figure 5.5 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); exponential rational polynomial model (16 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 5.6 — Saturation Pressure (p_{sat}) — Exponential-Rational Polynomial (16 coefficients).

Parameter	Value
Sum of Squared Residuals, $(psia)^2$	9.852
Standard Deviation, psia	1,357.11
Variance, $(psia)^2$	1,841,747
Average Absolute Relative Error, %	12.75

Case 6 — Saturation Pressure — Exponential-Rational Polynomial (10 coefficients)

The sixth model is the Exponential-Rational Polynomial (10-coefficient) model. This last model is actually a bit of a hybrid, designed to have as few coefficients as necessary to provide a certain level of performance. This model is given by:

$$\ln p_{sat} = \frac{[c_1 + c_2 \ln T]}{\begin{bmatrix} c_3 + c_4 \ln \gamma_g + c_5 \ln API + c_6 \ln R_{sb} + c_7 \ln R_{sb} \ln \gamma_g + \\ c_8 \ln API \ln \gamma_g + c_9 \ln API \ln R_{sb} + c_{10} \ln API \ln R_{sb} \ln \gamma_g \end{bmatrix}} \quad \text{.............................(5.6a)}$$

Where again, Eq. 5.6a is designed to be "compact," particularly in the denominator where the terms have been expanded and collected. Fitting Eq. 5.6a to the data in the project database yields:

$$\ln p_{sat} = \frac{[44.109 - 0.967 \ln T]}{\begin{bmatrix} 15.591 - 5.783 \ln \gamma_g - 1.772 \ln API - 0.687 \ln R_s - 2.098 \ln R_s \ln \gamma_g + \\ 3.177 \ln API \ln \gamma_g + 0.366 \ln API \ln R_s - 4.335 \times 10^{-2} \ln API \ln R_s \ln \gamma_g \end{bmatrix}} \quad \text{.........(5.6b)}$$

Eq. 5.6b has an *AARE* of 13.47 percent (from **Table 5.7**), but much more intriguing is the visual similarity of the performance of Eq. 5.6b shown on **Fig. 5.6** and that of Eq. 5.5b shown on **Fig. 5.5**. These relations are exceptionally similar, despite Eq. 5.6b having 6 less coefficients that Eq. 5.5b. Of course, Eq. 5.5b provides a better statistical fit (an *AARE* of 12.75 percent from **Table 5.7**), but compared visually on **Figs. 5.5 and 5.6**, the performance of Eqs. 5.5b and 5.6b are almost indistinguishable (the outlier points for each model are slightly different, but the body of the trends are essentially the same).

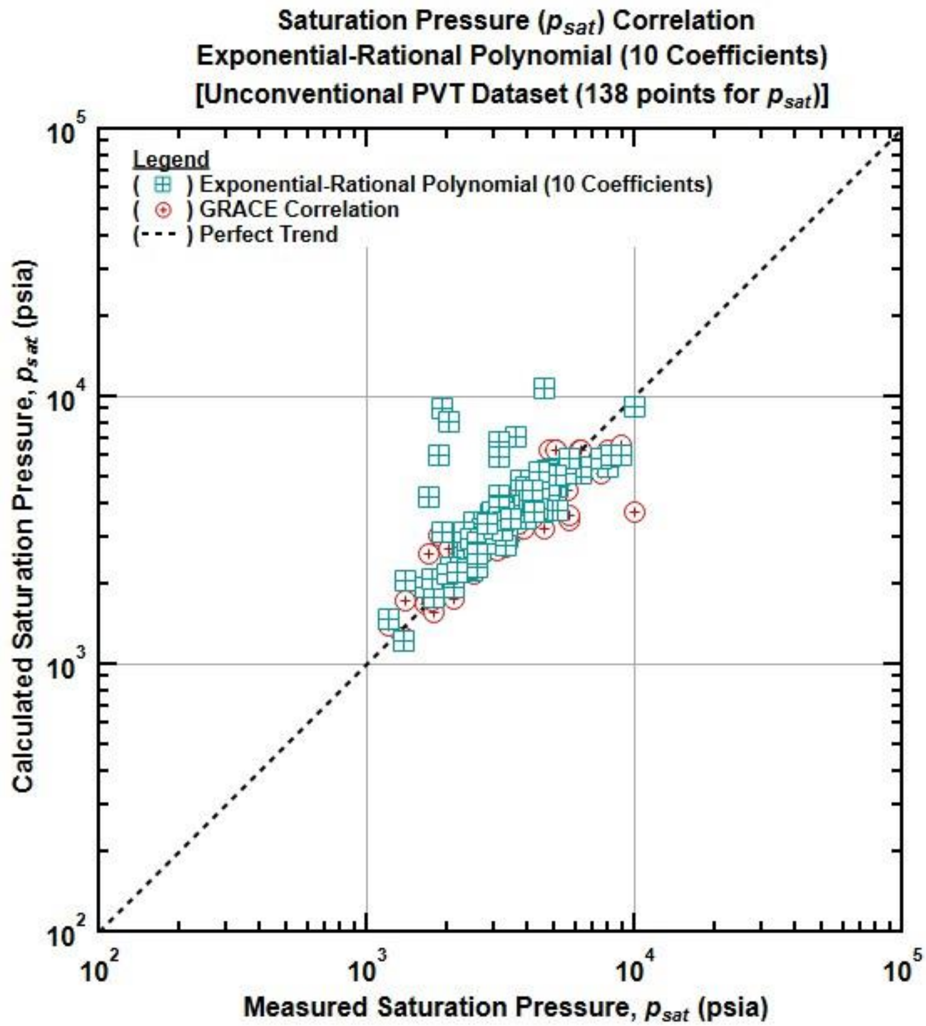

Figure 5.6 — Log-log plot of calculated versus measured saturation pressure (p_{sat}); exponential rational polynomial model (10 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 138 data points).

Table 5.7 — Saturation Pressure (p_{sat}) — Exponential-Rational Polynomial (10 coefficients).

Parameter	Value
Sum of Squared Residuals, (psia)^2	10.966
Standard Deviation, psia	1,474.72
Variance, (psia)^2	2,174,804
Average Absolute Relative Error, %	13.47

Summary — Comparison of Correlation Results for Saturation Pressure

The results for all p_{sat} correlation cases developed in this work are shown in **Table 5.8**.

Table 5.8 — Statistics: Saturation Pressure (p_{sat}) — Historical and New Correlations

Source	Standard Deviation, psia	Variance, $(psia)^2$	Average Absolute Relative Error, % (this work)
Standing	2,431	5,907,833	20.0
Petrosky and Farshad	1,057	1,117,756	21.1
Lasater	957	915,193	21.3
Vasquez and Beggs	2,543	6,465,340	20.3
Glaso	2,023	4,092,484	20.0
Al-Marhoun	2,432	5,915,237	24.2
Kartoatmodjo and Schmidt	3,337	11,132,681	21.1
Velarde et al	664	440,298	27.3
Dindoruk and Christman	2,107	4,437,844	20.3
Exponential-Linear (8 coefficients)	1,109	1,230,013	22.0
Exponential-Linear (16 coefficients)	1,364	1,859,755	12.7
Exponential-Quadratic (12 coefficients)	1,078	1,153,004	13.4
Exponential-Rational (8 coefficients)	1,535	2,355,080	14.2
Exponential-Rational (16 coefficients)	1,357	1,841,747	12.8
Exponential-Rational (10 coefficients)	1,475	2,174,804	13.5
Non-Parametric Solution (GRACE)	1,197	1,432,965	13.2

The results provided in **Table 5.8** confirm that all of the historical correlations re-fitted to the database for this work do not achieve sufficient accuracy to warrant application. These re-fitted historical relations have an averaged *AARE* of 21.0 percent (ignoring the Velarde et al result), compared to an *AARE* of 13.2 percent for the non-parametric (GRACE) correlation, which is thought to be the best result that can be achieved without "over-fitting" the data.

Further, the new correlations developed in this work have an averaged *AARE* of 13.5 percent (ignoring the Exponential-Linear 8-coefficient case), compared to the GRACE correlation which had an *AARE* of 13.2 percent. This performance suggests that the new correlations are approaching the accuracy of the non-parametric correlations. In fact, the Exponential-Linear 16-coefficient model and the Exponential-Rational 16-coefficient model both slightly out-performed the GRACE correlation in a statistical sense.

These results validate, at least in concept, the use of very "flexible" empirical models for the correlation and prediction of the saturation pressure (p_{sat}) for cases of unconventional reservoir fluids including black oils, highly volatile oils, and retrograde gas condensate fluids.

5.3 Correlation of the Oil Formation Volume Factor (B_{ob})

First and foremost, it must be noted that there are only 46 datasets available in this work for the correlation of the Oil Formation Volume Factor (B_{ob}) (see **Appendix A**). Such a small dataset for unconventional reservoir fluids makes this work somewhat "theoretical" (as opposed to being practical) as a limited dataset should not be "extrapolated" to suggest that it represents universal behavior. Ironically, Standing (1947) achieved exceptional results in his early PVT correlation work with a dataset that was only about twice the size of the one available in this work.

