

PREDICTIVE UNCERTAINTY OF FLUID MODELS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF ENERGY
RESOURCES ENGINEERING
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Lívia Fulchignoni de Paiva

June 2023

Abstract

Fluid models play a critical role in a wide range of applications, such as oil and gas exploration and production, chemical process design and optimization, and CO₂ enhanced oil recovery projects. Accurate predictions of phase behavior and thermo-physical properties are key to ensuring the safety, performance, and profitability of these operations. However, predictions derived from fluid models inherently possess uncertainties due to model and parametric factors. These uncertainties are especially significant for hydrocarbon reservoir fluids, which exhibit complex phase behavior. This dissertation discusses the sources of uncertainty in reservoir fluid modeling, encompassing both empirical and compositional approaches, and provides a deeper understanding of their impact on fluid and flow model predictions.

Initially, we demonstrate the uncertainty in model selection by comparing nine traditional empirical models for solution gas-oil ratios against experimental data. The results reveal that specific fluid compositions under different pressure and temperature conditions are better represented by distinct models. Consequently, we propose a new model that delivers superior overall performance.

We also explore the uncertainty in input parameters for compositional fluid models, specifically highlighting how subjective choices of optimization algorithms and initial guesses impact the equation of state (EoS) regression process. As a result, EoS predictions remain uncertain even after tuning the uncertain inputs to a limited set of experimental data points. We present results for two hydrocarbon reservoir fluids, treating five properties of the heaviest carbon fraction as design variables. Although all considered optimization algorithms and initial guesses match experimental data for gas and liquid properties, the resulting EoS parameterizations lead to dramatically

different predictions of the fluid's thermophysical behavior in unsampled pressure and temperature regions.

Next, we investigate the uncertainty in input parameters for fluid models, introducing a framework to quantify the predictive uncertainty of multiphase pipe-flow models due to correlated random inputs. A case study evaluates the uncertainty in cumulative production for a reservoir with unknown fluid properties during the exploration phase. Global Sensitivity Analysis using Sobol's indices is employed to identify inputs significantly contributing to the model's predictive uncertainty. This framework facilitates improved risk management and informed decision-making within the energy industry.

Lastly, we discuss the calibration of inherently uncertain flow models when field data is available. In particular, we propose a heuristic method for optimizing tuning factors applied to calculated pressure and temperature gradients and demonstrate its efficiency through a real case study.

Acknowledgements

I would like to take this opportunity to express my sincere gratitude to the individuals who have held my hand during my PhD journey and transformed this academic pursuit into a fulfilling and enlightening voyage. Without them, the path would have been boring and unsuccessful.

I would like to thank Daniel Tartakovsky, my advisor, for his empathetic understanding, support, and invaluable research insights that guided me throughout this challenging process. While allowing me to have the freedom and flexibility to delve into my research interests, he would lead me beyond my comfort zone towards more intellectually stimulating and rewarding discussions that ultimately elevated the quality of my work.

I would like to thank Professor Tony Kovscek for his incredible teaching skills that transformed my fear of Thermodynamics into genuine enjoyment for the subject. I am grateful and honored to have had the opportunity to engage in thought-provoking research discussions with Dr. Khalid Aziz. His research has inspired me throughout my academic journey and I still cannot wrap my head around him being in my reading committee. I am grateful to Professor Tapan Mukerji for his interest in my research and for accepting my invitation to be part of my orals committee. I also thank Professor Hai Wang for graciously accepting my invitation to Chair my PhD defense.

I am thankful to my research group for the sense of community and camaraderie. I am also grateful to the department's staff for their dedication to the students' well-being and efforts in organizing engaging social activities.

I have been fortunate to forge great relationships during my time at Stanford. To Alaa, Isaac, Jabs, Josue, Lama, Madalsa, Margariete, Rayan, Rasim, Weiyu, Xiaoyu,

and Zitong, for warming my heart and balancing long periods of studying with great get-togethers. To Nancy, for inspiring conversations that ended up with a next-day, wonderful trip to Wisconsin.

I would like to extend my gratitude to Petrobras for fostering a culture that pursues technical excellence and the academic development of its employees. The collective wisdom, expertise, and encouragement of my colleagues have left a lasting impact on both my academic and professional journey. I am particularly grateful to Roberto Carlos, for promoting a work environment where I felt valued and motivated to excel and for consistently supporting the team. I also wish to express my heartfelt appreciation to Roberto Fonseca, for his belief in my potential and for his important mentorship from the outset. I am deeply thankful for Dr. Elisio's attentive listening, career advice, and steadfast support that led to my PhD proposal being accepted by the company.

My warmest gratitude to my friends. Our constant exchanges through shared photos, stories, songs, heartfelt phone conversations and memorable Zoom meetings, have been truly a joy, and I cannot overstate the importance of your care and friendship. To my dear childhood friends, Barbara, Danielle, Joice, and Nicolle, thank you for the therapy group always just a call away. To Gabi, Natasha, Pepeu, Rachel, Rafael, Thadeu, and Yann, thank you for the inevitable joy and laughter that we share whenever we connect - even if that means on WhatsApp.

My love and appreciation to Matt, my partner in so many ways. Be it playing tennis, traveling, trying new restaurants, or just binge watching Netflix, his presence has brightened my days and enriched my life.

This dissertation is dedicated to my family, who made it all possible. To Vinicius, Rodrigo, and Flavio for the cherished family gatherings, brimming with affection and encouragement. To my sister, Carla, for our unique connection that transcends distance and for being my source of laughter, joy, and heartfelt conversations. To my parents, Adriana and Roberto, for what words cannot describe.

Contents

Abstract	iv
Acknowledgements	vi
1 Introduction	5
1.1 Objectives	7
1.2 Dissertation outline	8
2 Review of reservoir fluid modeling	10
2.1 Black Oil fluid modeling	10
2.2 Compositional fluid modeling	14
2.2.1 Characterization of hydrocarbon reservoir fluids	17
2.2.2 Thermodynamic calculations	19
2.2.3 Code implementation validation	22
2.3 Oil PVT Experiments	26
3 Uncertainty in fluid model selection	30
3.1 Introduction	31
3.2 Literature Review	33
3.3 Methodology	36
3.3.1 Data description	36
3.3.2 Numerical implementation	36
3.3.3 Error metrics	37
3.4 Results	38

3.5	Proposed model for R_s	47
3.5.1	Numerical calculation example	51
3.5.2	Error analysis	52
3.6	Conclusion	56
4	Uncertainty in compositional fluid modeling	58
4.1	Introduction	59
4.2	Methodology	61
4.2.1	Problem formulation	61
4.2.2	Alternative minimization strategies	63
4.2.3	Numerical implementation	70
4.2.4	Data description	70
4.3	Results	72
4.3.1	Optimization algorithm	72
4.3.2	Initial guess	89
4.4	Discussion	91
4.5	Conclusions	93
5	Uncertainty in empirical fluid modeling	95
5.1	Introduction	96
5.1.1	Sources of uncertainty in flow simulations	97
5.1.2	Uncertainty quantification for flow simulations	100
5.1.3	Global sensitivity analysis	101
5.2	Methodology	102
5.2.1	Probabilistic characterization of uncertain fluid properties	102
5.2.2	Monte Carlo simulations	104
5.2.3	Global sensitivity analysis	104
5.2.4	Case study	105
5.3	Results	109
5.3.1	Monte Carlo simulations	109
5.3.2	Sensitivity Analysis	114
5.4	Conclusion	115

6 Uncertainty in flow modeling	119
6.1 Introduction	119
6.2 Methodology	122
6.2.1 Optimization problem formulation	123
6.2.2 Alternative optimization strategies	124
6.2.3 Case study	126
6.3 Proposed optimization algorithm	127
6.4 Results	130
6.5 Conclusions	135
7 Overall Conclusions and Future Work	139
7.1 Conclusions and Discussion	139
7.2 Future work	140
A Implemented R_s BO models	142

List of Tables

2.1	Molar composition of an intermediate gravity oil and C ₇₊ fractions properties (from Table 7.7 in [67])	23
2.2	Well-defined components properties (adapted from Table 5.2 in [67])	23
2.3	Non-zero binary interaction coefficients	24
2.4	Final characterized fluid mixture (reproduces Table 7.12 in [67])	27
3.1	Summary of the data set used to evaluate solution gas-oil ratio models	36
3.2	Summary of the R _s prediction errors for Group I	40
3.3	Summary of the R _s prediction errors for Group II	40
3.4	Summary of the R _s prediction errors for Group III	41
3.5	Summary of the R _s prediction errors for Group IV	41
3.6	Summary of the R _s prediction errors for Group V	42
3.7	Summary of the R _s prediction errors for Group VI	42
3.8	Summary of the R _s prediction errors for the test set	54
3.9	Proposed R _s model prediction errors, by group.	54
4.1	Properties of the reservoir fluid components (CO ₂ , N ₂ , and hydrocarbons up to C ₁₉ fraction) [89]	62
4.2	Molar composition and plus fraction properties of reservoir fluids A [133] and B [97]	71
4.3	EoS regression results obtained with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	76

4.4	EoS predictions and experimental data for the saturation pressure and the API gravity of the residual oil from the differential liberation at 104°C and 76°C for fluids A and B, respectively. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	80
4.5	EoS-regression results obtained via the two-step regression in which the optimal values of the decision variables obtained with ADAM [64], CMA-ES [55], or DSM [57] serve as the initial guess for the subsequent DFP minimization [33, 40].	89
5.1	Parameters and their numerical values used in the multiphase flow simulations.	108
5.2	Flow and black-oil constitutive relations used in the multiphase-flow simulations.	109
5.3	Descriptive statistics of the oil flow rate throughout production life. .	111
6.1	Boundary conditions from nine production tests.	128
6.2	Pressure and temperature measurements at the PDG, TPT, and outlet during nine production tests.	128
6.3	Comparison of computational time (normalized) for the tuning factors regression performed with different optimization algorithms: ADAM, DFP, CMA-ES, HJ, and proposed heuristic algorithm.	134

List of Figures

2.1	Representation of the Black Oil system. Reservoir fluid properties are computed at any pressure and temperature condition (a) from information obtained at standard conditions (b).	11
2.2	Splitting results: mole fraction and density distribution; validated against Figure 7.3 in [67], which is reproduced here in Figure 2.3. . .	24
2.3	Reproduction of Figure 7.3 in [67].	25
2.4	Splitting results: critical pressure and critical temperature; validated against Figure 7.4 in [67], which is reproduced here in Figure 2.5. . .	25
2.5	Reproduction of Figure 7.4 in [67].	26
2.6	Bubble point pressure determination through a constant composition expansion experiment	28
3.1	Experimental and predicted solution gas-oil ratio from a differential liberation for reservoir fluids representative of (a) Group I, (b) Group II, (c) Group III, (d) Group IV, (e) Group V, and (f) Group VI. Pressures are plotted below and at the bubble point.	39
3.2	Experimental R_s data and new p_f curve fitting	44
3.3	Pressure profile for a well+pipeline production system considering Standing's [118] and Lasater's [68] R_s models.	46
3.4	Temperature profile for a well+pipeline production system considering Standing's [118] and Lasater's [68] R_s models.	46
3.5	Gas volume fraction profile for a well+pipeline production system considering Standing's [118] and Lasater's [68] R_s models.	47

3.6	Root Mean Square Error (RMSE) of R_s predictions given by nine models, grouped by the CO_2 content in the surface gas of the reservoir fluid sample.	48
3.7	Root Mean Square Error (RMSE) of P_b predictions given by eight models, grouped by the CO_2 content in the surface gas of the reservoir fluid sample.	52
3.8	Experimental solution gas-oil ratio from a differential liberation and new model predictions for reservoir fluids representative of (a) Group I, (b) Group II, (c) Group III, (d) Group IV, (e) Group V, and (f) Group VI.	55
4.1	Convergence history of fluid A's normalized design variables ($P_{\text{cr},C_{\text{plus}}}^*$, $T_{\text{cr},C_{\text{plus}}}^*$, $\omega_{C_{\text{plus}}}^*$, $M_{C_{\text{plus}}}^*$, and $c_{C_{\text{plus}}}^*$) for alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	73
4.2	Convergence history of fluid B's normalized design variables ($P_{\text{cr},C_{\text{plus}}}^*$, $T_{\text{cr},C_{\text{plus}}}^*$, $\omega_{C_{\text{plus}}}^*$, $M_{C_{\text{plus}}}^*$, and $c_{C_{\text{plus}}}^*$) for alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	74
4.3	Objective function (squared relative error) evaluated for fluid B at convex combinations between the solutions obtained by the covariance matrix adaptation evolution strategy (CMA-ES) [55] ($\theta = 0$) and the direct search method (DSM) [57] ($\theta = 1$). The curve is generated using 200 objective function evaluations.	75
4.4	Convergence history of the objective function F for (a) fluid A and (b) fluid B, and four optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	77

4.5	Fluid A EoS predictions and experimental data at 104°C for (a) solution gas-oil ratio, (b) oil formation volume factor, (c) gas compressibility factor, (d) gas formation volume factor, (e) oil specific gravity, and (f) gas specific gravity. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	78
4.6	Fluid B EoS predictions and experimental data at 76°C for (a) solution gas-oil ratio, (b) oil formation volume factor, (c) gas compressibility factor, (d) gas formation volume factor, (e) oil specific gravity, and (f) gas specific gravity. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].	79
4.7	Phase envelopes (vapor liquid equilibrium) for the EoS models before and after regression of (a) fluid A and (b) fluid B. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57]. The markers indicate the critical point.	81
4.8	Comparison between our EoS implementation and <code>PVTsim</code> predictions for gas-oil solubility ratio of (a) Fluid A and (b) Fluid B.	82
4.9	Comparison between our EoS implementation and <code>PVTsim</code> predictions for oil formation volume factor of (a) Fluid A and (b) Fluid B.	83
4.10	Comparison between our EoS implementation and <code>PVTsim</code> predictions for gas compressibility factor of (a) Fluid A and (b) Fluid B.	84
4.11	Comparison between our EoS implementation and <code>PVTsim</code> predictions for gas formation volume factor of (a) Fluid A and (b) Fluid B.	85

4.12	Comparison between our EoS implementation and PVT <code>sim</code> predictions for oil density of (a) Fluid A and (b) Fluid B.	86
4.13	Comparison between our EoS implementation and PVT <code>sim</code> predictions for gas specific gravity of (a) Fluid A and (b) Fluid B.	87
4.14	Comparison between our EoS implementation and PVT <code>sim</code> predictions for phase envelopes of (a) Fluid A and (b) Fluid B.	88
4.15	Phase envelopes (vapor liquid equilibrium) obtained for (a) fluid A and (b) fluid B via the two-step EoS regression in which the optimal values of the decision variables obtained with ADAM [64], CMA-ES [55], or DSM [57] serve as the initial guess for the subsequent GD2O minimization [3]. The markers indicate the critical point.	90
4.16	Fitted probability density function (PDF) for the C_{plus} fraction properties: (a) critical pressure, (b) critical temperature, (c) acentric factor, (d) molecular weight, and (e) volume shift parameter.	92
5.1	Physical properties that provide input for a steady-state black-oil flow simulation.	99
5.2	Schematic of a simplified production well implemented in the flow simulator.	106
5.3	(a) Pressure, (b) temperature, and (c) liquid holdup behaviors at the steady-state regime in the wellbore and production line for watercut values of 0%, 40%, and 80%, representing the beginning, middle and end of the production lifetime, respectively.	107
5.4	Convergence analysis of MCS after 10 years of production: (a) sample mean, μ , and standard deviation, σ , of the cumulative production at the end of the well's lifetime as function of the number of MC realizations, N ; (b) the absolute difference between two sample standard deviations, $ \sigma(N) - \sigma(N - 100) $, computed from N and $N - 100$ MC realizations.	110
5.5	Probability density function of the oil production flow rate throughout the well's production lifetime, at several times \mathcal{T}	112

5.6	Cumulative oil production through the well's production lifetime considering as time step for the numerical integration of the flow rates: (a) one month and (b) one year.	113
5.7	First-order and total-order Sobol's indices for the input parameters (γ_0 , γ_g , GOR, WC) of flow simulation model (a) at the well opening, (b) after five years of production, and (c) after ten years of production.	116
6.1	Schematic of a simplified production well implemented in the flow simulator.	127
6.2	Convergence history of the tuning factors (ϕ_P^{well} , ϕ_P^{pipe} , ϕ_T^{well} , and ϕ_T^{pipe}) using: (a) proposed heuristic method (this work), (b) adaptive moment estimation (ADAM) [64], (c) Davidon-Fletcher-Powell method (DFP) [33, 40], (d) covariance matrix adaptation evolution strategy (CMA-ES) [55], and (e) Hooke-Jeeves (HJ) [57].	133
6.3	Convergence history of the Mean Absolute Percentage Error (MAPE) using the proposed heuristic method (this work), adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and Hooke-Jeeves (HJ) [57].	134
6.4	Convergence history of the tuning factors (ϕ_P^{well} , ϕ_P^{pipe} , ϕ_T^{well} , and ϕ_T^{pipe}) using the proposed heuristic for the initial guess of four optimization algorithms: (a) adaptive moment estimation (ADAM) [64], (b) Davidon-Fletcher-Powell method (DFP) [33, 40], (c) covariance matrix adaptation evolution strategy (CMA-ES) [55], and (d) Hooke-Jeeves (HJ) [57].	136
6.5	Convergence history of the Mean Absolute Percentage Error (MAPE) using the proposed heuristic for the initial guess of four optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and Hooke-Jeeves (HJ) [57].	137

A.1 Adaptation of Lasater's [68] original chart for the relationship between the stock-tank oil gravity ($^{\circ}$ API) and the effective molecular weight (M_o) superposed by the curve fittings proposed by Cragoe [31], Almeida [14], and Pipesim [®] v.2015	145
--	-----

Nomenclature

Abbreviations

ADAM Adaptive moment estimation method

ANOVA Analysis of variance

BO Black Oil

C_{plus} Hydrocarbon plus fraction

CCUS Carbon capture utilization and sequestration

CMA-ES Covariance matrix adaptation evolution strategy

DFP Davidon-Fletcher-Powell method

DSM Direct search method

EOR Enhanced oil recovery

EoS Equation of state

GSA Global sensitivity analysis

HM History Matching

IAM Integrated asset model

IPR Inflow performance relationship

LHS Latin Hypercube Sampling

MCS Monte Carlo Simulation

O&G Oil and Gas

PDF Probability density function

PDG Permanent downhole gauge

PVT Pressure-volume-temperature

QoI Quantity of Interest

RMSE Root mean square error

SCN Single carbon number

UQ Uncertainty quantification

WAG Water alternating gas

Superscripts

* Normalized variable

l Liquid

sc Standard conditions

v Vapor

Subscripts

cr Critical

b Bubble point

comp Components

d Diameter

dg Dissolved gas

fg Free gas

g Gas

m Mixture

meas Measurements

o Oil

Pen Peneloux correction

sat Saturation

Variables

μ Mean values vector

Σ Covariance matrix

γ Specific gravity

\hat{y} Predicted value

μ Viscosity

ω Acentric factor

Ω_a, Ω_b Cubic EoS parameters

σ Standard deviation

γ_{API} Stock-tank API gravity

\mathbf{x} Design variables vector

\mathbf{z} Composition vector of fluid mixture

\mathcal{E} Percentage error

ρ Density

a	Attraction parameter
B	Formation volume factor
b	Covolume parameter
c	Volume shift parameter
e	EoS estimate
F	Objective function
M	Molecular weight
N	Total number
P	Absolute pressure
p	Relative pressure
R	Universal gas constant
R_s	Solution gas-oil ratio
T	Temperature
V	Volume
v	Molar volume
w	Weight
y_{exp}	Experimental value
Z	Gas compressibility factor
BIP	Binary interaction parameter
GOR	Reservoir gas-oil ratio
iter	Iterations
WC	Watercut

Chapter 1

Introduction

Fluid models play a critical role in a wide range of applications, including the exploration and production of oil and gas resources, chemical process design and optimization, and CO₂ enhanced oil recovery projects. Accurate predictions of phase behavior are crucial in these applications, as they directly affect the operations' performance and profitability.