As most of the historical correlations were re-fitted quite well to the project database in **Chapter IV**, only a single, "new" correlation will be presented, and that correlation will be analogous to the models presented for the correlation of the saturation pressure. Specifically, the "Exponential-Quadratic" polynomial model will be used where the oil formation volume factor is correlated against all of the independent variables in the dataset [*i.e.*, $B_{ob} = f(p_{sat}, R_{sb}, \gamma_g$ (or SG), API, T)].

The Exponential-Quadratic Polynomial (15-coefficient) model used to correlate B_{ob} is given as:

$$\begin{aligned}\ln B_{ob} = \quad & \left[c_1 + c_2 \ln T + c_3 \left[\ln T\right]^2\right] \\ \times & \left[c_4 + c_5 \ln API + c_6 \left[\ln API\right]^2\right] \\ \times & \left[c_7 + c_8 \ln R_s + c_9 \left[\ln R_{sb}\right]^2\right] \\ \times & \left[c_{10} + c_{11} \ln \gamma_g + c_{12} \left[\ln \gamma_g\right]^2\right] \\ \times & \left[c_{13} + c_{14} \ln p_b + c_{15} \left[\ln p_{sat}\right]^2\right]\end{aligned} \qquad \text{(5.7a)}$$

Fitting Eq. 5.7a to the data in the project database yields:

$$\begin{aligned}\ln B_{ob} = \quad & \left[1.550 - 0.510 \ln T + 1.194 \times 10^{-3} \left[\ln T\right]^2\right] \\ \times & \left[-0.287 - 0.682 \ln API - 5.456 \times 10^{-2} \left[\ln API\right]^2\right] \\ \times & \left[0.360 - 0.121 \ln R_s + 1.126 \times 10^{-2} \left[\ln R_{sb}\right]^2\right] \\ \times & \left[0.414 - 1.094 \times 10^{-2} \ln \gamma_g - 0.101 \left[\ln \gamma_g\right]^2\right] \\ \times & \left[0.770 + 0.521 \ln p_b + 1.287 \times 10^{-3} \left[\ln p_{sat}\right]^2\right]\end{aligned} \qquad \text{(5.7b)}$$

The results for Eq. 5.7b are shown **Fig.5.7**, and with the exception of 2 exaggerated outlying (calculated) data points, the visual correlation of Eq. 5.7b is excellent. Eq. 5.7b has an *AARE* of 5.02 percent (from **Table 5.9**), compared to the GRACE correlation for B_{ob}, which has an *AARE* of 3.41 percent (from **Table 5.10**).

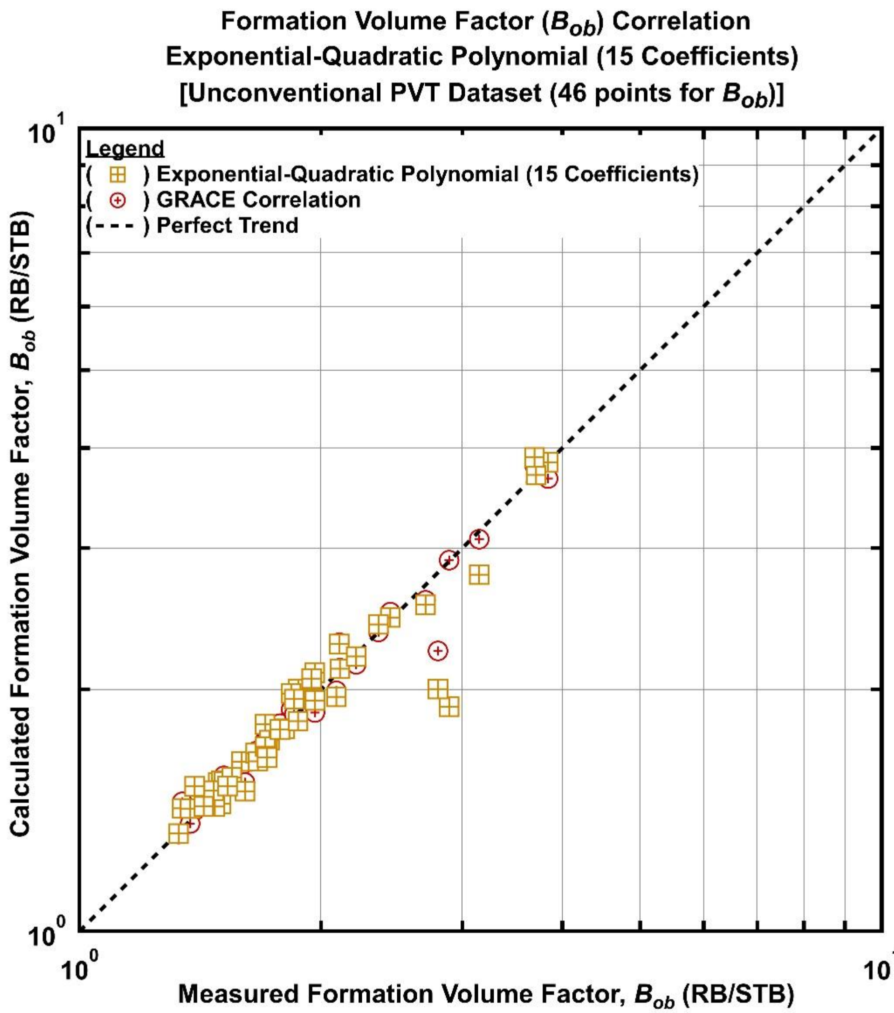

Figure 5.7 — Log-log plot of calculated versus measured oil formation volume factor (B_{ob}); exponential-quadratic polynomial model (15 coefficients) and the GRACE correlation (unconventional reservoir fluid data set — 46 data points).

Table 5.9 — Statistics: Oil FVF (B_{ob}) — Exponential-Quadratic Polynomial (15 coefficients).

Parameter	Value
Sum of Squared Residuals, $(RB/STB)^2$	0.368
Standard Deviation, RB/STB	0.598
Variance, $(RB/STB)^2$	0.358
Average Absolute Relative Error, %	5.02

Summary — Comparison of Correlation Results for Oil Formation Volume Factor

The results for all B_{ob} correlation cases developed in this work are shown in **Table 5.10**.

Table 5.10 — Statistics: Oil FVF (B_{ob}) — Historical and New Correlation

Source	Standard Deviation, RB/STB	Variance, $(RB/STB)^2$	Average Absolute Relative Error, % (this work)
Standing	0.603	0.372	4.97
Petrosky and Farshad	0.610	0.363	5.29
Vasquez and Beggs	0.585	0.343	5.52
Glaso	0.607	0.368	5.12
Al-Marhoun	0.608	0.370	5.33
Kartoatmodjo and Schmidt	0.624	0.389	5.09
Velarde et al	0.601	0.361	5.54
Dindoruk and Christman	0.620	0.384	4.30
This Study Exponential-Quadratic (15 coef.)	0.598	0.358	5.02
Non-Parametric Solution (GRACE)	0.622	0.387	3.41

The results provided in **Table 5.10** show remarkable consistency in the re-fitted historical correlations for B_{ob} — these results have an averaged *AARE* of 5.15 percent and the best fit historical model is the Dindoruk and Christman result, which has an *AARE* of 4.30 percent. Recalling, the GRACE correlation has an *AARE* of 3.41 percent and the Exponential-Quadratic Polynomial 15-coefficient model has an *AARE* of 5.02 percent.

Restating the caveat that these results are based on only 46 datasets, this work suggests that a prediction of the oil formation volume factor using these data and results should yield acceptable engineering accuracy. However, it must be noted that additional data must be incorporated into such correlations before a strong case for widespread application can be made.

CHAPTER VI

CONCLUSIONS AND RECOMMENDATIONS

6.1 Summary

The primary goal of this work is to present new empirical correlations for the saturation pressure and the oil formation volume factor for black oil, volatile oil, and retrograde condensate reservoir fluids obtained from unconventional (shale) reservoirs. Corporate donors provided the majority of the laboratory PVT data for this study, which covered shale plays in the United States and Latin America. All of these data are provided in **Appendix A**.

As part of a confidentiality agreement with the donor companies the data may be shared publicly, but must remain completely anonymous as a requirement of confidentiality (*i.e.*, no other information regarding the donated data may be released). A total of 138 data cases were used in this work (about 20 of the submissions were discarded as incomplete, inconsistent and/or erroneous). The data analysis in this work was performed using MS Excel and its built-in tools (*e.g.*, "Solver"), as well as various Python codes and Python library modules.

While the goal of this work may seem simple enough — developing accurate correlations that have high fidelity (*i.e.*, the vast majority of the data are matched), this proved to be a very challenging effort. Essentially all of the historical saturation pressure correlations tested on the unconventional reservoir fluids dataset failed in either accuracy or fidelity, or both. This is somewhat expected as *all of the prior work in this area focused on correlations applied to black oil or slightly volatile oil cases* — in this work there are numerous (very) volatile oil cases, as well as many retrograde gas condensate cases.

As a test of the fidelity of the data, non-parametric regression analysis was used to establish baseline statistics — for reference, non-parametric regression should always yield the lowest error metrics, and any regression that achieves lower error statistics are deemed to have been "over-fitted." The results of the non-parametric regression were plotted on all "calculated versus measured" correlation plots along with a "perfect correlation trend" (45-degree trendline) for reference with the results from the historical correlations and the new correlations proposed in this work.