For instance, in reservoir fluid modeling, the estimation of the amount of oil and gas that can be recovered from a reservoir depends on the prediction of fluid properties and behavior under various conditions. Similarly, reliable predictions of thermodynamic properties and phase behavior of fluid mixtures are essential in chemical processes. In addition, CO₂ enhanced oil recovery projects depend on accurate predictions of phase behavior to optimize CO₂ injection and increase oil recovery efficiency.

Nonetheless, for complex fluid systems, uncertainty permeates the fluid modeling effort. This dissertation employs hydrocarbon reservoir fluids as a vehicle for exploring uncertainty in fluid modeling, with the understanding that the methodologies and analyses presented herein can be seamlessly extended to other fluid mixtures. Hydrocarbon reservoir fluids are arguably the most compositionally complex organic mixture [71], and thus serve as a nice example. They are a mixture of various components, including light, medium, and heavy hydrocarbon fractions, as well as nonhydrocarbon compounds. These components can exhibit behavior above their

critical pressure and temperature, requiring sophisticated modeling approaches that go beyond classical fluid mechanics [51].

Uncertainty is interpreted here in its broader sense, i.e. as the lack of certainty due to imperfect, incomplete or unknown information. It applies to predictions of future events, to physical measurements, to mathematical models predictions, or to any parameter whose value is not known “for sure”. Uncertainty in fluid modeling arises from the limitations of our ability to fully comprehend or model the complex physics and chemistry of fluid systems.

BO fluid models, for instance, rely on equations that were fit to a specific data set, which can lead to significant errors when applied to different fluids or operating conditions. In compositional fluid models, uncertainties arise from the lack of knowledge of the fluid’s full composition, components parameters, and interaction parameters. The properties of the heaviest hydrocarbon fraction are often obtained from BO models or estimated through regression techniques, which can propagate through the models and lead to large predictive uncertainties in the outputs, including phase behavior and transport properties.

Thus, fluid behavior predictions should be accompanied by a measure of their uncertainty. Put simply, the predictive uncertainty of the fluid behavior must be assessed and quantified to enable more informed decisions. Such uncertainty quantification effort aims to estimate the effects of epistemic uncertainty, which comprises both structural uncertainty (uncertainty about the accuracy of a particular model) and parametric uncertainty (uncertainty about parameters and driving forces in a particular model) [122]. Epistemic uncertainty can be reduced by collecting more data. Conversely, aleatory or irreducible uncertainty refers to the limits of predictability of inherently random phenomena and therefore cannot be reduced [122].

While some level of uncertainty is unavoidable, it is crucial to quantify epistemic uncertainties to make more informed decisions. Neglecting to account for uncertainties can lead to false confidence in the models’ predictions and ultimately result in poor decision-making. Therefore, accurate modeling of hydrocarbon reservoir fluids requires careful consideration of these uncertainties to properly quantify and manage their impact on fluid and flow model predictions.

Predictions obtained from fluid models are often used to inform critical business decisions that can have significant financial implications, potentially worth millions or billions of dollars. Furthermore, these models may impact the safety and environmental sustainability of the systems being designed, underscoring the importance of obtaining accurate and reliable predictions. Ultimately, a failure to account for uncertainties in these models may pose human and environmental risks, highlighting the critical need for effective uncertainty quantification methods.

Therefore, this dissertation investigates the sources of epistemic uncertainty associated with reservoir fluid modeling, both in Black Oil (BO) and compositional approaches, and provides a better understanding of their impact on fluid and flow model predictions. The dissertation emphasizes the importance of quantifying predictive uncertainties of fluid models and proposes novel methods for their estimation. These methods are demonstrated using real-world data and case studies.

1.1 Objectives

The main objectives of this dissertation are:

- To investigate the sources of uncertainty associated with reservoir fluid modeling, both in BO and compositional approaches, and to provide a better understanding of their impact on fluid and flow models predictions.
- To emphasize the importance of quantifying predictive uncertainties of fluid models and to propose novel methods for their estimation.
- To demonstrate the proposed methods using real-world data and case studies.

These overall objectives are broken into the following specific objectives:

- To investigate the applicability of traditional solution gas-oil ratio models on an experimental data set containing CO₂-rich reservoir fluids and propose a more accurate model that is suitable for these types of fluids.

- To demonstrate the impact of inaccurate solution gas-oil ratio predictions on the multiphase flow simulation of a production system.
- To investigate the uncertainty arising from the equation of state (EoS) regression process due to the choices of optimization algorithm and initial guess and its impact on EoS predictions.
- To propose the probabilistic treatment of EoS predictions.
- To develop a framework for global sensitivity analysis in multiphase flow simulation results to identify the model inputs contributing most to the predictive uncertainty.
- To develop a framework for quantifying the predictive uncertainty of multiphase pipe-flow models due to correlated random inputs.

1.2 Dissertation outline

This dissertation investigates uncertainties in reservoir fluid modeling, focusing on both BO and compositional approaches. The following chapters present original contributions to this field, addressing specific sources of uncertainty and proposing novel methods to quantify and manage them.

Chapter 1 provides an introduction to the research, outlining the motivations and objectives of this dissertation.

Chapter 2 reviews the theoretical basis for reservoir fluid modeling, covering the two main approaches: BO and compositional. The BO model is empirical, simpler and computationally cheaper, while the compositional approach is based on thermodynamic relations and is therefore more theoretically robust. The concepts covered in this chapter are central to this research.

Chapter 3 investigates the uncertainty associated with the selection of solution gas-oil ratio (R_s) models. Many models are available in the literature, but their accuracy varies depending on fluid compositions and operational conditions. For some compositions, none of the models are accurate enough. To address this issue, we

propose a generalist R_s model that is simple and ready for implementation, and shows overall better performance on the training and test sets. Chapter 3 is a compilation of two papers published in [41] and [42].

Chapter 4 investigates predictive uncertainties from compositional fluid models. Specifically, we explore the equation of state (EoS) tuning to experimental data, which consists on a nonlinear and non-convex multivariate optimization problem. We demonstrate how different optimization algorithms and initial guesses can significantly affect the predictive capabilities of the resulting models in unsampled pressure and temperature regions. To address this issue, we propose a probabilistic treatment of the design variables to quantify EoS predictive uncertainty. Chapter 4 reproduces a paper accepted by the Geoenergy Science and Engineering journal.

Chapter 5 discusses the uncertainty associated with fluid models inputs and their influence on the results of flow simulations of a production system. The BO approach is used for the fluid modeling. We present a framework for quantifying the predictive uncertainty of multiphase pipe-flow models due to correlated random inputs. We demonstrate the framework's effectiveness through a case study, which quantifies the uncertainty of cumulative production for an oil reservoir whose fluid properties are unknown during the exploration phase. Global Sensitivity Analysis through Sobol's indices is performed to identify the model inputs contributing the most to predictive uncertainty. Chapter 5 reproduces a paper published in [44].

Chapter 6 investigates the calibration of pipe-flow models, addressing the inherent uncertainties present in these models. In the oil and gas industry, it is a common practice to use tuning factors to match field data and enhance the predictive capability of flow models. In particular, we propose a heuristic method tailored to optimize these tuning factors applied to calculated pressure and temperature gradients. A real case study demonstrates the proposed heuristic efficiency. Chapter 6 reproduces a paper to be submitted.

Chapter 7 summarizes the main conclusions of this dissertation and offer suggestions for future work.

Chapter 2

Review of reservoir fluid modeling

Reservoir fluid modeling plays a pivotal role in understanding and predicting the behavior of thermophysical of hydrocarbons mixtures within subsurface reservoirs, as well as during production, transport, and processing. Accurate fluid modeling is essential for making informed decisions related to reservoir management, production strategies, and facilities design, among other areas. This chapter provides a review of the two main approaches for reservoir fluid modeling: the Black Oil (BO) and the compositional. Both approaches have advantages and limitations, which are discussed in the following sections. The BO fluid modeling is used in Chapters 3 and 5, while the compositional approach is employed in Chapter 4.

2.1 Black Oil fluid modeling

The Black Oil approach for fluid modeling, referred here as the BO approach, consists of simple empirical equations for reservoir fluid properties proposed based on experimental or field data. In today's terms, it is a data-driven approach.

The primary advantage of the BO approach is its computational efficiency. Due to the simplified representation of the fluid and the low number of required parameters, black oil simulations can be carried out relatively quickly, making it a popular choice for flow simulations requiring low computational costs.

The main limitation of the BO approach is that its models should be used only

within the range of validity for which they were developed, i.e., for fluid compositions, pressure, and temperature conditions that align with those included in the dataset used for a model's development. Nonetheless, it is common for such models to be used indiscriminately. Extrapolations can be dangerous and lead to unphysical results, as demonstrated in Chapter 3.

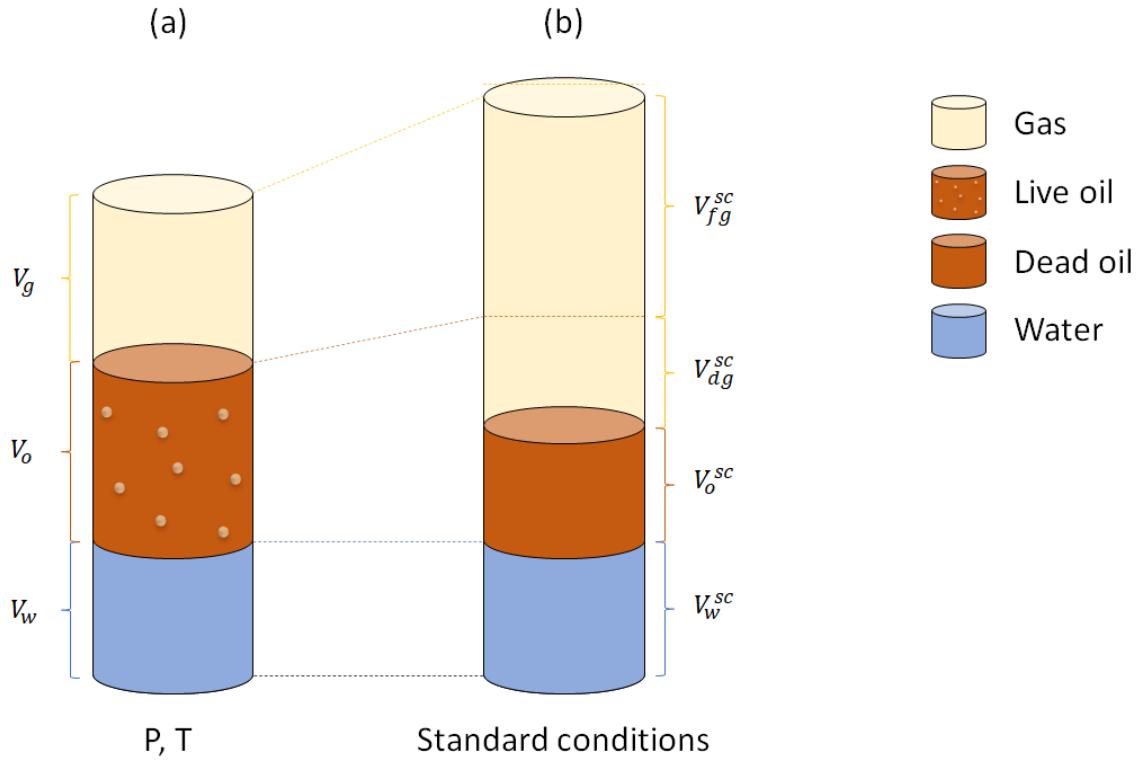


Figure 2.1: Representation of the Black Oil system. Reservoir fluid properties are computed at any pressure and temperature condition (a) from information obtained at standard conditions (b).

The BO approach characterizes each phase that can be present in a reservoir fluid mixture (oil, gas, and water) through simple inputs, typically obtained at standard conditions. Essentially, this approach can be conceived of as a compositional model, wherein each phase is represented by one pseudocomponent. The foundational premises of the BO approach are concisely detailed as follows:

- Component constitution: The system is structured around three distinct components, namely oil, gas, and water components. Each component is defined under standard conditions.
- Phase partitioning: The system accommodates the existence of three unique phases, namely oil, gas, and water phases.
- Solubility dynamics: The gas component can be dissolved in the oil phase, but not in the water phase. The oil and water components are strictly confined to the oil and water phases, respectively.
- Thermodynamic equilibrium: The system instantaneously achieves a state of thermodynamic equilibrium.

Figure 2.1 provides a visualization of the BO system. The diagram illustrates the relationship between the volumes (V) of oil (o), gas (g), and water (w) phases at a specific pressure (P) and temperature (T) condition (Figure 2.1a) and under standard conditions (sc) (Figure 2.1b). In Figure 2.1a, the system is at a pressure below the bubble point and a temperature near the reservoir temperature, hence the presence of the gas phase with volume V_g . The oil phase contains dissolved gas, thus being a “live oil”, and has volume V_o . The water phase volume is denoted by V_w . As this closed system transitions to standard conditions, a change in volumes is noticed. The free gas (fg) phase expands to a volume of V_{fg}^{sc} . The dissolved gas (dg) in V_o is liberated, its volume at standard conditions is V_{dg}^{sc} . Therefore, the gas phase total volume under standard conditions is given by $V_g^{sc} = V_{fg}^{sc} + V_{dg}^{sc}$. Meanwhile, the volume of the oil and water phases under standard conditions are denoted by V_o^{sc} and V_w^{sc} , respectively. Since there is no dissolved gas in the oil phase under standard conditions, it is called “dead oil”.

The phase behavior of the hydrocarbon fluid mixture is computed from information obtained at standard conditions through BO properties:

- Solution gas-oil ratio (R_s) - represents the volume of dissolved gas at standard conditions divided by the respective volume of stock-tank oil at standard conditions:

$$R_s = \frac{V_{dg}^{sc}}{V_o^{sc}}. \quad (2.1)$$

For a constant temperature, as pressure decreases, more gas is released from the oil, which is crucial to understanding the fluid's multiphase behavior.

- Oil formation volume factor (B_o) - represents the volume of oil with dissolved gas at a specific pressure and temperature condition divided by its respective volume of stock-tank oil at standard conditions:

$$B_o = \frac{V_o}{V_o^{sc}}. \quad (2.2)$$

B_o accounts for the expansion or contraction of oil due to pressure and temperature changes (including the release or dissolution of gas).

- Gas compressibility factor (Z) - represents the deviation of real gas behavior from ideal gas behavior under specific temperature and pressure conditions:

$$Z = \frac{Pv}{RT}. \quad (2.3)$$

The oil density (ρ_o) is computed from B_o and R_s at any pressure and temperature condition according to:

$$\rho_o = \frac{\rho_o^{sc} + \rho_g^{sc} R_s}{B_o}. \quad (2.4)$$

The gas formation volume factor (B_g) is obtained from Z:

$$B_g = \frac{V_g}{V_{fg}^{sc}} = \frac{P^{sc} T}{P T^{sc}} Z. \quad (2.5)$$

Finally, the gas density is computed from B_g at any pressure and temperature condition:

$$\rho_g = \frac{\rho_g^{sc}}{B_g}. \quad (2.6)$$

Typically, the oil and gas phase densities at standard conditions and the reservoir gas-oil ratio provide the essential inputs to the BO equations that characterize hydrocarbon fluid mixtures across a wide range of pressures and temperatures. Some BO models might include additional inputs, such as the CO₂ mole fraction of surface gas [41, 48]. Pressure and temperature at which fluid properties are computed are also required as inputs.

The generalized Black Oil approach relaxes some modeling premises of the BO approach. The interested reader is referred to its description in [16].

2.2 Compositional fluid modeling

The compositional approach provides a more detailed representation of the fluid's components and their interactions using equations of state (EoS) based on thermodynamic theory. In particular, this approach relies on the principle that fugacities of each component must be the same across all phases at equilibrium.

EoS models are widely employed in the oil and gas industry to predict the phase behavior of hydrocarbon fluids. The main advantage of the compositional approach lies in its rigorous theoretical foundation. As the EoS connects thermodynamic theory to the mechanical state of systems of interest [67], this approach offers theoretical robustness to the fluid modeling. Moreover, the compositional approach provides flexibility in modeling a variety of reservoir fluids. It can handle natural gases, crude oils, condensates, volatile oils, and even complex mixtures involving non-hydrocarbon components, such as CO₂, H₂S, and N₂. This is particularly important when compared to the BO approach, in which models are limited to their (sometimes small) validity range.