Of the historical correlations for both the saturation pressure (p_{sat}) (based on 138 data points) and the oil formation volume factor (B_{ob}) (based on 46 data points), the Dindoruk and Christman (2000) correlation models yielded the best results (statistically) and visually. Several new, generalized models using an exponential transform and polynomial parametric formulations (*i.e.*, the "exponential-polynomial" models) were applied and validated for both p_{sat} and B_{ob}, and in general these "exponential-polynomial" models yielded much better statistical results than the historical correlations. Further, the proposed "exponential-polynomial" models also captured the behavior of the highly volatile oil cases as well as the retrograde gas condensate cases, where in contrast none of the historical correlation models performed well for these cases.

6.2 Conclusions

The following key observations/deliverables are the noted from this work:

- *Historical Correlations — Saturation Pressure* (p_{sat}): Historical correlations for the saturation pressure (p_{sat}) applied to the project dataset (138 points) for fluids from unconventional reservoir were <u>not</u> able to yield statistical results (**Chapter IV**) sufficient to warrant the use of <u>any</u> historical correlation for the prediction of saturation pressure (p_{sat}) for practical applications in reservoir and production engineering (all correlations yielded an *AARE* of 20 percent or higher [**Chapter IV**], where the non-parametric regression gave an *AARE* of 13.17 percent, [**Chapter V**]). Further, visual evidence from the "calculated versus measured" plots clearly indicates that the historical correlations for saturation pressure (p_{sat}) — all of which were developed for "black oils", cannot capture the phase behavior of highly volatile oils nor retrograde gas condensate systems.

- *Historical Correlations — Oil Formation Volume Factor* (B_{ob}): In the case of the oil formation volume factor (B_{ob}), the project database contained only 46 useable datasets. Using these data to re-fit the historical correlations for the oil formation volume factor (B_{ob}) was actually a quite successful exercise, with almost all historical correlations performing in an acceptable manner, and for the correlation of Dindoruk and Christman (**ChapterIV**), the results were very good (*AARE* = 4.3 percent). For comparison, the non-parametric regression gave an *AARE* of 3.41 percent for B_{ob} (**Chapter V**).
- *New Correlations — General*: In this work a "family" of "exponential polynomial" correlations were proposed where the expression is written in $\ln(y)$ (hence the "exponential" name as these expressions must be exponentiated) and the independent variables (also in logarithmic forms) are represented by linear, quadratic, and rational polynomials as illustrated below:

  Linear Form:

  $$\ln(y) = [c_1 + c_2 \ln(x_1)] \times [c_3 + c_4 \ln(x_2)] \times [c_5 + c_6 \ln(x_3)] \times \ldots$$

  Quadratic Form:

  $$\ln(y) = [c_1 + c_2 \ln(x_1) + c_3 \ln(x_1)^2] \times [c_4 + c_5 \ln(x_2) + c_6 \ln(x_2)^2] \times \ldots$$

  Rational Polynomial Form:

  $$\ln(y) = \frac{[c_1 + c_2 \ln(x_1) + c_3 \ln(x_1)^2] \times [c_4 + c_5 \ln(x_2) + c_6 \ln(x_2)^2] \times \ldots}{[d_1 + d_2 \ln(x_1) + d_3 \ln(x_1)^2] \times [d_4 + d_5 \ln(x_2) + d_6 \ln(x_2)^2] \times \ldots}$$

  And, of course, there can be variations of these forms based on the influence of a given independent variable (*e.g.*, temperature might be better represented by a linear, rather than quadratic function).
- *New Correlations — Saturation Pressure* (p_{sat}): The "exponential polynomial" correlations constructed for the saturation pressure (p_{sat}) were remarkably effective, with most correlations performing in the 13-14 percent *AARE* range (recall that the non-parametric result for p_{sat} had an *AARE* of 13.17 percent). The only poor performer was the 8

coefficient "exponential-linear" model which had an *AARE* of about 22 percent). As might be envisioned for relations with so much flexibility, the 16 coefficient relations had *AARE* values slightly less than the non-parametric relation, indicating the possibility of minor over-fitting. In short, these new models provided an exceptional to outstanding performance as correlators for the saturation pressure (p_{sat}) for reservoir fluids from unconventional reservoirs.

- *New Correlations — Oil Formation Volume Factor* (B_{ob}): As a reminder, there were only 46 useable datasets for the oil volume factor (B_{ob}), which led to the belief that a single "exponential-polynomial" model would be sufficient. In this case a 15-coefficient "exponential-quadratic" form was used, which gave an *AARE* of 5.2 percent, but yield an excellent visual correlation of the data (see **Fig. 5.7** in **Chapter V**). Recalling, the non-parametric regression gave an *AARE* of 3.4 percent for B_{ob}, and the Dindoruk and Christman correlation gave an *AARE* of 4.4 percent (**Chapter V**). Given this performance, all of these reasonable should all be considered reasonable (and realistic), despite the relatively small number of data (only 46 useable datasets).

6.3 Recommendations for Future Work

Based on the developments in this work, the following recommendations for future work are proposed for others to consider:

- *Database*: The database for unconventional reservoir fluids must be expanded to 100s if not 1000s of data — the single most significant limitation of the present work is the relatively small size of the database that was created for this work.
- *Correlations*: The new correlation relations proposed for this work should be considered more of an exercise in necessity in order to have the flexibility required to address the complexities of the different reservoir fluids in the database (*e.g.*, black oils, volatile oils, and retrograde gas condensates). Future work should consider less empiricism and more physics (*i.e.*, correlation models derived (at least in part) from thermodynamics).

REFERENCES

Al-Marhoun, M.A. (1988) "PVT Correlations for Middle East Crude Oils." *Journal of Petroleum Technology* 40 650—666. https://doi.org/10.2118/13718-PA

Breiman, L. and Friedman, J.H. (1985): "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 580-619.

https://users.soe.ucsc.edu/~draper/eBay-Google-2013-breiman-friedman-1985.pdf

Dindoruk, B. and Christman, P.G. (2004) "PVT Properties and Viscosity Correlations for Gulf of Mexico Oils." *SPE Reservoir Evaluation and Engineering* 7 427-437.

https://doi.org/10.2118/89030-PA

El-Banbi, A., Alzahabi, A., and El-Maraghi, A. (2018) *PVT Property Correlations — Selection and Estimation*, Elsevier.

https://www.elsevier.com/books/pvt-property-correlations/el-banbi/978-0-12-812572-4

Glaso, O. (1980) "Generalized Pressure-Volume-Temperature Correlations." *Journal of Petroleum Technology* 32 785-795. https://doi.org/10.2118/8016-PA

Kartoatmodjo, T. and Schmidt, Z. (July 1994) "Large Data Bank Improves Crude Physical Property Correlations," *Oil and Gas Journal* 51-55.

Lasater, J.A. (1958) "Bubble Point Pressure Correlation." *Journal of Petroleum Technology* 65-67. https://doi.org/10.2118/957-G

McCain, W.D., Soto, R.B., Valko, P.P., and Blasingame, T.A. (1998) "Correlation of Bubblepoint Pressures for Reservoir Oils--A Comparative Study." Paper presented at the 1998 SPE Eastern Regional Meeting, Pittsburgh, PA. https://doi.org/10.2118/51086-MS

Petrosky, G.E. Jr. (1990) "PVT Correlations for Gulf of Mexico Crude Oils," MS Thesis, U. of Southwestern Louisiana, Lafayette, Louisiana.

Petrosky, G.E. Jr., and Farshad, F. (1998) "Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils." *SPE Reservoir Evaluation and Engineering* 1 416—420.

https://doi.org/10.2118/51395-PA

Standing, M.B. (1947) "A Pressure-Volume-Temperature Correlation for Mixtures of California Oils and Gases," *API Drilling and Production Practice*.

Vasquez, M., and Beggs, H.D. (1980) "Correlations for Fluid Physical Property Prediction." *Journal of Petroleum Technology* 968-970. https://doi.org/10.2118/6719-PA

Velarde, J.J. (1996) "Correlation of Black Oil Properties at Pressures Below the Bubble-Point," MS Thesis, Texas A&M University, College Station, TX.

https://blasingame.engr.tamu.edu/0_TAB_Grad/TAB_Grad_Thesis_Archive/MS_013_VELARDE_Javier_TAMU_Thesis_(Aug_1996).pdf

Velarde, J.J., Blasingame, T.A., and McCain, W.D. (1997) "Correlation of Black Oil Properties At Pressures Below Bubble Point Pressure — A New Approach." Paper presented at the Annual Technical Meeting, Calgary, Alberta. https://doi.org/10.2118/97-93

Xue, G., Datta-Gupta, A., Valko, P.P., and Blasingame, T.A. (1997) "Optimal Transformations for Multiple Regression: Application to Permeability Estimation From Well Logs." SPE Formation Evaluation 12 85–94. https://doi.org/10.2118/35412-PA

APPENDIX A

PVT DATA USED IN THIS STUDY (ANONYMIZED)

The database used in this study was obtained from companies operating in unconventional oil and gas reservoirs and is anonymized as per agreement with the companies that donated these data. The samples are taken from several US and international unconventional (shale) plays.

A.1 Saturation Pressure Database

Table A-1 — Table of data PVT for unconventional oil and gas reservoirs used for the development of the saturation pressure correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (^{o}API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)
1	203	45.49	0.825	1504	3765
2	198	47.59	0.773	2003	4365
3	203	57.82	0.744	23828	5678
4	169	48.75	0.774	3982	6225
5	169	46.41	0.776	1328	3565
6	167	47.69	0.794	2681	4745
7	325	53.00	0.717	3955	4237
8	179	76.41	0.715	2335	5596
9	160	42.70	0.798	860	1921
10	240	38.02	0.937	1472	2505
11	180	54.78	0.779	5067	5031
12	175	48.80	0.744	5128	5100
13	182	50.80	0.437	4058	4898
14	168	56.87	0.834	20481	3615
15	190	45.02	0.759	1575	3615
16	207	47.86	0.762	4925	4735
17	212	46.42	0.787	3889	4565
18	183	64.10	0.743	6221	5166
19	195	58.27	0.757	7467	5072
20	187	52.24	0.788	4780	4900
21	214	56.59	0.725	14011	5705
22	197	45.59	0.853	1705	3965
23	190	52.13	0.770	3637	4832
24	193	46.48	0.812	1669	3815
25	215	40.20	0.809	996	2579

Table A-1 — (continued) Table of data PVT for unconventional oil and gas reservoirs used for the development of the saturation pressure correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (^{o}API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)
26	215	42.10	1.300	1096	2497
27	250	38.70	0.896	1632	3469
28	226	52.70	0.794	16843	3102
29	225	55.80	0.704	4850	3852
30	221	54.50	0.719	5069	3773
31	231	58.10	0.794	1102	2904
32	259	42.90	0.926	1683	2951
33	226	66.90	0.716	17428	3122
34	225	48.00	0.801	1362	3650
35	180	43.00	0.745	4585	7990
36	185	44.40	0.746	6132	7525
37	196	50.40	0.822	37504	4590
38	185	45.70	0.783	19218	10000
39	181	43.60	0.756	3714	6300
40	180	42.00	0.808	3772	6310
41	150	38.80	0.770	4518	8947
42	182	43.40	0.769	5747	8120
43	173	44.20	0.796	3328	5240
44	234	47.50	0.750	39340	1919
45	234	52.80	0.717	28662	1844
46	234	48.80	0.721	54104	2030
47	234	53.10	0.712	28763	3111
48	234	62.87	0.700	58504	1707
49	168	44.45	0.888	1012	2465
50	158	43.00	0.900	869	2265
51	160	38.70	0.724	927	2685
52	175	42.50	0.722	2650	5065
53	163	41.02	0.965	1168	2613
54	166	42.10	0.715	1419	3220
55	162	40.90	0.715	1262	2950
56	156	41.30	0.694	4012	5510
57	161	40.10	0.939	868	2678
58	132	37.64	0.756	747	3150
59	178	39.60	0.917	865	2335

Table A-1 — (continued) Table of data PVT for unconventional oil and gas reservoirs used for the development of the saturation pressure correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (°API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)
60	126	37.70	0.886	520	2188
61	147	41.90	0.843	705	2290
62	153	40.40	0.826	661	2370
63	168	43.30	0.853	1000	2875
64	174	42.60	0.938	1100	3328
65	159	42.80	0.776	757	2721
66	189	45.10	0.894	1500	3478
67	188	45.10	0.909	1500	3259
68	240	39.00	1.008	612	1657
69	240	38.10	0.833	485	1389
70	240	36.70	0.944	483	1778
71	240	34.20	0.810	683	2057
72	245	35.17	0.887	1145	2674
73	237	34.60	1.058	915	2530
74	212	42.20	0.932	1275	2304
75	270	43.60	0.755	1230	2866
76	225	37.29	0.951	336	1211
77	225	34.60	0.994	275	1370
78	235	35.46	0.920	558	2122
79	238	43.58	0.720	1035	3064
80	225	30.00	0.650	638	2125
81	225	30.00	0.650	628	2100
82	228	30.00	0.900	762	2455
83	228	30.00	0.900	757	2440
84	228	30.00	0.650	764	2459
85	221	42.50	0.900	513	1764
86	225	34.60	0.994	275	1370
87	205	49.90	0.754	4889	5070
88	210	81.90	0.723	38878	2429
89	235	61.83	0.796	4912	3998
90	233	55.73	0.816	3912	4298
91	230	64.78	0.790	9809	3918
92	220	73.47	0.736	17933	3129
93	220	64.59	0.796	6097	4395

Table A-1 — (continued) Table of data PVT for unconventional oil and gas reservoirs used for the development of the saturation pressure correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (°API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)
94	218	88.88	0.775	19602	3188
95	247	74.50	0.701	85802	2577
96	151	40.50	0.918	550	1991
97	159	37.80	0.864	575	2388
98	164	39.10	0.822	825	3044
99	166	38.10	0.875	875	3297
100	166	40.40	0.993	735	2588
101	171	44.70	0.804	912	3000
102	175	44.80	0.901	1400	3200
103	159	40.80	0.897	745	2568
104	175	44.80	0.901	1200	3200
105	290	78.80	0.746	9177	3906
106	290	52.60	0.729	2901	4343
107	330	66.88	0.740	2329	3969
108	326	53.70	0.718	3187	4122
109	327	57.70	0.728	6979	4848
110	330	59.00	0.724	5771	3946
111	326	55.60	0.718	3568	4391
112	320	52.90	0.722	2449	4276
113	319	55.00	0.735	3262	4379
114	250	77.21	0.741	28130	2814
115	291	82.20	0.764	9308	3910
116	323	56.70	0.729	4573	4373
117	302	49.80	0.711	1142	3866
118	304	48.90	0.724	1224	3721
119	308	51.10	0.729	1464	3859
120	313	51.50	0.715	1944	4176
121	315	53.30	0.731	2492	4317
122	312	52.00	0.728	2711	4792
123	316	50.60	0.725	2504	4421
124	309	51.10	0.730	2102	4034
125	307	49.80	0.742	1701	4292
126	150	39.80	0.771	1175	3452
127	150	33.70	0.795	540	2172

Table A-1 — (continued) Table of data PVT for unconventional oil and gas reservoirs used for the development of the saturation pressure correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (^{o}API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)
128	161	47.20	0.759	6007	5675
129	161	41.40	0.767	3220	5120
130	297	81.41	0.732	2574	4431
131	298	83.82	0.797	3094	4550
132	310	78.46	0.752	7436	5160
133	310	80.31	0.747	6560	4520
134	298	76.96	0.798	2529	4355
135	285	74.56	0.793	1567	4148
136	305	76.54	0.810	2285	4213
137	220	53.10	0.730	7493	4385
138	293	74.68	0.737	2533	4149

A.2 Formation Volume Factor Database

Table A-2 — Table of data PVT for unconventional oil and gas reservoirs used for the development of the oil formation volume factor correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (°API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)	Formation Volume Factor at p_{sat} (bbl/STB)
1	203	45.49	0.825	1504	3765	1.874
2	198	47.59	0.773	2003	4365	2.108
3	169	46.41	0.776	1328	3565	1.704
4	167	47.69	0.794	2681	4745	2.441
5	160	42.70	0.798	860	1921	1.392
6	240	38.02	0.937	1472	2505	2.091
7	190	45.02	0.759	1575	3615	1.837
8	207	47.86	0.762	4925	4735	3.838
9	187	52.24	0.788	4780	4900	3.709
10	197	45.59	0.853	1705	3965	1.965
11	193	46.48	0.812	1669	3815	1.947
12	215	40.20	0.809	996	2579	1.708
13	215	42.10	1.300	1096	2497	1.726
14	250	38.70	0.896	1632	3469	2.112
15	231	58.10	0.794	1102	2904	2.889
16	259	42.90	0.926	1683	2951	2.214
17	225	48.00	0.801	1362	3650	2.798
18	168	44.45	0.888	1012	2465	1.593
19	158	43.00	0.900	869	2265	1.490
20	160	38.70	0.724	927	2685	1.502
21	175	42.50	0.722	2650	5065	2.360
22	163	41.02	0.965	1168	2613	1.658
23	166	42.10	0.715	1419	3220	1.804
24	162	40.90	0.715	1262	2950	1.716
25	156	41.30	0.694	4012	5510	3.150
26	161	40.10	0.939	868	2678	1.514
27	132	37.64	0.756	747	3150	1.343
28	178	39.60	0.917	865	2335	1.549
29	126	37.70	0.886	520	2188	1.329
30	147	41.90	0.843	705	2290	1.502

Table A-2 — (continued) Table of data PVT for unconventional oil and gas reservoirs used for the development of the oil formation volume factor correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (°API)	Separator Gas Gravity (air = 1)	Solution-Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)	Formation Volume Factor at p_{sat} (bbl/STB)
31	153	40.40	0.826	661	2370	1.475
32	168	43.30	0.853	1000	2875	1.669
33	174	42.60	0.938	1100	3328	1.705
34	159	42.80	0.776	757	2721	1.608
35	189	45.10	0.894	1500	3478	1.954
36	188	45.10	0.909	1500	3259	1.966
37	240	39.00	1.008	612	1657	1.471
38	240	38.10	0.833	485	1389	1.374
39	240	36.70	0.944	483	1778	1.429
40	240	34.20	0.810	683	2057	1.530
41	245	35.17	0.887	1145	2674	1.779
42	237	34.60	1.058	915	2530	1.714
43	212	42.20	0.932	1275	2304	1.873
44	270	43.60	0.755	1230	2866	1.852
45	205	49.90	0.754	4889	5070	3.691
46	161	41.40	0.767	3220	5120	2.700