The most pronounced disadvantage of the compositional approach is its computational cost. It requires solving a series of non-linear equations to obtain phase equilibrium, which can be time-consuming particularly for mixtures with a large number of components. Additionally, this approach requires knowledge of the fluid mixture composition, a piece of information that may not always be readily accessible. Despite its theoretical basis, empirical equations may be necessary to describe partially characterized components such as the C_{plus} fraction. In such cases, the compositional approach is considered to be semi-empirical. Lastly, the EoS model often necessitates fine-tuning to align with experimental data, a procedure referred to as EoS regression. As revealed in Chapter 4, this regression process can yield disparate fluid models, depending on subjective choices of an optimization algorithm and an initial estimate.

An EoS can be defined as a relationship between pressure (P), temperature (T), and volume (V) that describes the thermophysical state of a fluid system, such that $\mathcal{F}(P, T, V) = 0$. Originally proposed for pure components, EoS models were later extended to accommodate fluid mixtures through the use of mixing rules.

The van der Waals EoS (vdW) [126] was the first to predict accurately vapor-liquid equilibrium by incorporating corrections in the ideal gas equation of state, $PV = nRT$. These corrections aimed to account for the size of real gas molecules and the presence of intermolecular forces. They depend on empirical constants determined experimentally for each substance and require mixing rules for fluid mixtures.

Numerous authors have proposed modifications to the original vdW equation, with those proposed by Soave, Redlich, and Kwong (SRK) [115] and Peng and Robinson (PR) [96] being the most popular in the oil and gas industry. These modified EoS models are part of a class known as cubic EoS, as they involve cubic equations in terms of the compressibility factor $Z = Pv/RT$. Cubic EoS models can be represented by the following general formula:

$$Z^3 - (1 + B - uB)Z^2 + (A + wB^2 - uB - uB^2)Z - (AB + wB^2 + wB^3) = 0, \quad (2.7a)$$

where

$$A = \frac{aP}{R^2 T^2}, \quad (2.7b)$$

$$B = \frac{bP}{RT}, \quad (2.7c)$$

a and b are the attraction and covolume parameters, respectively, u and w are constants specific to each cubic EoS model, and R is the universal gas constant. For fluid mixtures, the attraction and covolume parameters are expressed as a_m and b_m , respectively, and are calculated via mixing rules derived from the parameters a_i and b_i of individual components (as exemplified in Eqs. 2.7g and 2.7h).

In particular, we implement the modified Peng-Robinson EoS [96, 105] in Chapter 4 because of its widespread usage in modeling reservoir fluids and simplicity. For this EoS, $u = 2$ and $w = -1$ in Eq. 2.7a. The attraction and covolume parameters of the i th component ($i = 1, \dots, N_{\text{comp}}$) are computed from the acentric factor ω_i as

$$a_i = \Omega_a \frac{R^2 T_{\text{cr},i}^2}{P_{\text{cr},i}} \left[1 + m_i \left(1 - \frac{T}{T_{\text{cr},i}} \right)^{1/2} \right]^2, \quad (2.7d)$$

$$b_i = \Omega_b \frac{RT_{\text{cr},i}}{P_{\text{cr},i}}, \quad (2.7e)$$

$$m_i = \begin{cases} 0.37464 + 1.54226\omega_i - 0.26992\omega_i^2, & \omega_i \leq 0.49 \\ 0.379642 + 1.48503\omega_i - 0.164423\omega_i^2 + 0.016666\omega_i^3, & \omega_i > 0.49, \end{cases} \quad (2.7f)$$

with $\Omega_a = 0.45724$ and $\Omega_b = 0.07780$.

The attraction, a_m , and covolume, b_m , parameters are computed for the fluid mixture of N_{comp} components with composition $\mathbf{z} = (z_1, \dots, z_{N_{\text{comp}}})^\top$ according to the mixing rules

$$a_m = \sum_{i=1}^{N_{\text{comp}}} \sum_{j=1}^{N_{\text{comp}}} z_i z_j (a_i a_j)^{1/2} (1 - \text{BIP}_{ij}), \quad (2.7g)$$

$$b_m = \sum_{i=1}^{N_{\text{comp}}} z_i b_i, \quad (2.7h)$$

where BIP_{ij} is the binary interaction parameter (BIP) between the i th and j th components.

For given P and T , the molar volume v predicted with Eq. (2.7) is shifted to v_{Pen} according to [95, 62]

$$v_{\text{Pen}} = v - \sum_{i=1}^{N_{\text{comp}}} z_i c_i, \quad (2.8)$$

where c_i is the volume shift parameter for the i th component. Finally, density can be computed based on the components molecular weights.

Other equations of state follow a different approach, and are known as non-cubic. These are out of the scope of this work. For a thorough overview, [66] discusses the main equations of state currently in use.

2.2.1 Characterization of hydrocarbon reservoir fluids

The hydrocarbon reservoir fluid mixture represents a complex fluid system that challenges classical fluid mechanics in accurately modeling its behavior [51]. These naturally occurring mixtures of hydrocarbons contain hundreds or thousands of components that cannot all be identified [23]. The fluid mixture is characterized by multiphase, multicomponent interactions between various hydrocarbon components (light, medium, and heavy) and nonhydrocarbon components, which may exist in supercritical states depending on pressure and temperature conditions. For example, the critical point of pure CO₂ is 1070 psi and 87.8°F, indicating that in certain production or injection systems, CO₂ may exist in a supercritical state within the pressure-temperature (P,T) space of interest. Furthermore, the characterization of the heaviest hydrocarbon fractions is often performed with limited information. Such heaviest hydrocarbon fractions are typically grouped into a C_{plus} fraction.

In order to model effectively and analyze hydrocarbon reservoir fluids, it is essential to address the challenges posed by their complex nature. One approach to simplifying this complexity involves the application of splitting and lumping procedures, that are described below.

Pseudoization procedure

Accurate phase behavior simulations in equations of state (EoS) calculations for hydrocarbon mixtures often face challenges due to the numerous fractions required to describe the mixture [5]. The computational cost and resources needed for compositional reservoir simulation increase significantly with the number of components used to represent the reservoir fluid [5]. Consequently, property estimation of the fluid is typically based on a limited number of compounds, referred to as pseudocomponents. This simplified representation may not necessarily correspond to the actual mixture.

Also, numerous authors argue that splitting the heaviest hydrocarbon fraction is essential for improved characterization [27, 92, 132], as treating this fraction as a single pseudocomponent may result in significant errors in EoS predictions of pressure-volume-temperature (PVT) behavior [67]. Splitting consists of describing the molar distribution and properties of a mixture of components, which were grouped together into a single pseudocomponent. Various methods have been proposed for distributing the molar content of the heaviest hydrocarbon fraction into a more detailed description.

Pedersen et al. [93] recommend extending the plus fraction up to C₂₀₀ to make sure all carbon number fractions present in significant amounts are accounted for. However, due to the large number of components, the computational cost of EoS computation becomes increasingly prohibitive. As a result, lumping procedures are required. Lumping involves grouping components together into pseudocomponents. Alternative lumping methods are available in the literature.

The determination of pseudocomponents and use of an appropriate model to describe a mixture by a certain number of pseudocomponents is an engineering art in the prediction of properties of petroleum mixtures [103]. Riazi et al. [104] show that different C₇₊ characterization methods and number of C₇₊ fractions can have significant impact on the estimation of physical properties and phase behavior. Even after the splitting and lumping procedures, EoS predictions may not align with experimental data. The discrepancy may stem from experimental data inputs or equation of state parameters [76].

Equation of State regression

In order to improve the accuracy of EoS predictions, tuning procedures are commonly recommended. The EoS regression consists of fitting the EoS model to experimental data, i.e., solving a nonlinear, non-convex, multivariate optimization problem. The latter requires one to select design variables, domains of each variable, and weights assigned to individual measurements. The design variables can be chosen from a broad range of parameters, including those associated with splitting and lumping, critical properties, binary interaction coefficients, acentric factors, among others. Several EoS regression procedures have been proposed in the literature (e.g., [30, 133, 27, 4, 8]).

Tuning EoS models to the experimental PVT data can be more of an art than a science, and it requires the use of appropriate software programs [36]. This is at least partially a result of the EoS models being highly nonlinear and the number of adjustable parameters in the regression being large. Chapter 4 describes the EoS regression and demonstrates that despite being more robust and theoretically sound than the BO approach, the compositional approach has some degree of empiricism.

2.2.2 Thermodynamic calculations

Vapor-liquid equilibria

For each i th component of the mixture with mole fraction z_i , the Rachford-Rice equation (Eq. 2.9) establishes a condition for vapor-liquid equilibrium:

$$\sum_i \frac{z_i(1 - K_i)}{K_i + (1 - K_i)l} = 0, \quad (2.9)$$

where

$$K_i = \frac{y_i}{x_i}, \quad (2.10)$$

y_i and x_i are the i th component's mole fractions in the vapor and liquid phases, respectively, and l is the liquid mole fraction of the mixture. The first K-value K_i in

a vapor-liquid equilibria calculation is typically guessed using an empirical correlation. Wilson's equation [134] is typically used for this initial guess. Recently, neural networks have also been used for calculating initial K-values [141].

Eq. 2.9 is solved iteratively for l . If the fluid is in the two-phase region, the EoS provides the molar volumes of the vapor and liquid mixtures (if both phases are present), v^v and v^l , at the specified pressure and temperature condition. Finally, fugacities of the i th component in the vapor and liquid phases, \hat{f}_i^v and \hat{f}_i^l , are computed.

Employing the general form of the cubic equation of state in Eq. 2.7a, the fugacity coefficient ϕ can be computed after solving the EoS according to Eq. 2.11 (recall that $u = 2$ and $w = -1$ for the modified Peng-Robinson EoS [96, 105]).

$$\begin{aligned} \ln \phi_i = & \beta_i(Z - 1) - \ln(Z - B) \\ & + \frac{A}{B\sqrt{u^2 - 4w}} (\beta_i - \delta_i) \ln \frac{2Z + B(u + \sqrt{u^2 - 4w})}{2Z + B(u - \sqrt{u^2 - 4w})}, \end{aligned} \quad (2.11)$$

where A and B are defined in Eqs. 2.7b and 2.7c respectively, while

$$\beta_i = \frac{T_{ci}/P_{ci}}{\sum_j x_j T_{cj}/P_{cj}}, \quad (2.12)$$

$$\delta_i = \frac{2\sqrt{a_i}}{a} \sum_j x_j \sqrt{a_j} (1 - \text{BIP}_{ij}). \quad (2.13)$$

A noteworthy point is that the liquid composition is being utilized in Eqs. 2.12 and 2.13. However, when determining the vapor fugacity, the vapor composition, y_i , should be implemented in place of the liquid composition, x_i .

Fugacities in both phases are obtained from fugacity coefficients according to

$$\hat{f}_i^v = \phi_i^v y_i P \quad (2.14a)$$

$$\hat{f}_i^l = \phi_i^l x_i P \quad (2.14b)$$

The system is in equilibrium when $\hat{f}_i^v = \hat{f}_i^l$ for all components. If the system

is not in equilibrium for the current liquid mole fraction value, new K-values are estimated and the entire process is repeated. A more detailed explanation of the vapor-equilibrium calculation can be found in [67].

Phase envelope

The phase envelope construction consists of a series of bubble point and dew point calculations. They are particular cases of the general flash calculation procedure, since the composition of either the liquid or vapor phase and the liquid molar fraction are known. Therefore, simplifications can be done to the Rachford-Rice equation (Eq. 2.9). At the bubble point ($x_i = z_i$ and $l = 1$), it becomes

$$\sum_i K_i z_i - 1 = 0 \quad (2.15)$$

At the dew point ($y_i = z_i$ and $l = 0$),

$$\sum_i \frac{z_i}{K_i} - 1 = 0 \quad (2.16)$$

Dew and bubble points are calculated by solving Eqs. 2.15 and 2.16, respectively, at specified pressures or temperatures around the phase envelope. However, the number of solutions for a given initial state is not known in advance; there may be a single solution, multiple solutions or no solution. Particularly at high pressures, the possibility of multiple solutions and convergence to the so-called “trivial” solution make the saturation point calculation more complex. Such complexities have been reported in the literature (e.g., [77, 80]). Thus, carefully chosen initial guesses are required for convergence to the correct solution.

Several procedures for the phase envelope construction have been proposed (e.g., [80, 130, 142]). Zier vogel and Poling’s [142] procedure is used in this work. It consists of calculating bubble and dew point points by iterating on a single variable - pressure or temperature, depending on the value of $|d(\ln P)/d(\ln T)|$ at each point. The critical point is located by the intersection of the bubble and dew point curves.

PVT properties

“PVT properties” is a term commonly used to describe the volumetric behavior of reservoir fluids as functions of pressure and temperature [88]. PVT experiments on reservoir fluids offer valuable insights into the volumetric changes that occur within the reservoir, wells, pipelines, and processing plants. Experimental data is essential for validating fluid models.

PVT properties can be calculated from an EoS by performing a series of vapor-liquid equilibrium calculations. This dissertation focuses on PVT properties measured in a differential liberation experiment, which is described in Section 2.3.

2.2.3 Code implementation validation

We validate our implementation of the aforementioned thermodynamic equations against results in [67]. The fluid mixture original composition is given in Table 2.1 along with the molecular weight and the density for the C₇₊ fractions. The benchmark case consists of the characterization and phase envelope construction for an intermediate gravity crude oil. We compare our results in terms of the splitted and lumped composition properties and phase envelopes. The goal is to demonstrate that our code is able to reproduce accurately the reference solution, and therefore can be used for the upcoming analyses.

Properties of well-defined components (C₁ through C₆ and nonhydrocarbons) are reported in Table 2.2. Properties of the heavy end fractions (C₇ to C₁₁₊) are calculated according to [87]. Binary interaction coefficients are zero, except if specified otherwise in Table 2.3. The SRK equation of state is used.

The C₁₁₊ plus fraction is expanded up to C₂₀₀ according to Pedersen et al.’s [93] procedure. Figures 2.2 and 2.4 compare the molar fraction, specific gravity, critical pressure and temperature distribution along the fractions carbon numbers against the reference plots in Figures 2.3 and 2.5.

Table 2.1: Molar composition of an intermediate gravity oil and C₇₊ fractions properties (from Table 7.7 in [67])

Component	Mole fraction (%)	Molecular weight (g/mol)	Density (g/cm3)
CO ₂	0.45		
N ₂	1.64		
C ₁	45.85		
C ₂	7.15		
C ₃	6.74		
i-C ₄	0.84		
C ₄	3.11		
i-C ₅	1.03		
C ₅	1.65		
C ₆	2.52		
C ₇	3.77	92	0.7294
C ₈	4.28	106	0.7509
C ₉	2.70	120	0.7739
C ₁₀	1.69	137	0.7835
C ₁₁₊	16.58	288	0.8835

Table 2.2: Well-defined components properties (adapted from Table 5.2 in [67])

Component	Critical pressure (bar)	Critical temperature (K)	Acentric factor (-)	Molecular weight (g/mol)
CO ₂	73.8	304.1	0.239	44.01
N ₂	33.9	126.2	0.039	28.013
C ₁	46.0	190.40	0.011	16.043
C ₂	48.8	305.4	0.099	30.070
C ₃	42.5	369.8	0.153	44.094
i-C ₄	36.5	408.2	0.183	58.124
C ₄	38.0	425.2	0.199	58.124
i-C ₅	33.8	460.4	0.227	72.15
C ₅	33.7	469.7	0.251	72.151
C ₆	30.1	507.5	0.299	86.178

Table 2.3: Non-zero binary interaction coefficients

	N ₂	CO ₂
C ₁	0.02	0.12
C ₂	0.06	0.15
C ₃	0.08	0.15
C _{n>3}	0.08	0.15

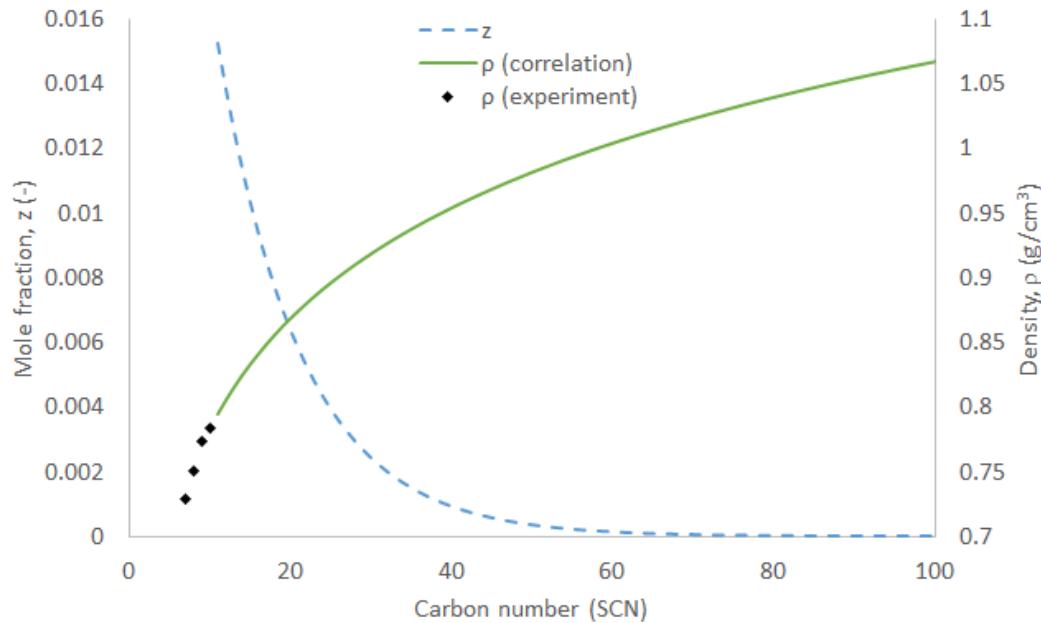


Figure 2.2: Splitting results: mole fraction and density distribution; validated against Figure 7.3 in [67], which is reproduced here in Figure 2.3.

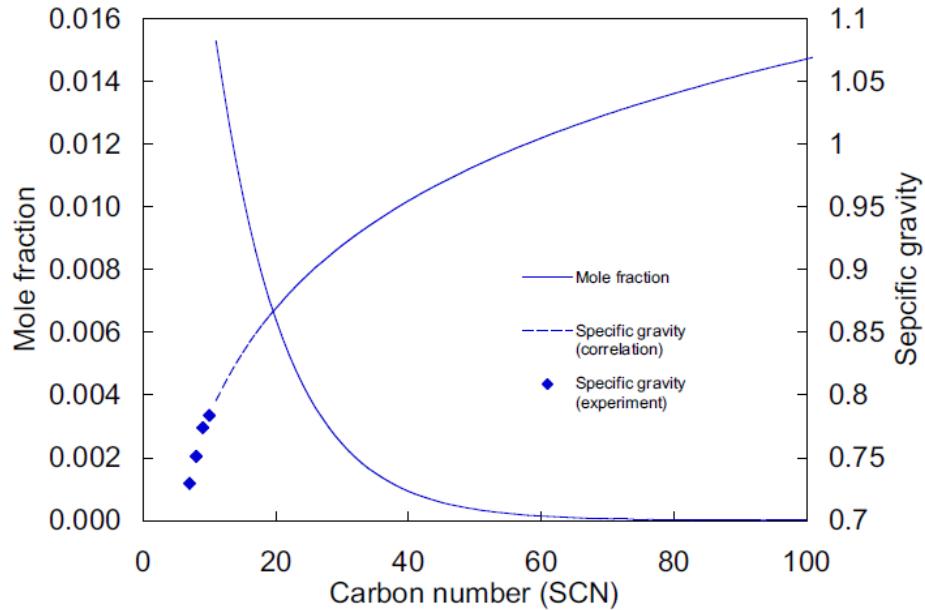


Figure 2.3: Reproduction of Figure 7.3 in [67].

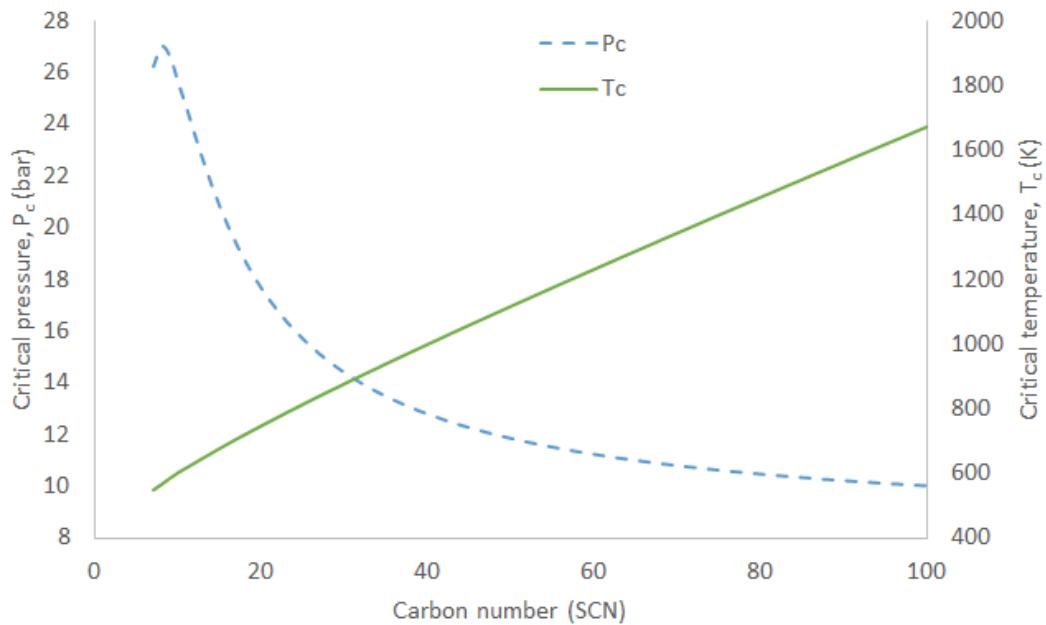


Figure 2.4: Splitting results: critical pressure and critical temperature; validated against Figure 7.4 in [67], which is reproduced here in Figure 2.5.

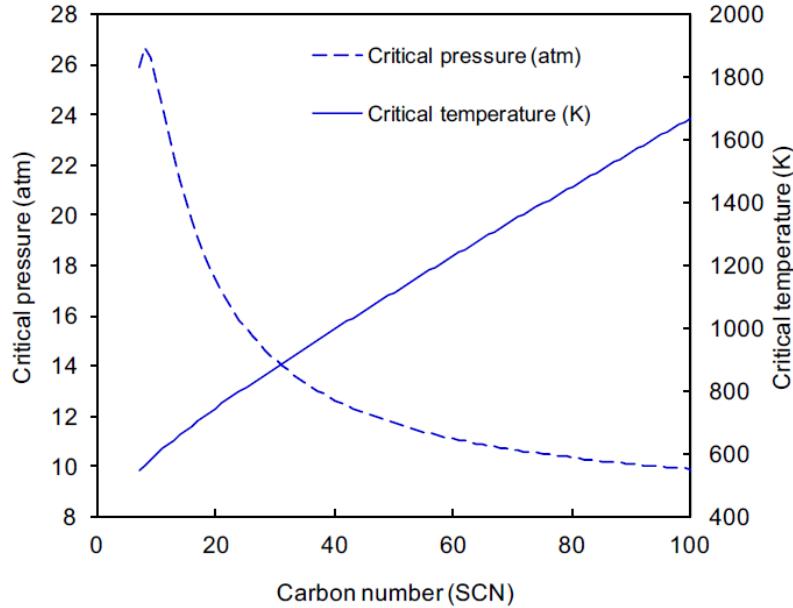


Figure 2.5: Reproduction of Figure 7.4 in [67].

C_{7+} components are lumped into four pseudocomponents. Group properties are computed through mass fraction weighting, as recommended by [87]. Table 2.4 reports the final fluid characterization. The four pseudocomponents are referred to as PS1, PS2, PS3, and PS4. Their molar fractions and properties match those in [67].

Phase envelopes and predictions for PVT properties measured in a differential liberation are validated against commercial simulator `PVTsim` results in Section 4.3.1.

2.3 Oil PVT Experiments

Standard PVT experiments performed on reservoir oils usually include constant composition expansion, differential liberation, and multistage-separator tests [133].

In the constant composition expansion experiment, a specialized cell is filled with a known mass of the fluid mixture sample. The experiment is carried out under a constant temperature. First, the cell is pressurized to a pressure above the saturation pressure, ensuring that the fluid is single phase. The cell's volume is then increased, resulting in a reduction in pressure. Pressure and cell volume measurements are

Table 2.4: Final characterized fluid mixture (reproduces Table 7.12 in [67])

Component	Mole fraction (%)	Critical pressure (bar)	Critical temperature (K)	Acentric factor (-)	Molecular weight (g/mol)
CO ₂	1.64	73.8	304.1	0.239	44.01
N ₂	0.45	33.9	126.2	0.039	28.013
C ₁	45.85	46.0	190.40	0.011	16.043
C ₂	7.15	48.8	305.4	0.099	30.070
C ₃	6.74	42.5	369.8	0.153	44.094
i-C ₄	0.84	36.5	408.2	0.183	58.124
C ₄	3.11	38.0	425.2	0.199	58.124
i-C ₅	1.03	33.8	460.4	0.227	72.15
C ₅	1.65	33.7	469.7	0.251	72.151
C ₆	2.52	30.1	507.5	0.299	86.178
PS1	13.97	26.2	577.91	0.401	113.49
PS2	7.4	21.1	674.22	0.647	206.30
PS3	5.01	16.4	792.01	0.902	318.72
PS4	2.64	12.8	1026.8	0.999	554.00

recorded at each step of volume expansion and subsequently plotted (as in Figure 2.6). The bubble point pressure (p_b) and volume can be read from the intersection of the near-linear and non-linear segments of the curve, respectively representing the single-phase and two-phase regions.

The differential liberation, or differential depletion, begins by transferring reservoir fluid to a cell maintained at a constant temperature, often the reservoir temperature, and a pressure above the reservoir pressure. The pressure is reduced until the saturation pressure is achieved. The sample is then equilibrated at the first selected pressure below the saturation pressure. Subsequently, the free gas is removed from the cell at a constant pressure, and its composition is measured, allowing for the calculation of the remaining liquid phase composition in the cell. These steps are repeated for further pressure reduction steps until the desired range of pressures has been explored.

Primary results from a differential liberation experiment performed on an oil mixture are: solution gas-oil ratio (R_s), oil formation volume factor (B_o), gas formation

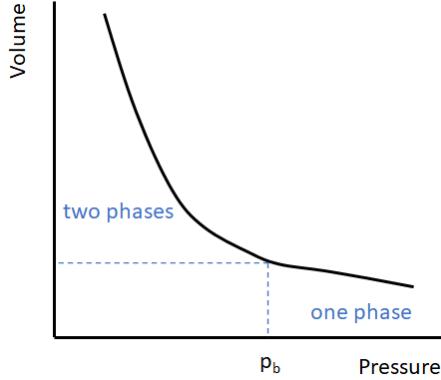


Figure 2.6: Bubble point pressure determination through a constant composition expansion experiment

volume factor (B_g), oil density (ρ_o), gas specific gravity (γ_g), and gas compressibility factor (Z). Such properties are described in Section 2.1. Due to the changing fluid composition in the PVT cell during a differential liberation, McCain's correction [74] is employed to relate the experimentally measured volumes with those expected for the reservoir fluid composition. Results from differential liberation experiments are used in Chapter 3 as the data set in which traditional BO models are evaluated and in Chapter 4 to tune equations of state.

The separator test provides a basis for converting residual-oil information measured in a differential liberation to a stock-tank-oil basis [133]. In the course of this test, the oil fluid sample is placed into a cell at specific pressure and temperature conditions, typically simulating varying operational settings that might be encountered in real-world production scenarios. The fluid subsequently separates into gas and liquid phases. The gas phase is extracted from the cell, and its volume is measured at standard conditions. The remaining liquid is then subjected to a decrease in pressure and temperature in what is termed as the second separation stage. This process liberates additional dissolved gas. The gas phase is once again removed from the cell and its volume measured at standard conditions. This process is repeated at each separation stage. The oil from the final separation stage at standard conditions is referred to as stock-tank oil. The term 'stock-tank' indicates that under atmospheric conditions, the oil does not liberate additional gas. It is common that

different separator conditions are reproduced in separator tests to identify the specific condition that maximizes stock-tank oil production.

Chapter 3

Uncertainty in fluid model selection

The content of this chapter is a compilation from the following publications:

Lívia Paiva Fulchignoni, Alcino R Almeida, and Roberto da Fonseca Jr. A novel model for the solution gas-oil ratio suitable for CO₂-rich reservoir fluids. *Results in Engineering*, 16(100681), 2022.

Lívia Paiva Fulchignoni, Alcino R Almeida, and Roberto Fonseca Jr. Evaluation of the solution gas-oil ratio correlation proposed by Lasater applied to oils with high CO₂ and high GOR. *Brazilian Journal of Petroleum and Gas*, 15(3), 2021.

This chapter investigates the impact of model selection uncertainty in the context of BO reservoir fluid modeling, with a focus on reservoir fluids that have high CO₂ fraction in their compositions. Such fluids have unique thermophysical properties that require specialized BO models for accurate characterization. However, BO models that have been extensively used to model the reservoir fluid behavior originally did not cover this particular composition, and their extrapolation to this subset of fluids has been little explored in the literature. We first evaluate the performance of nine such traditional BO models for the solution gas-oil ratio (R_s) property of reservoir fluids using an experimental database containing 1457 data points covering a wide

range of CO₂ molar content and gas-oil ratio (GOR) values. Results indicate that existing models yield poor predictions for reservoir fluids with high CO₂ content, with high deviations from experimental data and inconsistencies being observed. A real well+pipe production system is used to illustrate the impact of the R_s model selection on results from multiphase pipe-flow models, and a difference of up to 20°C on the temperature profile is observed. To address this limitation, a novel R_s model with superior overall performance is proposed. The proposed model outperforms the existing models on both training and test sets, even for reservoir fluids with low or no CO₂ content.

3.1 Introduction

Carbon capture utilization and sequestration (CCUS) is recognized as a key technology area for putting global CO₂ emissions on a sustainable path [58]. It consists on the capture of CO₂ from large point sources or the atmosphere for subsequent use or storage. If CO₂ storage is achieved through injection into geological formations, reservoir engineering practices largely follows those well established in petroleum engineering [22], which include models for fluid, flow, geochemical, and geomechanical properties of production/injection systems. Given that captured CO₂ is at present predominantly used for enhanced oil recovery (EOR) [72], the thermophysical behavior of CO₂-rich reservoir fluids has received growing attention.

In particular, the prediction of a reservoir fluid's thermophysical behavior at different pressure and temperature conditions is essential for the estimation of the flow behavior in subsurface reservoirs, wells and pipes. The bubble point, for instance, determines the onset of two-phase flow, a landmark in flow simulations that indicates that single-phase liquid models are no longer valid, and more sophisticated models that would include effects such as phase slippage are therefore required. Accurate models for the fluid properties are thus of extreme importance for multiphase flow modeling of production and injections systems related to subsurface reservoirs, and affect areas such as the fluid lifting, transportation and pumping. Their impact permeates both design and operation phases, including production/injection predictions,

monitoring and optimization, and field layout design.

Even though obtaining such fluid properties experimentally is possible, those analyses are neither cheap nor quick, as they entail the collection and transport of fluid samples from the field to a lab. To overcome those drawbacks, methods to estimate the fluid properties at any pressure and temperature have been proposed. Among them is the Black Oil (BO) approach, which attempts to model the properties of the oil and gas multiphase system through empirical models. The BO models have as independent variables reservoir fluid properties that are easy to obtain, mainly the stock-tank oil gravity (γ_{API}) and the surface gas specific gravity (γ_g). Nonetheless, some authors (e.g., [48, 129]) introduce additional variables for better characterization of the fluid. For example, [129] consider the bubble point and the solution gas-oil ratio at the bubble point of the reservoir fluid as input parameters to their models.

An alternative to the BO approach are the computational packages that output such fluid properties for any pressure and temperature conditions, based on the reservoir fluid's composition and a suitable thermodynamic equation-of-state. There are quite a few of those commercial packages in the industry. Nonetheless, even the computational approach comes with a degree of empiricism in the equation of state regression [46]. Moreover, in some cases the BO approach is preferred due to its low computational cost and ease of implementation.

Those BO models are meant to be applied only to fluids whose properties are covered in the sample set used in the development of the given model. Attempts to extrapolate a model outside its domain of application might result in unpredictable – and often disastrous – deviations from the actual value of a property, thus compromising numerical flow simulations entirely. However, sometimes extrapolation is the only option – especially during the early stages of a new project when fluid properties and reservoir depletion data are not available. This may be critical to the economics (i.e., the rate of return and the projected revenues) of a development project due to poor production forecasts.

For instance, reservoir fluids from Brazil's offshore pre-salt region in general present high gas-oil ratio (GOR) and high CO₂ content in their composition. These will further increase along production time due to the use of water alternating gas

(WAG), an enhanced oil recovery (EOR) method. Regrettably, little is known about the prediction of properties associated with high-CO₂ reservoir fluids, and the literature on this matter is sketchy at best. Under these circumstances, an investigation of the applicability of conventional BO models to fluids with this peculiar characteristic becomes necessary. In particular, this paper focuses on the solution gas-oil ratio (R_s) property, defined as the standard volume of gas dissolved in a unity standard volume of oil at a given pressure and temperature.

We evaluate the performance of nine BO R_s models on oils with high gas-oil ratio and high CO₂ content. The models were selected based on their relevance and/or on the fitting technique and diversity of the data points used for their development. They are, in alphabetic order: Al-Marhoun's [7], Al-Shammasi's [9], Alakbari et al.'s [11], Glasø's [48], Lasater's [68], Petrosky and Farshad's [98], Standing's [118], Vazquez and Beggs's [128], and Velarde et al.'s [129] models. A dataset (training) containing 1457 R_s experimental points is considered for the evaluation. Among them, 829 experimental points are from reservoir fluids with CO₂ molar content ranging from 11% to 38% (15% to 45% in the surface gas) and GOR from 1117 scf/STB to 2487 scf/STB, produced 200 km offshore Brazil, in the Santos Basin's pre-salt layer. After high prediction errors are found for this subset, a more general R_s model is proposed. This novel model is validated against a test (holdout) dataset containing 173 data points, with CO₂ molar content in the surface gas ranging from 0% to 44% and GOR from 19 to 2409 scf/STB. The mean average percentage error is 6% in the training set and 8% in the test set, whereas the root mean square error is 27 scf/STB and 71 scf/STB, respectively.

3.2 Literature Review

Standing [118] was the first to propose a specific model for the bubble point pressure (p_b) of oils as a function of stock-tank oil gravity, gas specific gravity, gas-oil ratio and temperature. Standing did not make clear, however, whether his model is suitable for pressures below the bubble point – although this kind of extrapolation is commonly made. This model was developed with the aid of an experimental data base containing

105 bubble points from oils produced in California, USA. His expression may be easily rearranged so as to provide an equation for R_s instead of p_b .

Several other BO models for R_s (or p_b) have been proposed since then. Lasater [68] developed a model for the bubble point by fitting 158 experimental data points in an equation based on Henry's Law. No analytical expression was supplied; the author rather relied on two graphical representations: one for the effective molecular weight of the tank oil as a function of its gravity and another relating the gas mole fraction in the reservoir oil to a pressure factor of the bubble point. Later, Whitson and Brûlé [133] proposed an equation describing Lasater's chart for the bubble point's pressure factor and recommended Cragoe's [31] model for estimating the effective molecular weight of the tank oil.

Vazquez and Beggs [128] also proposed a model for R_s . They used over 5000 experimental data points from fields around the world. By increasing the number of data points used to tune the equation, the authors claimed that their model results should be applicable to a wider range of oil properties.