A.3 Oil Viscosity Database

Table A-3 — Table of data PVT for unconventional oil and gas reservoirs used for the development of the oil viscosity correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (°API)	Separator Gas Gravity (air = 1)	Solution Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)	Oil Viscosity at p_{sat} (cp)
1	203	45.49	0.825	1504	3765	0.272
2	198	47.59	0.773	2003	4365	0.196
3	169	46.41	0.776	1328	3565	0.354
4	167	47.69	0.794	2681	4745	0.189
5	160	42.70	0.798	860	1921	0.289
6	240	38.02	0.937	1472	2505	0.182
7	190	45.02	0.759	1575	3615	0.211
8	207	47.86	0.762	4925	4735	0.062
9	187	52.24	0.788	4780	4900	0.154
10	197	45.59	0.853	1705	3965	0.246
11	193	46.48	0.812	1669	3815	0.256
12	168	44.45	0.888	1012	2465	0.467
13	158	43.00	0.900	869	2265	0.499
14	160	38.70	0.724	927	2685	0.435
15	175	42.50	0.722	2650	5065	0.243
16	163	41.02	0.965	1168	2613	0.300
17	166	42.10	0.715	1419	3220	0.300
18	162	40.90	0.715	1262	2950	0.340
19	156	41.30	0.694	4012	5510	0.140
20	161	40.10	0.939	868	2678	0.450
21	178	39.60	0.917	865	2335	0.457
22	147	41.90	0.843	705	2290	0.391
23	153	40.40	0.826	661	2370	0.429
24	168	43.30	0.853	1000	2875	0.280
25	174	42.60	0.938	1100	3328	0.292
26	159	42.80	0.776	757	2721	0.363
27	189	45.10	0.894	1500	3478	0.205
28	188	45.10	0.909	1500	3259	0.202
29	240	39.00	1.008	612	1657	0.352
30	240	38.10	0.833	485	1389	0.488
31	240	36.70	0.944	483	1778	0.453
32	240	34.20	0.810	683	2057	0.476
33	245	35.17	0.887	1145	2674	0.184

Table A-3 — (continued) Table of data PVT for unconventional oil and gas reservoirs used for the development of the oil viscosity correlations from anonymous sources.

Point	Reservoir Temperature (Deg F)	Stock-Tank Oil Gravity (°API)	Separator Gas Gravity (air = 1)	Solution Gas-Oil Ratio (scf/STB)	Saturation Pressure (psia)	Oil Viscosity at p_{sat} (cp)
34	237	34.60	1.058	915	2530	0.392
35	212	42.20	0.932	1275	2304	0.257
36	270	43.60	0.755	1230	2866	0.239
37	205	49.90	0.754	4889	5070	0.074
38	161	41.40	0.767	3220	5120	0.150

APPENDIX B

CORRELATION OF OIL VISCOSITY FOR UNCONVENTIONAL RESERVOIR FLUIDS USING A LIMITED DATABASE

This Appendix provides the additional work conducted to correlate the limited data that were available for oil viscosity at the saturation pressure (μ_{ob}). This is a limited dataset of 46 points as given in **Appendix A** (**Table A.3**). Traditional correlations of oil viscosity refer to the viscosity at a given pressure and temperature related to the so-called "dead oil viscosity" at a standard temperature and pressure (this is the approach of Chew and Connally [1959]). As there were no "dead oil viscosity" data for this study, the "dead oil viscosity" formulation for a given correlation was substituted into the μ_{ob} correlation and solved for all coefficients using <u>only</u> the μ_{ob} data. Obviously this is neither ideal, nor the likely to produce unique correlations, but this was a limitation of the dataset and is the reason this work is provided only as an auxiliary companion to the main correlations of p_{sat} and B_{ob} developed in the main body of this work.

For this work, it is assumed that oil viscosity at the saturation pressure (μ_{ob}) can be correlated using the same variables as the saturation pressure and the oil formation volume factor at the saturation pressure, *i.e.*, $\mu_{ob} = f(R_{sb}, \gamma_g$ (or SG), $API, T, p_{sat})$.

B.1 Literature Review of Previous Work

As noted above, most (essentially all) prior correlations of the correlations of oil viscosity refer to the "dead oil viscosity" (μ_{od}) (Chew and Connally [1959]). The Chew and Connally approach assumes that $\mu_{od} = f(API, T)$ and $\mu_{ob} = f(\mu_{od}, R_{sb})$, and since there are no data for μ_{od} in the database used for this work, the correlation relation for μ_{od} (all methods) becomes an "intermediate" or "proxy" correlation and μ_{od} is computed as an input for the given μ_{ob} correlation as part of the regression process.

The historical correlations for oil viscosity that are used in this work are summarized in **Table B.1**:

Table B.1 — Historical Correlations for Oil Viscosity (μ_{ob}) used in This Work.

Correlation	μ_{od} form	μ_{ob} form
Petrosky and Farshad (1990, 1998)	$\mu_{od} = f(API, T)$	$\mu_{ob} = f(\mu_{od}, R_{sb})$
Kartoatmodjo and Schmidt (1994)	$\mu_{od} = f(API, T)$	$\mu_{ob} = f(\mu_{od}, R_{sb})$
Dindoruk and Christman (2004)	$\mu_{od} = f(API, T)$	$\mu_{ob} = f(\mu_{od}, R_{sb})$

As process, the coefficients for each correlation are solved using regression on the available data $\mu_{ob} = f(R_{sb}, API, T)$, where the "dead oil viscosity" (μ_{od}) is simply an "intermediate" or "implicit" variable in the μ_{ob} formulation (*i.e.*, there are no μ_{od} data, this is simply a correlation with the given μ_{ob} correlation).

As a "test" of the relevance of the oil viscosity at saturation pressure data, the oil viscosity at saturation pressure (μ_{ob}) is plotted versus the saturation pressure (p_{sat}) in **Fig. 3.1**, and a reasonable correlative trend is observed. Obviously, μ_{ob} is not solely a function of p_{sat} — there are temperature and composition effects that must be addressed, the simple rendering of μ_{ob} versus p_{sat} shown in **Fig. 3.1** is used as a data observation/validation tool.

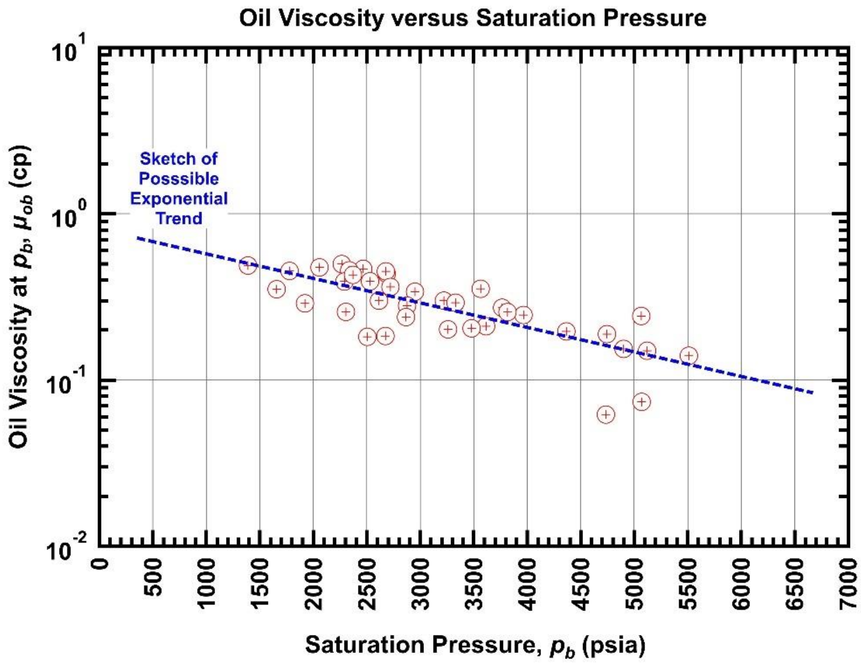

Figure B.1 — Semilog plot of measured oil viscosity versus saturation pressure for data taken from the unconventional reservoir fluid database.

B.2 Refitted Historical Correlations for Oil Viscosity

In this section regression methods are used to re-fit the historical correlations for oil viscosity to the available database of viscosity data (*i.e.*, **Appendix A** (**Table A.3**)). All of the historical correlations use a "dead oil viscosity" (μ_{od}) correlation — as there are no μ_{od} data available, the μ_{od} correlation is used implicitly inside of the μ_{ob} correlation. This process yields relevant, but probably not unique correlations for μ_{ob} behavior. In the following section a completely general (exponential-polynomial) formulation is applied to the $\mu_{ob} = f(R_{sb}, \gamma_g$ (or SG), API, T, $p_{sat})$ dataset.

The refitted historical correlations are provided below, and it is noted that regression algorithms in Python and MS Excel were used to complete the optimization (re-fitting) of these correlations.