Glasø [48] included the effect of paraffinic compounds and contaminants (namely CO_2 , N_2 and H_2S) in oil. He observed that "the effect of nonhydrocarbons on saturation pressure can range from minimal to extreme, depending on the type of nonhydrocarbon, the quantity with which it is found in the reservoir oil, temperature, and stock-tank oil gravity". This observation corroborates the motivation of the present work, given the unusually high contents of a nonhydrocarbon (CO_2) in the fluids produced from the pre-salt layers in the Santos Basin. Glasø's models are based on regression of data of reservoir fluids from the North Sea.

Al-Marhoun [7] presented a model for the bubble point pressure based on fitting 160 experimental data points from 69 Middle East reservoirs. Petrosky and Farshad [98] developed a model for the bubble point pressure by introducing three additional fitting parameters to Standing's [118] model. The equation coefficients were fitted against data from the Gulf of Mexico's crude oils.

Verlarde et al. [129] proposed a model for R_s based on 2097 experimental values of 195 PVT analysis, after verifying that published models did not match the concave up – point of inflection – concave down shape of the R_s curve as pressure declined

below the bubble point for a reservoir oil with high initial gas-in-solution. They also proposed a model for the bubble point pressure that is independent from the R_s one.

Al-Shammasi [9] published a model for the bubble point pressure based on 1243 data points (from published literature and from unpublished data from Kuwaiti reservoirs). The author also evaluated thirteen published models against the database and proposed new coefficients for each model, based on the overall data set.

Regarding the effect of nonhydrocarbons on saturation pressure, Ostermann et al. [84] experimentally compared predictions given by several models with the measured bubble point pressure of crudes containing high N_2 and CO_2 concentrations, from the Alaskan North Slope. The study emphasized that the presence of nonhydrocarbons in oils has a significant effect on the prediction of bubble point pressures, so that correction factors must be applied.

More recently, novel models for the bubble point pressure have been proposed using machine learning techniques (e.g., [102, 10, 1, 56, 47, 11]). In particular, Heidarian et al. [56] developed a p_b model exclusively for Middle East crudes using a genetic algorithm and compared its results with predictions from 16 existing BO models. Ghorbani et al. [47] compare the performance of four machine learning techniques for predicting the bubble point pressure of oils from the Ahvaz field, in Iran. They also compare the four models' results with predictions from 15 existing models. Alakbari et al. [11] proposed a bubble point pressure model using a polynomial neural network trained on a worldwide dataset, whose predictions were compared against 21 models available in the literature. They also performed a sensitivity analysis on such models and verified that some ([83, 54, 49, 37]) predicted inconsistent behaviors and that extrapolations can lead to unphysical estimates.

Despite the recent efforts for improving the reservoir fluid behavior prediction through empirical equations, the newer models have not seen widespread implementation - they are not included in the most popular commercial steady-state multiphase flow simulators. PIPESIM version 2017, for instance, implements the following correlations for the solution gas-oil ratio: Lasater [68], Standing [118], Vazquez and Beggs [128], de Ghetto et al. [34], Glasø [48], and Petrosky and Farshad [98].

3.3 Methodology

3.3.1 Data description

Our experimental database consists of 1457 R_s points, corresponding to 171 PVT analyses of reservoir fluids from 112 Brazilian wells. This dataset is proprietary to Petrobras and is therefore subject to confidentiality constraints.

The R_s data points were obtained after applying McCain's [74] correction to the volumes measured in differential liberation experiments at each pressure reduction step. The reservoir fluids with CO₂ molar fraction above 8% are all from the Santos Basin pre-salt region. Table 3.1 describes the samples in terms of the stock-tank API gravity (γ_{API}), the surface gas specific gravity (γ_g), the stock-tank gas-oil ratio (GOR) and the bubble point pressure (p_b).

Table 3.1 also shows the data set division into six groups according to the molar fraction of CO₂ in the stock-tank gas (y_{CO_2}). Group I is composed of 433 R_s points from reservoir fluids with up to 2% CO₂ in the stock-tank gas. Groups II to VI contain 1024 samples with CO₂% in the stock-tank gas ranging from 2% to 45%. The stock-tank gas H₂S and N₂ molar fractions were less than 1% throughout the data set, except for one sample with 4% of N₂.

Table 3.1: Summary of the data set used to evaluate solution gas-oil ratio models

Group	y_{CO_2} (mol %)	Number of PVT analysis	Number of R_s points	γ_{API}	γ_g	GOR (scf/STB)
I	≤ 2	62	433	[12.6, 49.4]	[0.57, 1.03]	[75, 1620]
II	2 to 5	25	195	[14.4, 38.7]	[0.60, 0.96]	[214, 1927]
III	5 to 15	19	182	[16.6, 28.3]	[0.63, 0.92]	[225, 1447]
IV	15 to 20	14	137	[24.1, 29.5]	[0.91, 0.94]	[1117, 1553]
V	20 to 30	41	391	[25.5, 30.3]	[0.94, 1.04]	[1070, 2109]
VI	43 to 45	10	119	[26.5, 27.9]	[1.10, 1.15]	[2255, 2487]
TOTAL		171	1457			

3.3.2 Numerical implementation

The nine R_s models [7, 9, 11, 48, 68, 98, 118, 128, 129] implemented in this Chapter are reported in Appendix A.

3.3.3 Error metrics

For the sake of completion, prediction errors of each R_s model were computed according to five different metrics, described below. In Eqs. 3.1 to 3.5, y_{exp} is the measured value, \hat{y} is the predicted value, and N is the sample size.

- Mean Percentage Error (MPE) – indicates the trend of the relative error

$$\text{MPE} = \frac{1}{N} \sum_{i=1}^N \frac{y_{\text{exp}} - \hat{y}_i}{y_{\text{exp}}} \quad (3.1)$$

- Standard deviation of the percentage error (σ_{PE}) – a measure of the dispersion of the relative error

$$\sigma_{\text{PE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\frac{y_{\text{exp}} - \hat{y}_i}{y_{\text{exp}}} - \text{MPE} \right)^2} \quad (3.2)$$

- Mean Absolute Percentage Error (MAPE) – indicates the relative error's amplitude

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^N \left| \frac{y_{\text{exp}} - \hat{y}_i}{y_{\text{exp}}} \right| \quad (3.3)$$

- Standard deviation on the absolute percentage error (σ_{APE}) – gives the dispersion on the absolute percent error.

$$\sigma_{\text{APE}} = \sqrt{\frac{1}{N} \sum_{i=1}^N \left(\left| \frac{y_{\text{exp}} - \hat{y}_i}{y_{\text{exp}}} \right| - \text{MAPE} \right)^2} \quad (3.4)$$

- Root Mean Square Error (RMSE) – gives the amplitude of the absolute error, but is more sensitive to larger errors than the MAPE

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_{\text{exp}} - \hat{y}_i)^2} \quad (3.5)$$

The RMSE was used as the comparison criterion between the evaluated models, due to its interpretability. The interested reader may find use in the other metrics.

3.4 Results

Because the bubble point pressure is not typically available in the BO modeling, Velarde et al.'s model [129] is considered with both the experimental and the calculated p_b . The first is denoted "Velarde et al. [129] - p_b exp." and the second "Velarde et al. [129] - p_b est.", in which p_b is estimated according to Eq. A.29.

Figure 3.1 presents a comparison between lab-measured and predicted R_s values for a well from each group. For the sake of clarity, only the five models with the lowest RMSE are shown. For the representative wells of Groups I to V (Figures 3.1a to 3.1d), there are R_s models able to model the experimental behavior. On the other hand, for the representative well of Group VI (Figure 3.1f), all models highly deviate from the experimental data.

Group VI deserves further attention for its peculiar PVT behavior. The R_s curve shown in Figure 3.1a differs from the other ones by its sharp decline as the pressure reduces near the bubble point region. This change in shape of the R_s curve according to the fluid's GOR is also reported in [129]. It is worth mentioning that, for Group VI, McCain's procedure [74] is even more relevant to correct the dissolved gas volumes measured by the differential liberation technique, due to the greater divergence between the gas-oil ratio at the bubble point obtained by the flash and differential liberations.

Table 3.2 summarizes the results obtained for all Group I wells and highlights the best and worst RMSE in green and red, respectively. Velarde et al.'s model [129] presented the best RMSE, outperforming all the others, whereas Alakbari et al. [11] yielded the poorest performance. As one could expect, Velarde et al. [129] performs better when the experimental bubble point pressure is used as input, instead of its estimate. Nonetheless, Velarde et al. [129] with estimated p_b presents the second best RMSE for Group I. The same does not hold for oils with higher CO₂ content.

Similarly, Table 3.3 reports the results for Group II wells. Again, Velarde et al.'s

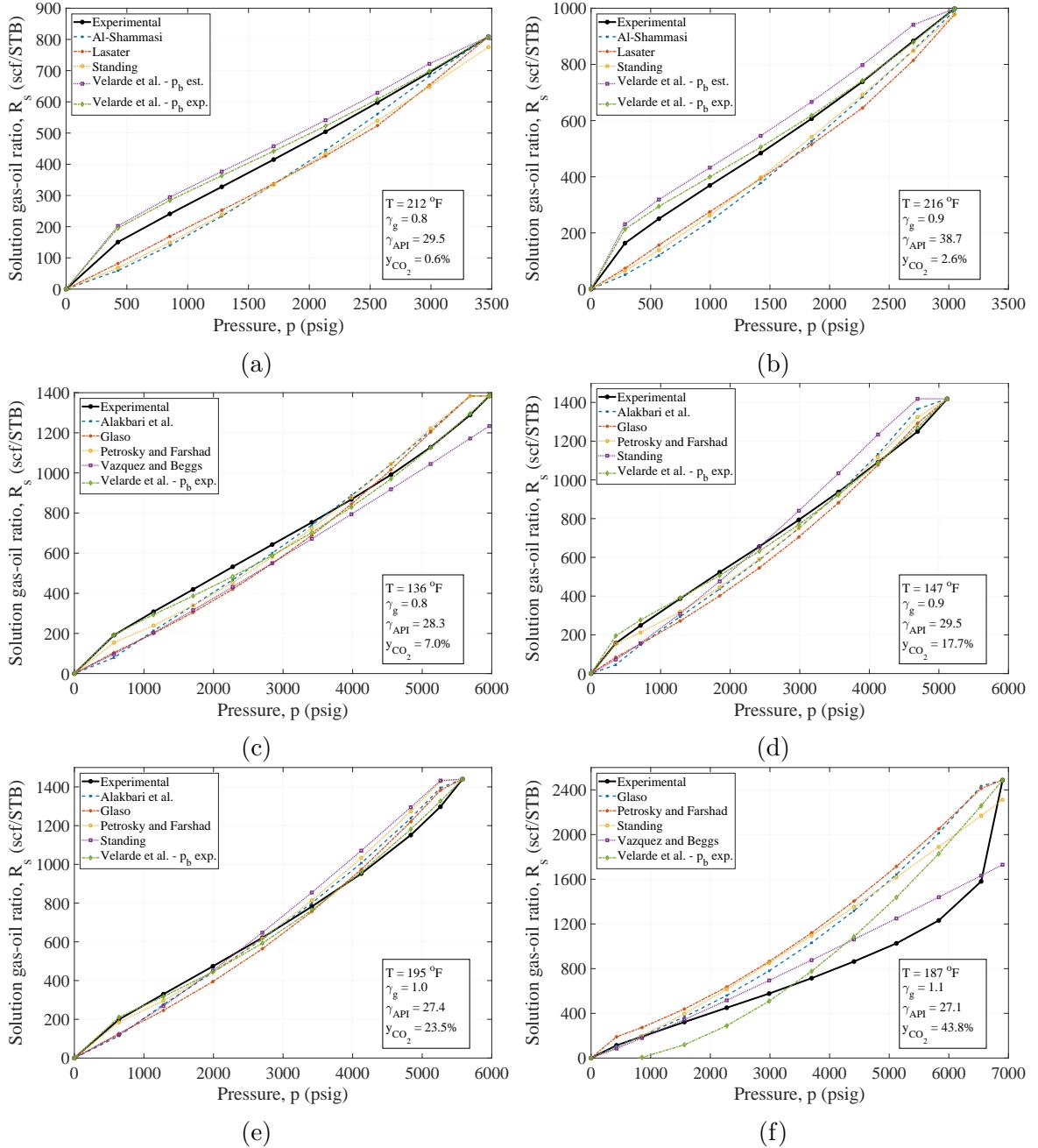


Figure 3.1: Experimental and predicted solution gas-oil ratio from a differential liberation for reservoir fluids representative of (a) Group I, (b) Group II, (c) Group III, (d) Group IV, (e) Group V, and (f) Group VI. Pressures are plotted below and at the bubble point.

Table 3.2: Summary of the R_s prediction errors for Group I

R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	30.17	23.19	30.97	22.11	68.71
Al-Shammasi [9]	10.07	23.77	19.26	17.19	58.73
Alakbari et al. [11]	54.69	93.51	55.63	92.95	87.11
Glasø [48]	23.13	16.25	24.65	13.84	70.10
Lasater [68]	3.25	18.05	13.43	12.49	45.87
Petrosky and Farshad [98]	0.28	23.44	14.72	18.25	52.31
Standing [118]	9.02	19.45	15.69	14.61	50.12
Vazquez and Beggs [128]	24.11	16.29	24.96	14.94	85.65
Velarde et al. [129] - p_b est.	-8.68	14.43	13.02	10.68	36.80
Velarde et al. [129] - p_b exp.	-3.57	11.94	8.41	9.20	20.87

model [129] presented the best performance: the lowest RMSE is obtained when the measured p_b is used and the second lowest RMSE when p_b is estimated according to the authors' proposed equation. For this group, Vazquez and Beggs' [128] model presented the highest RMSE. For Group III (Table 3.4), Velarde et al. [129] gives the best RMSE when the experimental p_b is used as an input, and the worst RMSE when p_b is estimated. The same is observed for Groups VI and V, as shown in Tables 3.5 and 3.6, respectively.

Table 3.3: Summary of the R_s prediction errors for Group II

R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	23.29	23.91	25.06	22.05	75.88
Al-Shammasi [9]	13.46	22.95	19.95	17.59	75.31
Alakbari et al. [11]	41.73	57.08	42.12	56.80	108.43
Glasø [48]	28.11	12.54	28.69	11.16	110.31
Lasater [68]	4.74	20.55	16.63	12.97	67.88
Petrosky and Farshad [98]	7.30	21.57	15.72	16.47	104.12
Standing [118]	12.29	20.37	18.11	15.42	87.65
Vazquez and Beggs [128]	27.26	15.56	27.26	15.56	143.36
Velarde et al. [129] - p_b est.	-13.03	15.00	15.94	11.86	58.88
Velarde et al. [129] - p_b exp.	-2.61	11.48	7.05	9.43	35.88

Table 3.4: Summary of the R_s prediction errors for Group III

R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	4.10	28.60	22.47	18.16	120.59
Al-Shammasi [9]	1.73	24.00	17.87	16.11	79.52
Alakbari et al. [11]	21.33	44.61	25.37	42.44	57.39
Glasø [48]	17.37	17.41	20.27	13.93	72.42
Lasater [68]	-7.75	23.51	18.60	16.34	111.46
Petrosky and Farshad [98]	-0.82	21.10	11.76	17.53	44.74
Standing [118]	-0.85	21.30	15.52	14.62	72.96
Vazquez and Beggs [128]	16.31	17.19	19.31	13.73	84.21
Velarde et al. [129] - pb est.	-32.18	21.84	32.18	21.84	151.52
Velarde et al. [129] - pb exp.	-0.59	7.08	4.41	5.57	18.53

Table 3.5: Summary of the R_s prediction errors for Group IV

R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	-9.06	28.14	24.36	16.75	197.98
Al-Shammasi [9]	7.10	22.04	16.75	15.99	98.33
Alakbari et al. [11]	9.84	17.21	13.10	14.88	66.94
Glasø [48]	14.84	15.82	16.15	14.48	83.39
Lasater [68]	-2.58	22.56	18.55	13.09	160.44
Petrosky and Farshad [98]	4.39	9.94	8.56	6.70	57.32
Standing [118]	3.40	18.90	14.29	12.83	93.97
Vazquez and Beggs [128]	20.99	12.62	20.99	12.62	118.64
Velarde et al. [129] - pb est.	-40.35	25.96	40.35	25.96	235.29
Velarde et al. [129] - pb exp.	-1.43	9.07	4.66	7.91	25.71

Table 3.6: Summary of the R_s prediction errors for Group V

R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	-17.34	30.37	28.68	20.01	251.91
Al-Shammasi [9]	4.04	21.86	16.54	14.86	107.65
Alakbari et al. [11]	6.07	17.04	11.83	13.68	76.33
Glasø [48]	10.65	15.75	14.06	12.79	76.35
Lasater [68]	-7.46	24.14	19.79	15.71	201.11
Petrosky and Farshad [98]	-0.37	8.55	6.86	5.12	66.63
Standing [118]	2.91	18.24	13.48	12.62	91.06
Vazquez and Beggs [128]	19.86	12.06	19.86	12.06	125.64
Velarde et al. [129] - pb est.	-48.20	30.38	48.20	30.38	281.68
Velarde et al. [129] - pb exp.	-1.41	6.99	4.25	5.73	28.00

Table 3.7: Summary of the R_s prediction errors for Group VI

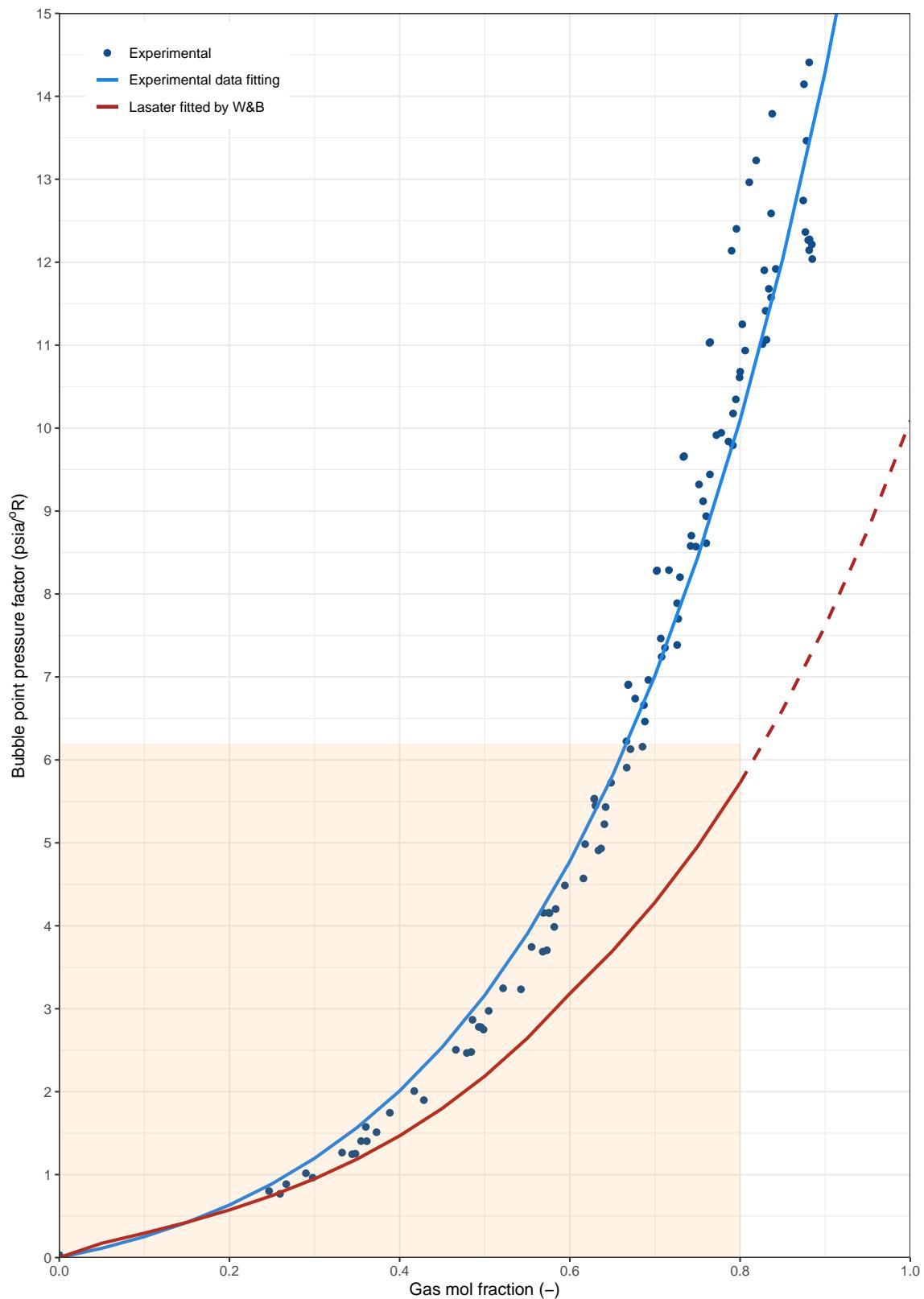
R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	-106.65	68.08	106.83	67.79	933.45
Al-Shammasi [9]	-46.33	35.75	49.87	30.62	540.53
Alakbari et al. [11]	-47.04	32.87	48.01	31.44	552.32
Glasø [48]	-44.26	33.13	44.55	32.74	527.19
Lasater [68]	187.33	935.25	322.96	897.49	11490.53
Petrosky and Farshad [98]	-55.15	24.74	55.15	24.74	525.23
Standing [118]	-43.74	31.10	45.44	28.56	484.97
Vazquez and Beggs [128]	-15.14	20.55	21.85	13.20	262.00
Velarde et al. [129] - Pb est.	-133.44	67.19	133.44	67.19	900.89
Velarde et al. [129] - Pb exp.	14.37	51.55	40.13	35.41	274.89

Finally, Table 3.7 shows the results for reservoir fluids in Group VI, whose CO₂ molar content in the surface gas ranges from 43 to 45%. All models incur high prediction errors. In fact, Lasater [68] and Velarde et al. [129] can predict negative values for R_s, which is physically unrealistic. An example of Velarde et al. [129] providing a negative R_s value is illustrated in Figure 3.1f. As can be expected, an unsuitable R_s model in a flow simulation can yield disastrous results [43].

In particular, for Lasater's correlation applied to oils in Group VI, the high errors are justified when its validity range is analyzed [42]. In terms of the primitive parameters (P , T , R_s , γ_{API} and γ_g), most of the experimental data is inside the ranges covered by Lasater. However, the domain of the parameter p_f goes up to the approximate value of 6.4 psia/°R in Lasater's data set whereas p_f values in our sample reaches up to 14.4 psia/°R. In fact, checking individual properties ranges in a correlation does not guarantee that a particular combination was covered in the original work.

For a graphical comparison between the experimental data and the Lasater's correlation, Figure 3.2 presents experimental data on the p_f vs y_g chart. For the experimental p_f the corresponding y_g was calculated from Eq. (A.15) using Eq. (A.18) for M_o . Figure 3.2 also shows Whitson and Brulé's fitting of Lasater's original p_f chart (Eq. A.17). The domain region of Lasater's original chart is highlighted in orange. It is clear that Lasater's chart is being used outside its domain for the specific kind of fluid considered within this work. Indeed, deviations become progressively larger as we move outside the correlation's domain. Lasater's correlation was based on systems essentially free of non-hydrocarbon material. The author suggests an error of 5% in the predicted bubble point pressure when the gas has 9.1% of CO₂ content¹. However, for the CO₂ content of our sample the resulting errors are unacceptable.

¹It is not specified whether this fraction is in molar or mass basis.

Figure 3.2: Experimental R_s data and new p_f curve fitting

Moreover, checking extrapolation of a correlation is important because the results obtained in a simulator may be the result of internal safeguards instead of the correlation itself. For instance, PIPESIM® limits the original range of p_f covered by Lasater. In order to avoid extrapolation, the imposition of a limit to the maximum value of y_g obtained from the fitting equations is a common practice in oil and gas multiphase flow simulators. In general, this limit value for y_g is 0.85, which corresponds to the highest y_g value given in a table presented by Lasater for “smoothed points” read from the p_f *versus* y_g chart.

In addition to assessing the fidelity of an BO equation when applied to a real fluid, it is important to verify the influence of the model selection on the estimation of relevant variables in real systems. As an example, we consider a well+pipeline production system from Brazil’s presalt region to evaluate the impact of the selected R_s correlation on pressure, temperature and void fraction profiles is constituted by a vertical well with 3.6 km depth and 5.8” ID tubing and a flexible pipeline (flowline and riser) of 5.6 km long and 8” ID. The reservoir fluid is characterized by an oil API gravity of 27.2° and a reservoir fluid gas-oil ratio of 440 Sm³/Sm³. The flashed gas has a CO₂ molar fraction of 44% and a specific gravity of 1.12. No water production is present. The boundary conditions are the pressures at the inlet (reservoir), the reservoir’s Productivity Index (PI), and the outlet (surface). Flow simulations were performed on the commercial simulator PIPESIM version 2015. Simulated pressure, temperature and void fraction profiles simulated with Standing’s [118] and Lasater’s [68] R_s models are shown in Figure 3.3, 3.4 and 3.5, respectively. The blue and the green curve distinguish from each other only for the model selected for the solution gas-oil ratio (R_s). The void fraction inside the well differs significantly, with Lasater’s model indicating single phase flow during approximately the first two-thirds of the well. However, there is no significant difference in the pressure profile since at those high pressures and fluid composition the gas and liquid densities are very similar inside the well. In its turn, the calculated temperature can differ up to 20°C near the wellhead due to the expansion of the gas (Standing’s case) against the friction heating in single phase flow (Lasater’s case).

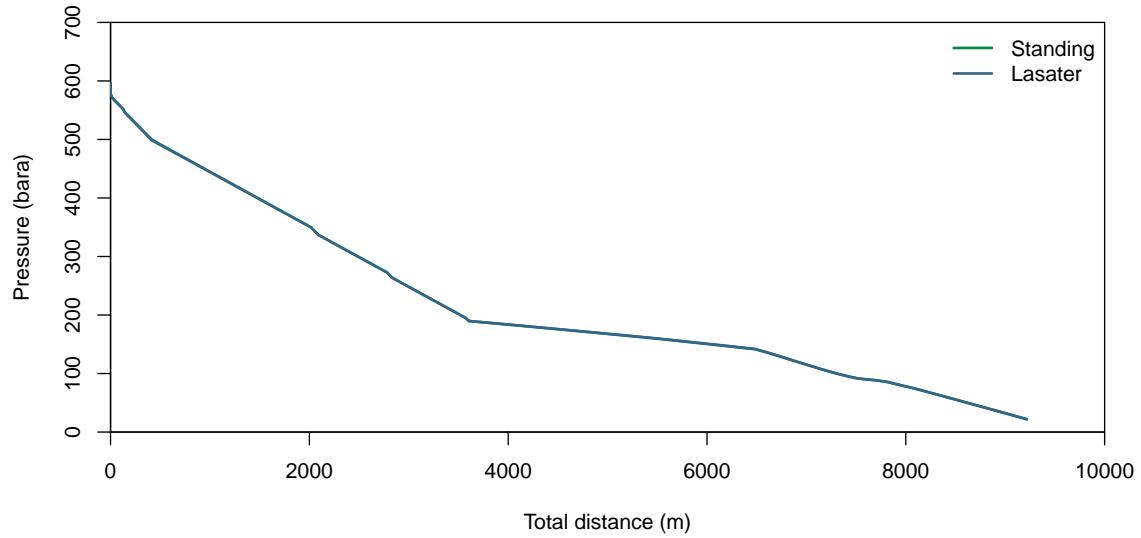


Figure 3.3: Pressure profile for a well+pipeline production system considering Standing's [118] and Lasater's [68] R_s models.

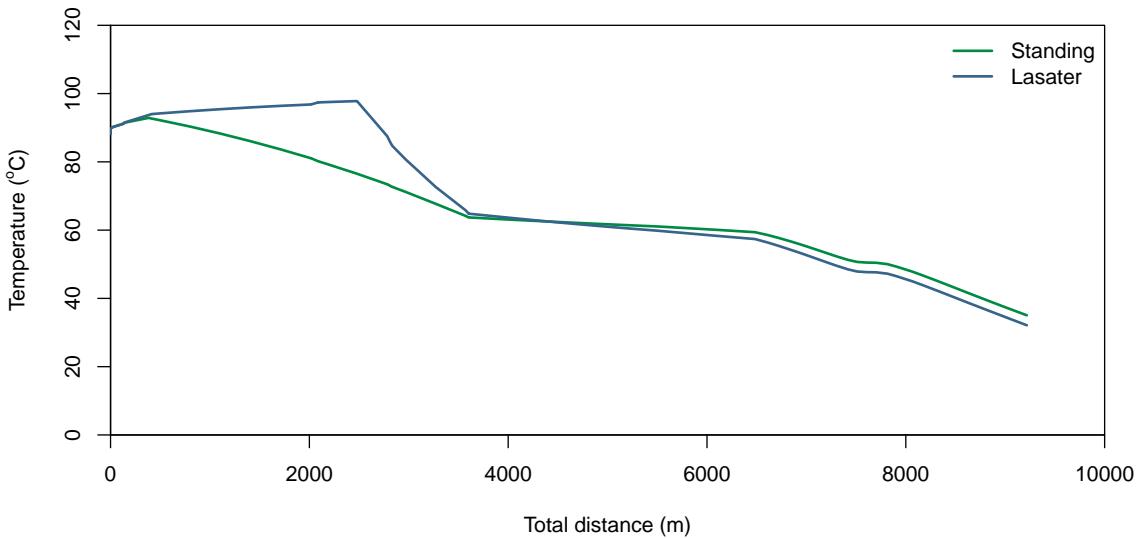


Figure 3.4: Temperature profile for a well+pipeline production system considering Standing's [118] and Lasater's [68] R_s models.

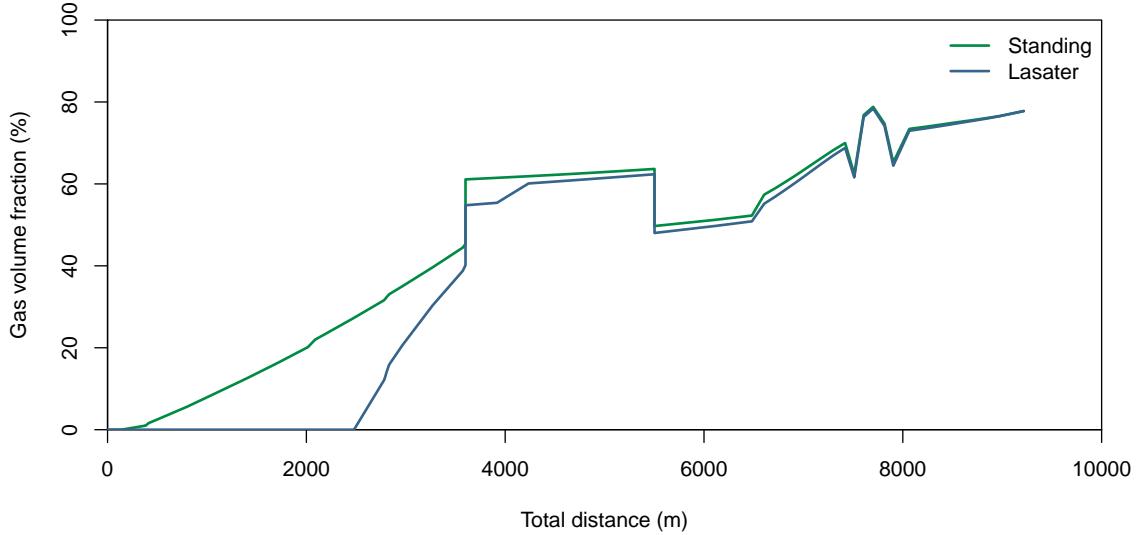


Figure 3.5: Gas volume fraction profile for a well+pipeline production system considering Standing’s [118] and Lasater’s [68] R_s models.

Figure 3.6 summarizes the previous results by showing the RMSE per model per group. There is an overall trend of increasing RMSE with increasing CO₂ content in the surface gas. The RMSE is particularly high for fluids in Group VI. Clearly, the R_s models implemented in this study are unable to represent the behavior of reservoir fluids from Group VI, i.e. reservoir fluids with high CO₂ content and high GOR. This observation motivates the proposition of a new R_s model that embraces this subset of reservoir fluids.

3.5 Proposed model for R_s

Plenty of models have been proposed in the literature for the solution gas-oil ratio (R_s). Those models are based on specific reservoir fluid samples, with particular properties and from different production regions around the world. This study demonstrated that CO₂-rich reservoir fluids lie outside the domain of application of such models. As shown in Section 3.4, extrapolating the limits of the aforementioned

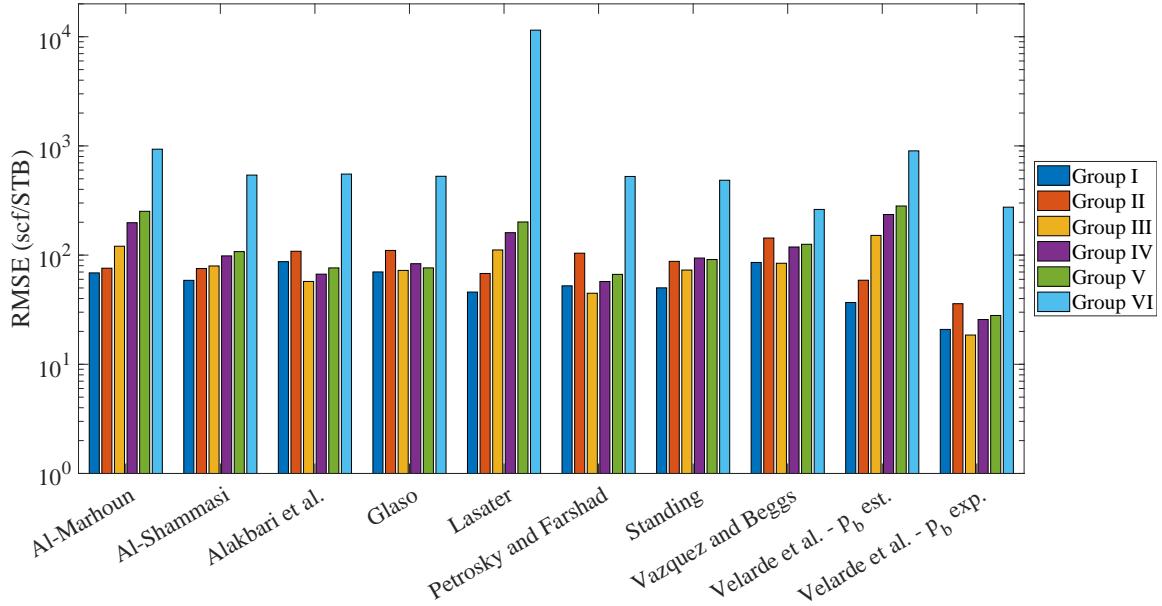


Figure 3.6: Root Mean Square Error (RMSE) of R_s predictions given by nine models, grouped by the CO₂ content in the surface gas of the reservoir fluid sample.

models results in unacceptable errors for this subset of reservoir fluids. Therefore, a model that embraces Brazilian pre-salt oils with high CO₂ content and high GOR needs to be developed.

Since Velarde et al.'s model [129] presented consistently good performance through Groups I to V given that p_b is known, it is used as the basis of our new model, being modified to account for the presence of CO₂. In fact, McCain et al. [73] also recommend this equation, after evaluating ten BO R_s models on 5359 data points from 718 reservoir fluids. In addition, inspired by Glasø's [48] and Jacobson's [60] corrections for the presence of contaminants on the pressure term, a correction factor f (Eq. 3.7g) that adjusts p_b in Eqs. (3.7d), (3.7e), and (3.7f) is introduced. This correction factor is a function of the CO₂ fraction in the surface gas and of the pressure in which R_s is being computed. Furthermore, the gas specific gravity was substituted by the hydrocarbon gas specific gravity, calculated from

$$\gamma_{g_{HC}} = \frac{\gamma_g - (y_{N_2}M_{N_2} + y_{CO_2}M_{CO_2} + y_{H_2S}M_{H_2S})/M_{air}}{1 - y_{N_2} - y_{CO_2} - y_{H_2S}} \quad (3.6)$$

where y_{N_2} , y_{CO_2} , and y_{H_2S} are the molar fractions of N₂, CO₂, and H₂S in the surface gas, respectively; and M_{N_2} , M_{CO_2} , M_{H_2S} , and M_{air} are the N₂, CO₂, H₂S, and air molecular weights at standard conditions, respectively.