Petrosky and Farshad Correlation for Oil Viscosity:

The Petrosky and Farshad correlation re-fitted to the unconventional database yields:

Dead Oil Viscosity: (at standard conditions)

$$\mu_{od} = 2.257 \times 10^{-7}\ T^{-2.7772} \log API^{[2.0073 \log T - 12.5688]} \quad \text{(B.1)}$$

Oil Viscosity: (at p_{sat})

$$\mu_{ob} = \hat{A}\,[\mu_{od}]^{\hat{B}} \quad \text{(B.2)}$$

where:

$$\hat{A} = 0.04864 + 1.03243 \times 10^{[-2.01634 \times 10^{-4} R_{sb}]} \quad \text{(B.3)}$$

$$\hat{B} = 1.56922 + 1.92503 \times 10^{[-0.0010435\ R_{sb}]} \quad \text{(B.4)}$$

The statistical results for the Petrosky and Farshad correlation are given in **Table B.2**.

Table B.2 — Statistics: Oil Viscosity (μ_{ob}) — Petrosky and Farshad Correlation

Parameter	Value
Sum of Squared Residuals, $(cp)^2$	5.450
Standard Deviation, cp	0.116
Variance, $(cp)^2$	0.013
Average Absolute Relative Error, %	28.77

The graphical results for the Petrosky and Farshad correlation are shown in **Fig. B.2**. In general, the Petrosky and Farshad is reasonably well-correlated about the perfect correlation trend. There is most probably some overfitting in the region of 0.5 cp, where the Petrosky and Farshad correlation is "better" (*i.e.*, over-fitted) compared to the GRACE correlation (which should always yield the lowest error statistics).

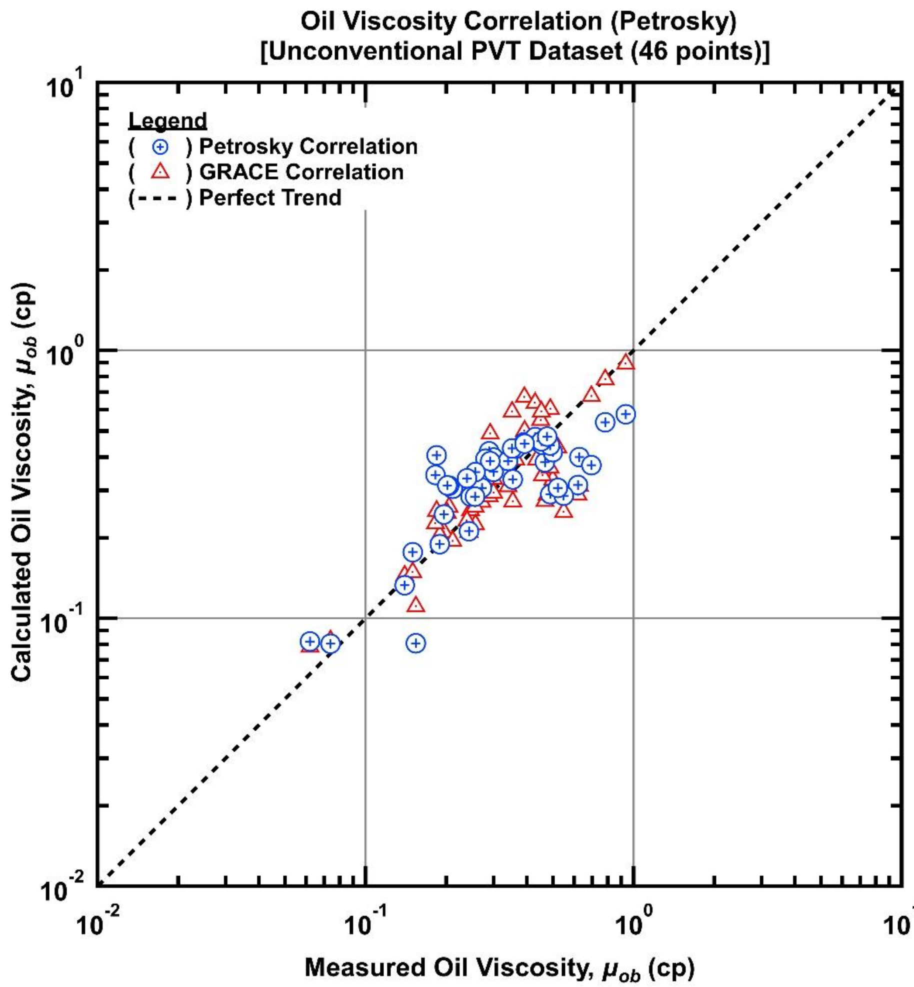

Figure B.2 — Log-log plot of calculated versus measured oil viscosity for the Petrosky and Farshad correlation model, applied using non-linear regression to the uncon-ventional reservoir fluid data set.

Kartoatmodjo and Schmidt Correlation for Oil Viscosity:

The Kartoatmodjo and Schmidt correlation re-fitted to the unconventional database yields:

Dead Oil Viscosity: (at standard conditions)

$$\mu_{od} = 1.3204 \times 10^{10}\ T^{-0.65188}\ \log API^{[-0.065912\ \log T - 27.35313]} \text{ ....(B.5)}$$

Oil Viscosity: (at p_{sat})

$$\mu_{ob} = -0.002134 + 11.494 X_2 - 8.501 \times 10^{-4} X_2^{\,2} \text{ ....(B.6)}$$

where:

$$X_2 = \left[1.65928 + 0.11995 \times 10^{[-5.98 \times 10^{-2} R_{sb}]}\right] \mu_{od}^{\;X_1} \text{ ....(B.7)}$$

$$X_1 = 0.23655 - 4.89068 \times 10^{[-1.76 \times 10^{-2} R_{sb}]} \text{ ....(B.8)}$$

The statistical results for the Kartoatmodjo and Schmidt correlation are given in **Table B.3**.

Table B.3 — Statistics: Oil Viscosity (μ_{ob}) — Kartoatmodjo and Schmidt Correlation

Parameter	Value
Sum of Squared Residuals, (cp)^2	13.07
Standard Deviation, cp	0.067
Variance, (cp)^2	0.004
Average Absolute Relative Error, %	36.97

The graphical results for the Kartoatmodjo and Schmidt correlation are shown in **Fig. B.3**. The first observation is that the Kartoatmodjo and Schmidt correlation model systematically deviates from the perfect correlation trend, indicating either a poor regression or a weak correlation model, or both. Speculation suggests that the complexity of the μ_{ob} model (*i.e.*, Eqs. B.6-B.8) is the cause of the relatively poor correlation fit for this particular case.

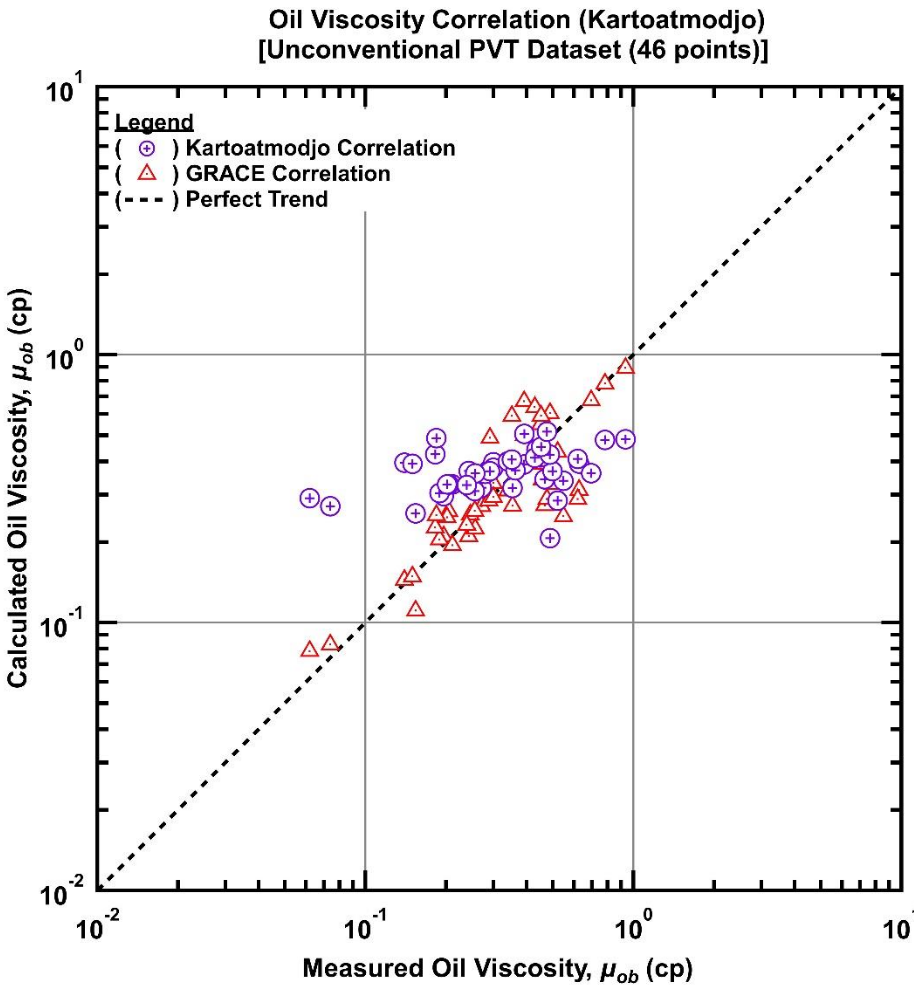

Figure B.3 — Log-log plot of calculated versus measured oil viscosity for the Kartoatmodjo and Schmidt correlation model, applied using non-linear regression to the unconventional reservoir fluid data set.

Dindoruk and Christman Correlation for Oil Viscosity:

The Dindoruk and Christman correlation re-fitted to the unconventional database yields:

Dead Oil Viscosity: (at standard conditions)

$$\mu_{od} = \frac{9.7583\times10^{9}\ T^{-2.8296} \log API^{[1.7026\times10^{1} \log T - 8.988]}}{7.648\times10^{-1}\ p_{sat}^{1.0348} + 0.01043\ R_{sb}^{-7.769\times10^{-4}}} \quad \text{(B.9)}$$

Oil Viscosity: (at p_{sat})

$$\mu_{ob} = \hat{A}[\mu_{od}]^{\hat{B}} \quad \text{(B.10)}$$

where:

$$\hat{A} = \frac{0.9708}{\exp\left[7.155\times10^{-4}\ R_s\right]} + \frac{\left[-9.375\times10^{-3}\right] R_s^{4.717\times10^{-1}}}{\exp\left[9.408\times10^{-4}\ R_{sb}\right]} \quad \text{(B.11)}$$

$$\hat{B} = \frac{0.34228}{\exp\left[-2.189\times10^{-5} R_{sb}\right]} + \frac{\left[-1.758\times10^{-2}\right] R_{sb}^{4.488\times10^{-1}}}{\exp\left[-2.753\times10^{-4} R_{sb}\right]} \quad \text{(B.12)}$$

The statistical results for the Dindoruk and Christman correlation are given in **Table B.4**.

Table B.4 — Statistics: Oil Viscosity (μ_{ob}) — Dindoruk and Christman

Parameter	Value
Sum of Squared Residuals, $(cp)^2$	5.598
Standard Deviation, cp	0.128
Variance, $(cp)^2$	0.017
Average Absolute Relative Error, %	34.33

The graphical results for the Dindoruk and Christman correlation are presented in **Fig. B.4** and are quite similar to those of Petrosky and Farshad, which also suggests a (very) slight "over-fitting" compared to the GRACE correlation as the Dindoruk and Christman correlation results also compare very well with GRACE results.

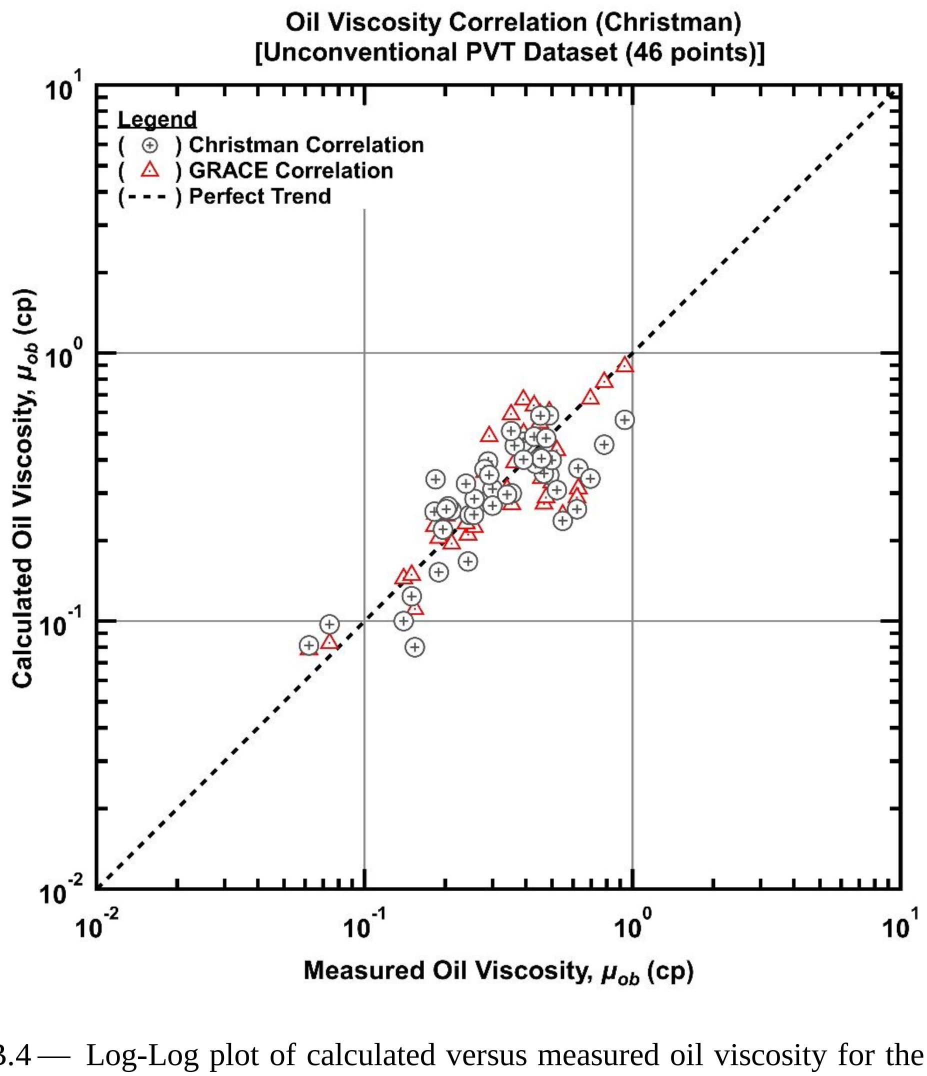

Figure B.4 — Log-Log plot of calculated versus measured oil viscosity for the Dindoruk and Christman correlation model, applied using non-linear regression to the unconventional reservoir fluid data set.

Statistical Comparison of Re-Fitted Historical Oil Viscosity Correlations:

Table B.5 provides the statistical summary of the historical oil viscosity correlations fitted to the database for unconventional reservoir fluids. It is important to note that there are only 46 data points used in these correlations for oil viscosity at the saturation pressure.. As comment, the Petrosky and Farshad and Dindoruk and Christman models performed well both statistically and in a visual sense. However; the Kartoatmodjo-Schmidt model clearly does not match the data nor the perfect correlation trend, suggesting that this model may not be appropriate.

Table B.5 — Statistics: Oil Viscosity (μ_{ob}) — All Historical Correlations (46 points)

Source	Standard Deviation, cp	Variance, $(cp)^2$	Average Absolute Relative Error, % (this work)
Petrosky and Farshad (1990, 1998)	0.116	0.013	28.8
Kartoatmodjo and Schmidt (1994)	0.067	0.004	37.0
Dindoruk and Christman (2004)	0.128	0.017	34.3
Non-Parametric Solution (GRACE)	0.186	0.0344	25.4

B.3 Development of New Correlations

As was discussed at the beginning of this Appendix, the historical models for oil viscosity (μ_{ob}) were compared for accuracy and for nature of the model match to the data (*e.g.*, the Kartoatmodjo and Schmidt model did not match the data nor the perfect correlation trend, Fig. B.4). Having completed this work, the next goal was to establish a best-fit correlation model where $\mu_{ob} = f(R_{sb}, \gamma_g$ (or SG), $API, T, p_{sat})$.

As a general correlation is desired (*i.e.*, a direct solution for μ_{ob} without having to have or compute the dead oil viscosity [μ_{od}]), several different types of models were considered, the most "flexible" of which is the exponential-quadratic model. This correlation is performed in logarithmic space (hence the "exponential" must be taken to obtain μ_{ob}) and each independent variable (R_{sb}, γ_g (or SG), API, T, p_{sat}) is expressed with a general quadratic (logarithmic) relation. The proposed general form of the exponential-quadratic model is:

$$\begin{aligned}\ln \mu_{ob} = \; & \left[c_1 + c_2 \ln T + c_3 [\ln T]^2\right] \\ \times & \left[c_4 + c_5 \ln API + c_6 [\ln API]^2\right] \\ \times & \left[c_7 + c_8 \ln R_{sb} + c_9 [\ln R_{sb}]^2\right] \\ \times & \left[c_{10} + c_{11} \ln \gamma_g + c_{12} [\ln \gamma_g]^2\right] \\ \times & \left[c_{13} + c_{14} \ln p_b + c_{15} [\ln p_b]^2\right]\end{aligned}$$

.... (B.13)

In fitting Eq. B.13 to the oil viscosity database, it became necessary to remove 8 (eight) points from the dataset. While this is neither ideal nor preferred, this was necessary in order for Eq. B.13 to converge to a statistically significant regression. In fact, given that there are 15 coefficients and only 38 data points, it is very likely that Eq. B.13 would overfit the dataset, but also that it would be difficult to converge statistically. The results for this correlation effort are given in Eq. B.14. The statistical summary is show in **Table B.6** and the correlation plot is shown in **Fig. B.5**.

$$\begin{aligned}\ln \mu_{ob} = \; & \left[1.549 - 0.509 \ln T + 1.194\times10^{-3} [\ln T]^2\right] \\ \times & \left[-0.287 - 0.682 \ln API - 5.456\times10^{-2} [\ln API]^2\right] \\ \times & \left[0.360 - 0.121 \ln R_{sb} + 1.126\times10^{-2} [\ln R_{sb}]^2\right] \\ \times & \left[0.414 - 1.019\times10^{-2} \ln \gamma_g - 0.101 [\ln \gamma_g]^2\right] \\ \times & \left[0.770 + 0.521 \ln p_b + 1.287\times10^{-3} [\ln p_b]^2\right]\end{aligned}$$