The coefficients of Velarde et al.'s equation [129] were fitted to match our experimental data according to the gradient descent method, using their original values as the initial guess. The proposed model for the solution gas-oil ratio is given in Eqs. (3.7a) to (3.7g).

$$R_{sr} = a_1 p_r^{a_2} + (1 - a_1) p_r^{a_3} \quad (3.7a)$$

$$p_r = \frac{p}{p_b} \quad (3.7b)$$

$$R_{sr} = \frac{R_s}{GOR} \quad (3.7c)$$

$$a_1 = A_0 \gamma_{g_{HC}}^{A_1} \gamma_{API}^{A_2} T^{A_3} (p_b f_{p_b})^{A_4} \quad (3.7d)$$

$$a_2 = B_0 \gamma_{g_{HC}}^{B_1} \gamma_{API}^{B_2} T^{B_3} (p_b f_{p_b})^{B_4} \quad (3.7e)$$

$$a_3 = C_0 \gamma_{g_{HC}}^{C_1} \gamma_{API}^{C_2} T^{C_3} (p_b f_{p_b})^{C_4} \quad (3.7f)$$

$$f_{p_b} = 1 + D_0 y_{CO_2}^{D_1} p^{D_2} \quad (3.7g)$$

where

$A_0 = 2.826\,773 \times 10^{-6}$	$B_0 = 8.902\,647 \times 10^{-5}$	$C_0 = 0.014815$	$D_0 = 0.028061$
$A_1 = 0.099827$	$B_1 = 0.515955$	$C_1 = -0.431815$	$D_1 = 5.237181$
$A_2 = 2.431229$	$B_2 = -0.346606$	$C_2 = -0.185931$	$D_2 = 1.043339$
$A_3 = -0.030813$	$B_3 = 0.421825$	$C_3 = 0.232081$	
$A_4 = 0.359656$	$B_4 = 1.153339$	$C_4 = 0.390657$	

and p is the pressure in psig, T is the temperature in °F, γ_{API} is the API gravity of the stock-tank oil in standard conditions, γ_g is the surface gas specific gravity in standard

conditions, y_{CO_2} is the molar fraction of CO_2 in the surface gas, p_b is the bubble point pressure in psig, GOR is the stock-tank gas-oil ratio in scf/STB, and R_s the solution gas-oil ratio in scf/STB. The values of p_b must refer to the same temperature T in which R_s is being computed. Section 3.5.1 provides a numerical calculation example for the prediction of a R_s point for the reservoir fluid represented in Figure 3.1f.

A drawback of the proposed model is the need for the bubble point pressure at the desired temperature as an input, which may not be available. In that case, if the bubble point pressure is unknown, it should be estimated according to an empirical model. For reservoir fluids with $y_{CO_2} > 15\%$, we recommend Glasø's p_b model [48], which was also the model of choice of Ostermann et al. [84] for reservoir fluids with high N_2 and CO_2 content. However, we disregard Glasø's [48] corrections for the presence of paraffin and CO_2 . Rather, a new correction factor f_{CO_2} (Eq. 3.8c) is proposed for the CO_2 presence as a function of the CO_2 molar content in the surface gas and the reservoir fluid GOR. Moreover, we also consider the hydrocarbon gas specific gravity, obtained according to Eq. (3.6), as an input to the p_b model. Glasø's [48] model is reproduced in Eqs. (3.8a) to (3.8e) with the aforementioned modifications for completeness, where P_b is the bubble point pressure in psia. The uppercase P distinguishes the absolute pressure in psia from p , the relative pressure in psig.

$$P_b = f_{CO_2} f_{N_2} f_{H_2S} 10^{1.7669 + 1.7447 \log P_b^* - 0.30218(\log P_b^*)^2} \quad (3.8a)$$

$$P_b^* = \left(\frac{GOR}{\gamma_{gHC}} \right)^{0.816} \frac{T^{0.172}}{\gamma_{API}^{0.989}} \quad (3.8b)$$

$$f_{CO_2} = 1.0 - 0.1297 y_{CO_2}^{0.5320} GOR^{-0.3332} T^{0.6235} \quad (3.8c)$$

$$f_{N_2} = 1.0 + [(-2.65 \times 10^{-4} \gamma_{API} + 5.5 \times 10^{-3})T + 0.0931 \gamma_{API} - 0.8295] y_{N_2} + (1.954 \times 10^{-11} \gamma_{API}^{4.699} T + 0.027 \gamma_{API} - 2.366) y_{N_2}^2 \quad (3.8d)$$

$$f_{H_2S} = 1.0 - (0.9035 + 0.0015 \gamma_{API}) y_{H_2S} + 0.019(45 - \gamma_{API}) y_{H_2S}^2 \quad (3.8e)$$

Besides p_b , the inputs to the proposed R_s model are: surface gas specific gravity, stock-tank oil API gravity, reservoir fluid gas-oil ratio, and pressure and temperature

in which the solution gas-oil ratio is desired. It is worth noting that GOR is not a common input in R_s models available in the literature. Nonetheless, in contrast to the difficulty of measuring p_b , we understand that GOR information can be easily obtained. In case a reservoir fluid sample is available, it can be directly measured through a flash liberation experiment. Otherwise, it can be estimated from the gas and oil production volumetric flowrates.

3.5.1 Numerical calculation example

This section provides a calculation example for the R_s prediction by the proposed model, which can be used for validating the numerical implementation of equations (3.6) and (3.7a)-(3.7g). The inputs specified below refer to the reservoir fluid represented in Figures 3.1f and 3.8f.

$$p = 5831.4 \text{ psig}$$

$$T = 186.8 \text{ }^{\circ}\text{F}$$

$$p_b = 6901.0 \text{ psig}$$

$$\gamma_{\text{API}} = 27.06^\circ$$

$$\gamma_g = 1.1252$$

$$\text{GOR} = 2487.4 \text{ scf/STB}$$

$$y_{\text{CO}_2} = 0.4383$$

$$y_{\text{N}_2} = 0.0024$$

$$y_{\text{H}_2\text{S}} = 0$$

Proposed model prediction:

$$\gamma_{\text{gHC}} = 0.8165$$

$$f_{p_b} = 4.1693$$

$$a_1 = 0.2874$$

$$a_2 = 32.2458$$

$$a_3 = 1.6273$$

$$R_s = 1350.7 \text{ scf/STB}$$

3.5.2 Error analysis

Figure 3.7 compares, for each CO₂ content group, the RMSE of the proposed p_b model with those given by Al-Marhoun [7], Al-Shammasi [9], Alakbari et al. [11], Glasø [48], Standing [118], Vazquez and Beggs [128], and Velarde et al. [129]. Glasø's [48] p_b modified model (Eqs. 3.8a-3.8e) predictions have a significantly lower error than predictions obtained by other popular p_b models for reservoir fluids with $y_{CO_2} > 15\%$ (Groups IV, V, and VI) in particular. As expected, as the CO₂ content in the surface gas decreases, the proposed modified Glasø model converges to the original model ([48]). For reservoir fluids with $y_{CO_2} < 15\%$, both original and modified Glasø models provide a competitive p_b estimate, but other models offer a more suitable alternative for calculating p_b .

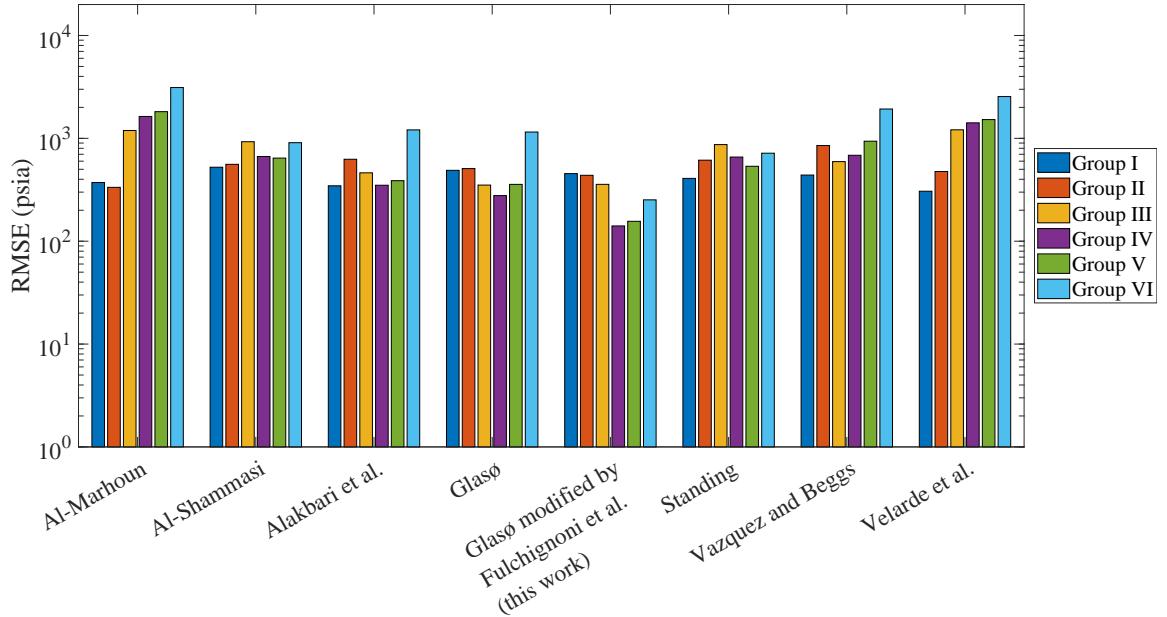


Figure 3.7: Root Mean Square Error (RMSE) of P_b predictions given by eight models, grouped by the CO₂ content in the surface gas of the reservoir fluid sample.

A seventh group, denoted “test group”, is introduced here for its evaluation against a new set of data. It contains 173 R_s points from 22 differential liberation experiments and 18 novel reservoir fluid compositions. The fluids in this group have CO₂ content ranging from 0% to 38% (0% to 44% in the surface gas), while their GOR ranges

from 19 scf/STB to 2409 scf/STB, the stock tank API gravity from 11.8° to 32.8°, and the surface gas specific gravity from 0.57 to 1.12. Table 3.8 shows the prediction errors for the test set obtained by the aforementioned nine R_s models and by the R_s proposed in this paper (Eqs. 3.7a to 3.7g), distinguishing two scenarios: when p_b is known (“Proposed model - p_b exp.”) and estimated (“Proposed model - p_b est.”). In addition, Table 3.9 presents the prediction errors of the proposed R_s model for all the groups. As expected, the novel model provides smaller errors if the bubble point pressure is known, compared to estimated p_b . In this case, the model is able to provide low prediction errors - the MAPE is below 10% for all the groups - and to match the experimental curve with excellent agreement.

Overall, the proposed R_s model outperforms the traditional models available in the literature. If p_b is known, the proposed model provides a RMSE of 17.76 scf/STB for Group I, while the best RMSE among the exiting models is 20.87 scf/STB (Velarde et al. [129]). For Groups II to VI, its RMSE is 37.04 scf/STB, 14.08 scf/STB, 18.81 scf/STB, 28.69 scf/STB, and 46.04 scf/STB; while the best performing existing model provides a RMSE of 35.88 scf/STB (Velarde et al. [129]), 18.53 scf/STB (Velarde et al. [129]), 25.71 scf/STB (Velarde et al. [129]), 28.00 scf/STB (Velarde et al. [129]), and 262.00 scf/STB (Vazquez and Beggs [128]), respectively.

On the other hand, when p_b is unknown, the error propagates from the p_b estimation to the R_s estimation, increasing the mismatch between the experimental and the predicted R_s values. Consequently, using a better model to estimate p_b will increase the accuracy of the estimated R_s . The proposed modification to Glasø’s p_b model ([48]) is recommended for reservoir fluids with surface gas CO₂ fraction above 15%. However, for smaller CO₂ contents, other models may be more suitable, as already observed in Figure 3.7.

The R_s curves from the differential liberation experiments shown in Figure 3.1 are reported again in Figure 3.8, along with the estimates obtained by Eqs. (3.7a) to (3.7g). The proposed R_s model is able to reproduce the experimental curve for reservoir fluids from all groups. It is especially worth noting that the proposed R_s model is able to reproduce the behavior of the experimental curve even in the pressure regions just below p_b , in contrast with the five models reported in Figure 3.1. Once more,

Table 3.8: Summary of the R_s prediction errors for the test set

R_s model	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Al-Marhoun [7]	-7.51	48.93	35.37	34.64	350.72
Al-Shammasi [9]	-0.42	28.30	22.08	17.69	196.36
Alakbari et al. [11]	26.53	104.27	40.93	99.50	198.42
Glasø [48]	10.74	23.92	21.05	15.65	148.94
Lasater [68]	5.83	131.76	43.89	124.37	2080.40
Petrosky and Farshad [98]	-4.18	21.34	14.57	16.15	146.88
Standing [118]	-1.30	23.77	18.68	14.76	165.33
Vazquez and Beggs [128]	17.07	16.75	19.65	13.64	121.23
Velarde et al. [129] - p_b est.	-37.39	36.33	38.88	34.73	347.25
Velarde et al. [129] - p_b exp.	1.78	26.69	12.01	23.90	92.39
Proposed R_s model - p_b est.	0.97	12.31	9.06	8.40	81.62
Proposed R_s model - p_b exp.	-3.33	12.09	8.35	9.36	71.31

Table 3.9: Proposed R_s model prediction errors, by group.

	Proposed R_s correlation with known p_b					Proposed R_s correlation with estimated p_b				
	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)	MPE (%)	σ_{PE} (%)	MAPE (%)	σ_{APE} (%)	RMSE (scf/STB)
Group I	-6.79	15.96	9.28	14.65	17.76	3.67	20.78	14.65	15.19	51.60
Group II	-4.46	12.85	7.59	11.28	37.04	7.57	15.21	11.34	12.66	73.49
Group III	1.22	5.41	3.53	4.28	14.08	0.69	14.62	6.78	12.97	25.07
Group IV	1.03	4.78	2.71	4.07	18.81	1.31	5.24	3.62	4.02	25.29
Group V	0.31	4.74	2.91	3.75	28.69	-0.12	5.11	3.77	3.45	33.97
Group VI	-0.59	5.51	4.38	3.39	46.04	70.56	10.43	7.37	7.40	158.64
Groups I to VI	-2.26	11.00	5.55	9.76	27.01	2.29	14.41	8.58	11.80	63.39
Test group	-3.33	12.09	8.35	9.36	71.31	0.97	12.31	9.06	8.40	81.62

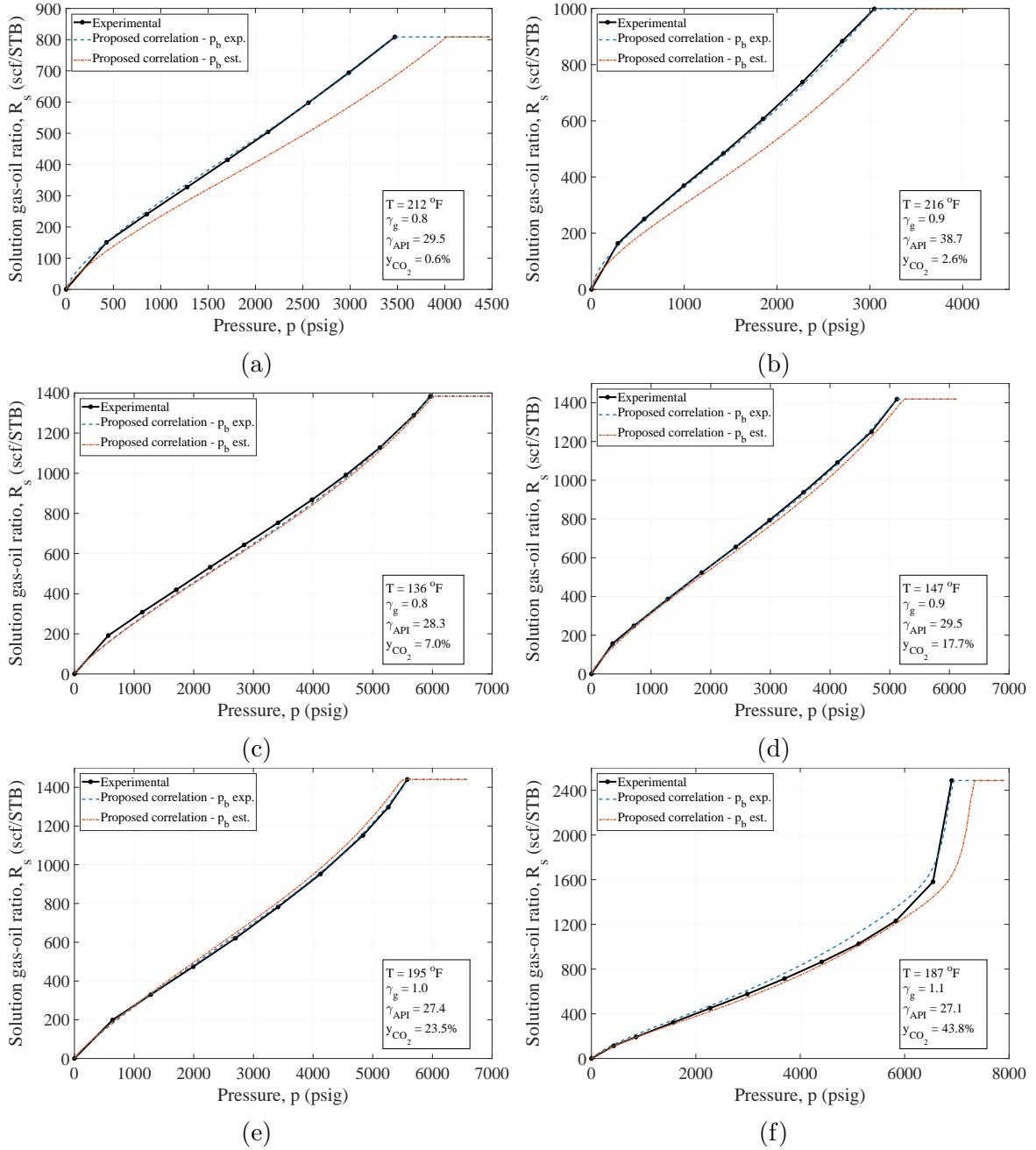


Figure 3.8: Experimental solution gas-oil ratio from a differential liberation and new model predictions for reservoir fluids representative of (a) Group I, (b) Group II, (c) Group III, (d) Group IV, (e) Group V, and (f) Group VI.

the error propagation of the p_b estimation into the R_s estimation can be observed. Nonetheless, even for unknown p_b the proposed model is able to provide satisfactory estimates. Despite the prediction of the R_s behavior specifically for reservoir fluids with high CO₂ content and high GOR being the main motivation for the development of the new model, it proved to be also suitable for reservoir fluids with little or no CO₂.