....(B.14)

Table B.6 — Statistics: Oil Viscosity (μ_{ob}) — Exponential-Quadratic Model (38 points)

Parameter	Value
Sum of Squared Residuals, $(cp)^2$	1.010
Standard Deviation, cp	0.112
Variance, $(cp)^2$	0.013
Average Absolute Relative Error, %	13.90
Average Absolute Relative Error, % (GRACE)	18.84

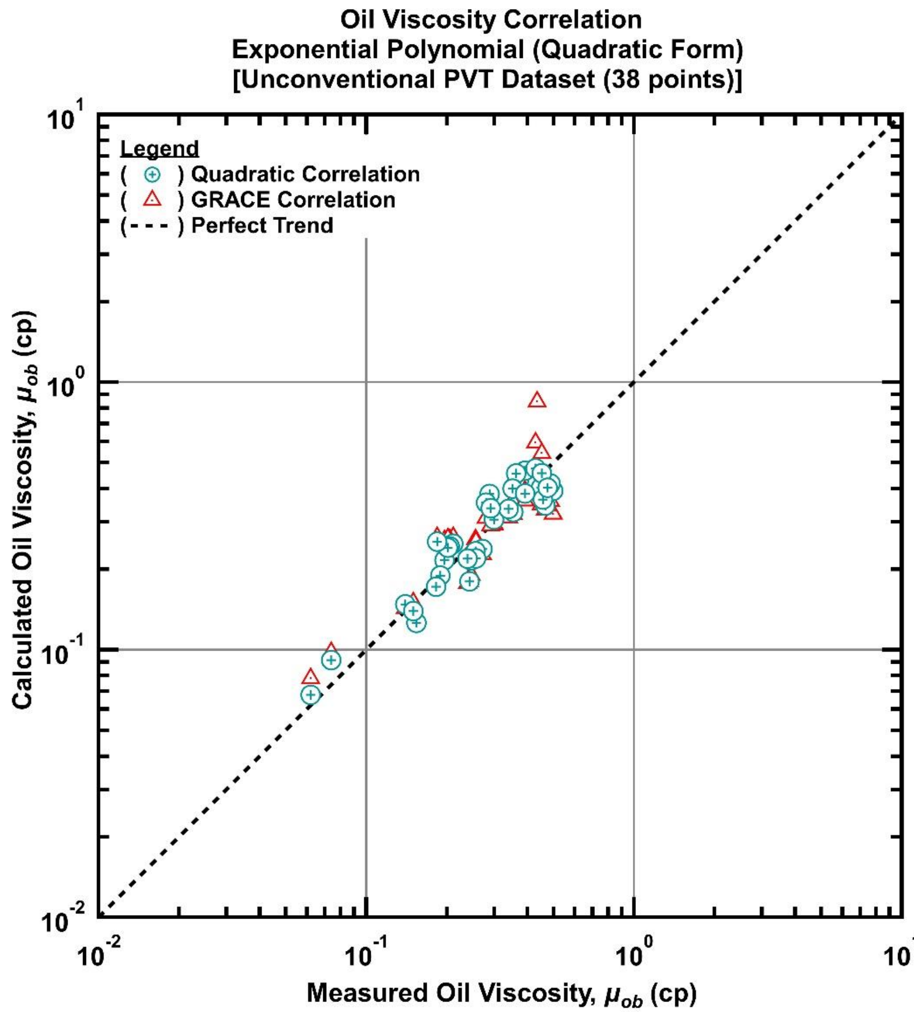

Figure B.5 — Log-log plot of calculated versus measured oil viscosity for the exponential-quadratic correlation model, applied using non-linear regression to the unconventional reservoir fluid data set. (38 points)

As given in **Table B.6**, the statistics related to fitting Eq. B.14 to the oil viscosity database clearly shows that the Exponential-Quadratic Model is overfitted (*i.e.*, AARE = 13.90 percent compared to the GRACE correlation at 18.84 percent — where the GRACE correlation is a non-parametric correlation and should always provide the lowest error metrics). However; this "overfitting" was expected, and Eq. B.14 is provided as a conceptual starting point for cases where (much) more data are available.

Statistical Comparison of All Viscosity Correlations:

The statistical results for all correlations are summarized in **Table B.7**. While it would be unwise to draw strong conclusions, it is reasonable to state that the new (Exponential-Quadratic) model is a "best fit" (see also **Fig. B.5**), the Petrosky and Farshad and the Dindoruk and Christman models also perform well and are likely robust and representative, but again, any strong conclusions should wait until there is a much larger dataset to establish more statistically consistent correlations.

Table B.7 — Statistics: Oil Viscosity (μ_{ob}) — Historical and New Correlations

Source	Number of Points	Standard Deviation, cp	Variance, $(cp)^2$	Average Absolute Error, % (this work)
Exponential-Quadratic Model (this study)	38	0.112	0.013	13.9
Non-Parametric Correlation (GRACE)	38	0.112	0.013	18.8
Petrosky and Farshad (1990, 1998)	46	0.116	0.013	28.8
Kartoatmodjo and Schmidt (1994)	46	0.067	0.004	37.0
Dindoruk and Christman (2004)	46	0.128	0.017	34.3
Non-Parametric Correlation (GRACE)	46	0.186	0.034	25.4

B.4 Summary, Conclusions, and Recommendations

Summary:

The refitting of the historical correlations for oil viscosity against the dataset available for this project was largely successful; as all of these correlations, with the exception of the Kartoatmodjo and Schmidt model, were successfully "re-fitted" to the dataset for fluids from unconventional reservoirs. It should be noted that the historical correlations only employ three variables: solution gas-oil ratio (R_s), stock-tank oil gravity (API), and temperature (T).

In addition, a new generalized model, the "exponential-quadratic" model incorporates all variables in the database [*i.e.*, $\mu_{ob} = f(R_{sb}, \gamma_g$ (or SG), API, T, $p_{sat})$] was proposed and fitted to the database. The "exponential-quadratic" model did yield the best statistics, but this relation also "over-fitted" the data — the statistical results for this model were significantly below those of the non-parametric correlation, which is a definitive sign of "over-fitting."

Conclusions:

The following conclusions were observed for the work on oil viscosity in this Appendix:

- The Petrosky and Farshad (1990, 1998) model was successfully fitted to the database for this work and was probably over-fitted.
- The Kartoatmodjo and Schmidt (1994) was <u>not</u> successfully fitted to the database, most likely due to a poor regression and/or an inadequate model.
- The Dindoruk and Christman (2004) model was successfully fitted to the database and was possibly over-fitted (but not significantly).
- The exponential-quadratic model (this study) yielded the best statistical fit (by a significant margin) over all of the correlations tested, but was definitely (and substantially) over-fitted.

- The non-parametric correlation (GRACE) does not provide a functional correlation of the data, but was successfully deployed to establish the lower bound of error statistics. This approach is essential as it provides the basis of comparison for parametric models in order to establish over-fitting.

Recommendations:

The following recommendations are proposed for future work performed on correlating oil viscosity for fluids from unconventional reservoirs.

- A (much) larger database of oil viscosity (and oil formation volume factor) must be established for fluids from unconventional reservoirs.
- The parametric formulations of the historical oil viscosity correlations only include R_{sb}, *API,* and T — such formulations should include all of the variables typically available for black oil correlations (*i.e.*, R_{sb}, γ_g, *API*, T, and p_{sat}).

B.5 References

Chew, J.N. and Connally, C.A (1959) "A Viscosity Correlation for Gas-Saturated Crude Oils." *Trans.* 216, 23–25. https://doi.org/10.2118/1092-G

Breiman, L. and Friedman, J.H. (1985): "Estimating Optimal Transformations for Multiple Regression and Correlation," *Journal of the American Statistical Association*, 580-619. https://users.soe.ucsc.edu/~draper/eBay-Google-2013-breiman-friedman-1985.pdf

Dindoruk, B. and Christman, P.G. (2004) "PVT Properties and Viscosity Correlations for Gulf of Mexico Oils." *SPE Reservoir Evaluation and Engineering* 7 427-437. https://doi.org/10.2118/89030-PA

Kartoatmodjo, T. and Schmidt, Z. (July 1994) "Large Data Bank Improves Crude Physical Property Correlations," *Oil and Gas Journal* 51-55.

Petrosky, G.E. Jr. (1990) "PVT Correlations for Gulf of Mexico Crude Oils," MS Thesis, U. of Southwestern Louisiana, Lafayette, Louisiana.

Petrosky, G.E. Jr., and Farshad, F. (1998) "Pressure-Volume-Temperature Correlations for Gulf of Mexico Crude Oils." *SPE Reservoir Evaluation and Engineering* 1 416—420.

https://doi.org/10.2118/51395-PA

Xue, G., Datta-Gupta, A., Valko, P.P., and Blasingame, T.A. (1997) "Optimal Transformations for Multiple Regression: Application to Permeability Estimation From Well Logs." SPE Formation Evaluation 12 85–94. https://doi.org/10.2118/35412-PA