Given the crucial aspect of an accurate reservoir characterization in developing, monitoring, and optimizing production and reservoir management [32] and the complexity of the multiphase flow modeling and simulation in this context [50], the importance of using representative models for the fluid properties cannot be underestimated. Past R_s BO equations should not be used to model CO₂-rich reservoir fluids. Rather, the use of the proposed R_s model is strongly recommended for this subset.

3.6 Conclusion

Many attempts have been made to propose a general model for the solution gas-oil ratio over the past several decades, with varying degrees of success. However, none of these models used high CO₂ and high GOR reservoir fluids in their development nor were they evaluated for this particular subset of fluids. This Chapter fills this gap in the literature by analysing the performance of traditional R_s BO models on this subset of reservoir fluids and by proposing a general R_s model, suitable for CO₂-rich reservoir fluids. Major conclusions are summarized below.

- We investigated the performance of nine popular BO R_s models by comparing their predictions against 1457 experimental data points (training dataset), among which 829 data points are from reservoir fluids with CO₂ molar content ranging from 11% to 38% (15% to 45% in the surface gas) and GOR from 1117 scf/STB to 2487 scf/STB. The R_s models considered are Al-Marhoun's [7], Al-Shammasi's [9], Alakbari et al.'s [11], Glasø's [48], Lasater's [68], Petrosky and Farshad's [98], Standing's [118], Vazquez and Beggs's [128], and Velarde et al.'s [129].

- These existing R_s models yielded unacceptably high prediction errors for CO₂-rich reservoir fluids, with some even exhibiting physically inconsistent behaviors.
- We proposed a general R_s model, suitable for CO₂-rich reservoir fluids. It requires as input: pressure, temperature, bubble point pressure, surface gas specific gravity, stock-tank oil API gravity, and reservoir fluid gas-oil ratio.
- The proposed R_s model is general and outperforms traditional models even for reservoir fluids with low or no CO₂ content. This performance was evaluated on a test dataset containing 173 data points with CO₂ molar content in the surface gas ranging from 0% to 44% and GOR from 19 to 2409 scf/STB. Its mean average percentage error was 6% in the training set and 8% in the test set, and the root mean square error was 27 scf/STB and 71 scf/STB, respectively.
- We also proposed a model for the bubble point pressure specific for reservoir fluids whose stock-tank gas has a CO₂ molar content greater than 15%. Otherwise, for $y_{CO_2} < 15\%$, other models may be more suitable for estimating p_b .
- The bubble point pressure input to the proposed R_s model can be estimated through such empirical p_b models in case experimental values are not available. However, if a measured p_b is available, it should be used as input to avoid error propagation in the R_s estimation.
- If the proposed p_b model is used to estimate this input, the R_s model provides a mean average percentage error of 9% in the training set and 9% in the test set, and a root mean square error of 63 scf/STB and 82 scf/STB, respectively.
- Representative fluid models should be used in multiphase flow simulations, because an erroneous prediction of the fluid thermophysical behavior can have a significant negative impact on simulation results and, consequently, on economic and feasibility studies of the design system, flow assurance and equipment integrity analyses, and operation optimization decisions, among others. In this sense, the use of the proposed R_s model is strongly recommended for CO₂-rich reservoir fluids.

Chapter 4

Uncertainty in compositional fluid modeling

The content of this chapter follows closely the manuscript:

Lívia Paiva Fulchignoni, Daniel M. Tartakovsky. Uncertain characterization of reservoir fluids due to brittleness of equation of state regression. *Geoenergy Science and Engineering*, 2023 (accepted).

This chapter demonstrates that equations of state (EoS) predictions are inherently uncertain, even after tuning the model to a limited set of experimental data points. We investigate these uncertainties by explicitly describing the EoS regression as a nonlinear and non-convex multivariate optimization problem, considering five properties of the heaviest hydrocarbon fraction as design variables. We explore the impact of different optimization algorithms and initial guesses on the regression process and demonstrate that these subjective choices can significantly affect the predictive capabilities of the resulting models in unsampled pressure and temperature regions. Although all the optimization algorithms and initial guesses considered match experimental data for the gas and liquid properties, the resulting EoS parameterizations can yield vastly different predictions of the fluid's thermophysical behavior. To address this issue, we propose a probabilistic treatment of the design variables to quantify

the predictive uncertainty of the resulting fluid models.

4.1 Introduction

Estimation of phase equilibria and thermophysical properties of complex fluid mixtures is central to a plethora of applications such as design of fuel/biofuel blends [116], planning carbon capture and storage operations [137], and quantitative forecasting of multiphase flow in the oil and gas industry [111]. This task typically relies on an equation of state (EoS) to predict a complex fluid's pressure, volume and temperature (PVT) behavior at the macroscopic level. Parameters of a postulated EoS model are adjusted to match experimental data via a fitting/tuning procedure (regression), in which pre-selected variables are allowed to vary within a certain interval.

In a typical application, financial and operational constraints result in available experimental data that cover only a small subset of the PVT conditions of interest. The use of a data-tuned EoS in the unsampled PVT regions introduces uncertainty in predictions of fluid behavior. Partial knowledge of a complex fluid's composition is another source of predictive uncertainty of an EoS. For example, in petroleum fluid mixtures, it is only possible to make a component analysis for the lighter fractions of the reservoir fluid [91], while the heavier fractions are lumped together in the so-called C_{plus} fraction. Thus, petroleum compositions are most often reported to the C₇₊, C₁₀₊, or C₂₀₊ fractions and, in rare cases, to the C₃₀₊ fraction [90]. The data reported in this way pose an additional challenge to the EoS parameterization.

Commercial thermodynamic simulators, such as **WinProp**, **Multiflash** or **PVTsim**, offer modules for the EoS regression and for the C_{plus} fraction characterization (splitting and lumping procedures). For instance, **WinProp** allows the regression process to adjust the fluid components' critical pressure P_{cr} , critical volume V_{cr} , critical temperature T_{cr} , acentric factor ω , molecular weight M , among other properties. It also allows the binary interaction parameter (BIP) to be included as fitting variables. While a standard approach is to include into the tuning process the most uncertain variables, e.g., the parameters related to the characterization of the C_{plus} fraction, its optimal implementation remains unsettled.

Several EoS regression procedures have been proposed in the literature. Coats and Smart [30] state that it is usually sufficient to perform the cubic EoS regression on the Ω_a and Ω_b parameters associated to the methane and plus fraction components and on the methane-plus fraction BIP. Whitson and Brûlé [133] suggest fitting the Ω_a and Ω_b parameters of the next-to-heaviest C_{plus} fraction instead of those of the methane or, alternatively, fitting P_{cr} and T_{cr} . Christensen [27] proposes first to fit the M_{C7+} used for splitting and the volume shift parameters of the expanded components to then adjust the two most sensitive coefficients in empirical correlations for P_{cr} , T_{cr} and ω of the C_{7+} expanded components. Zurita and McCain [4] recommend tuning first M of the plus fraction and then P_{cr} , T_{cr} and ω of single carbon number (SCN) groups up to C_{45+} . Al-Meshari [8] modifies the latter method by adjusting ω of the heaviest multiple carbon number group to match the saturation pressure, instead of its critical properties. This ambiguity is a reason why the detailed process of the EoS parameter tuning is more of an art than a science [35].

Previous attempts to bring some rigor into the EoS regression involve the assessment of the ability of alternative tuned EoS models to reproduce experimental data. For example, a comparison of the performance of the alternative EoS regression procedures [30, 133, 27, 4, 8] on 30 fluid samples (19 oil and 11 gas) suggest that the approaches from [8] and [133] perform best for oil and gas fluids, respectively [13]. Such comparative studies ignore the impact of the selection of an optimization algorithm and initial guess for the design variables.

Our study focuses on these aspects of the EoS regression. The importance of this hitherto unexplored aspect of complex fluid modeling stems from the fact that the EoS regression is a nonlinear, non-convex, multivariate optimization problem that poses multiple local minima. Since different optimization algorithms employ alternative approaches to explore the search space, the choice of an optimization algorithm and its hyperparameters is known to affect the optimization trajectory and final destination. An initial guess can also impact the optimization path. In Section 4.2, we describe four representative optimization algorithms considered in this work: the adaptive moment estimation method (ADAM) [64], the Davidon-Fletcher-Powell method (DFP) [33, 40], the covariance matrix adaptation evolution strategy method (CMA-ES) [55], and

the direct search method [57] (DSM). Section 4.3 contains our assessment of the relative performance of these conceptually distinct optimization algorithms on the EoS tuning of two hydrocarbon reservoir fluids. Main conclusions drawn from this study are summarized in Section 4.5.

4.2 Methodology

To be specific, we consider the modified Peng-Robinson EoS [96, 105], described in Section 2.2, because of its widespread usage and simplicity. The properties of pure hydrocarbon components, generalized single carbon number fractions, and light gases are reported in Table 4.1.

For given P and T , the molar volume v predicted with Eq. (2.7) is shifted to v_{Pen} according to [95, 62], as described in Eq. 2.8. Finally, densities are computed based on the components molecular weights.

4.2.1 Problem formulation

The design variables used to fit the EoS (2.7) to data are $P_{\text{cr},i}$, $T_{\text{cr},i}$, ω_i , c_i , and the molecular weight M_i for the i th pseudocomponent corresponding to the heaviest hydrocarbon fraction, the C_{plus} fraction. The resulting set of design variables is $\{P_{\text{cr},C_{\text{plus}}}, T_{\text{cr},C_{\text{plus}}}, \omega_{C_{\text{plus}}}, c_{C_{\text{plus}}}, M_{C_{\text{plus}}}\}$. A design point $\mathbf{x} = (P_{\text{cr},C_{\text{plus}}}^*, T_{\text{cr},C_{\text{plus}}}^*, \omega_{C_{\text{plus}}}^*, c_{C_{\text{plus}}}^*, M_{C_{\text{plus}}}^*)^\top$ is a vector containing these design variables non-dimensionalized by their initial values, so that $\mathbf{x} = \mathbb{1}$ at the first iteration.

Eq. (2.7) is solved for a series of pressure and temperature conditions, under some of which the fluid mixture separates into a gas phase and a liquid phase. When two phases are present, the molar volumes of the gas and liquid phases, $v_{\text{Pen},g}$ and $v_{\text{Pen},l}$, are computed by the EoS, besides of the liquid fraction l obtained through the equality of the components fugacity in each phase, the fundamental criteria for phase equilibria. From these set of solutions, fluid properties measured experimentally are predicted. The optimization is carried by minimizing the relative error for N_{meas}

Table 4.1: Properties of the reservoir fluid components (CO_2 , N_2 , and hydrocarbons up to C_{19} fraction) [89]

Component	Critical pressure (bar)	Critical temperature (K)	Acentric factor	Molecular weight
CO_2	73.8	304.2	0.225	44.0
N_2	33.9	126.2	0.040	28.0
C_1	46.0	190.6	0.008	16.0
C_2	48.8	305.4	0.098	30.1
C_3	42.5	369.8	0.152	44.1
i-C ₄	36.5	408.1	0.176	58.1
C ₄	38.0	425.2	0.193	58.1
i-C ₅	33.8	460.4	0.227	72.2
C ₅	33.7	469.5	0.251	72.2
C ₆	29.7	507.4	0.296	86.2
C ₇	31.1	548.0	0.280	95.0
C ₈	28.8	575.0	0.312	106.0
C ₉	26.3	603.0	0.348	116.0
C ₁₀	24.2	626.0	0.385	133.0
C ₁₁	22.3	648.0	0.419	152.0
C ₁₂	20.8	668.0	0.454	164.0
C ₁₃	19.6	687.0	0.484	179.0
C ₁₄	18.6	706.0	0.516	193.0
C ₁₅	17.6	724.0	0.550	209.0
C ₁₆	16.6	740.0	0.582	218.0
C ₁₇	15.9	755.0	0.613	239.0
C ₁₈	15.3	767.0	0.638	250.0
C ₁₉	14.8	778.0	0.662	264.0

measurements,

$$\mathcal{E}_k(\mathbf{x}) = \frac{e_k(\mathbf{x}) - y_k}{y_k}, \quad k = 1, \dots, N_{\text{meas}}, \quad (4.1)$$

between $e_k(\mathbf{x})$, the fluid property predicted by the EoS with design variables \mathbf{x} , and y_k , the k th experimental data point. Following [3], we consider the objective function

$$F(\mathbf{x}) = \sum_{k=1}^{N_{\text{meas}}} w_k^2 \mathcal{E}_k^2(\mathbf{x}), \quad (4.2)$$

where w_k is the weight assigned to the k th measurement; we assign equal weights to all measurements, such that $w_1 = \dots = w_{N_{\text{meas}}} = 1/N_{\text{meas}}$.

The upper and lower bounds for each design variable are defined as a percentage of their initial guesses. A maximum variation of 20% is permitted for $P_{\text{cr},C_{\text{plus}}}^*$, $T_{\text{cr},C_{\text{plus}}}^*$, $\omega_{C_{\text{plus}}}^*$, and $M_{C_{\text{plus}}}^*$; and 50% for $c_{C_{\text{plus}}}^*$. Representing these limits by the vector $\Delta\mathbf{x}$, we formulate the EoS regression as a constrained optimization problem

$$\underset{\mathbf{x}}{\text{minimize}} \ F(\mathbf{x}), \quad \text{subject to} \quad 1 - \Delta x_i \leq x_i \leq 1 + \Delta x_i \quad \text{for all } i. \quad (4.3)$$

The initial values for $P_{\text{cr},C_{\text{plus}}}$ and $T_{\text{cr},C_{\text{plus}}}$ are obtained through Twu's correlation [125]; for $\omega_{C_{\text{plus}}}$ through Lee-Kesler's correlation [63]; for the volume shift parameter $c_{C_{\text{plus}}}$ for all components through Jhaveri-Youngren's correlation [62]. The initial guess for $M_{C_{\text{plus}}}$ is its measured value reported in the PVT analysis.

The stopping criteria for the optimization process is either the convergence of the objective function F within a tolerance of 10^{-8} for the change in $|F|$ between two consecutive iterations or a maximum number of 100 iterations.

4.2.2 Alternative minimization strategies

We discuss the four alternative minimization algorithms used in our comparative study in this section. They are selected because of both their popularity in the field and their conceptual dissimilarity from each other. For a fair comparison, their hyperparameters were optimized for this particular application.

The adaptive moment estimation method [64] ADAM is a first-order method that computes individual adaptive learning rates for each parameter from estimates of the first and second moments of the gradients of the objective function. At each iteration, the intermediate variables

$$\mathbf{m}_{t+1} \leftarrow \beta_1 \mathbf{m}_t + (1 - \beta_1) \nabla F(\mathbf{x}_t), \quad \hat{\mathbf{m}}_{t+1} \leftarrow \frac{\mathbf{m}_{t+1}}{(1 - \beta_1^{t+1})}, \quad (4.4a)$$

$$\mathbf{v}_{t+1} \leftarrow \beta_2 \mathbf{v}_t + (1 - \beta_2) \nabla F(\mathbf{x}_t) \odot \nabla F(\mathbf{x}_t), \quad \hat{\mathbf{v}}_{t+1} \leftarrow \frac{\mathbf{v}_{t+1}}{(1 - \beta_2^{t+1})} \quad (4.4b)$$

are computed in order to update the design point,

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \alpha \hat{\mathbf{m}}_{t+1} \oslash (\sqrt{\hat{\mathbf{v}}_{t+1}} + \epsilon). \quad (4.4c)$$

The symbols \odot and \oslash refer to the element-wise vector product and division, respectively. Algorithm 1 presents a pseudocode for the ADAM optimization algorithm for a deterministic objective function.

The Davidon-Fletcher-Powell method [33, 40] DFP is a second-order gradient descent method. It rests on Newton's method,

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - (\mathbf{H}_t)^{-1} \nabla F(\mathbf{x}_t), \quad (4.5)$$

where the inverse of the Hessian matrix \mathbf{H} is approximated by a symmetric and positive definite matrix \mathbf{Q} . Because the second-order information is approximated, it is called a quasi-Newton method. At each iteration t , the design point is updated according to

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \alpha_t \mathbf{Q}_t \nabla F(\mathbf{x}_t), \quad (4.6)$$

where α_t is a scalar step factor. Algorithm 2 presents a pseudocode for the DFP optimization algorithm.

Algorithm 1 Adaptive Moment Estimation Method (ADAM) optimization algorithm [64]

```

1: Require:  $\alpha > 0$ : hyperparameter for the step size
2: Require:  $\epsilon > 0$ : small value ( $\sim 10^{-8}$ ) to prevent division by zero
3: Require:  $\beta_1, \beta_2 \in [0, 1]$ : hyperparameters for the exponential decay rates for the
   moment estimates
4: Require:  $F(\mathbf{x})$ : Objective function with parameters  $\mathbf{x}$ 
5: Require:  $\mathbf{x}_0$ : Initial guess
6:  $\mathbf{m}_0 \leftarrow 0$  (Initialize 1st moment vector)
7:  $\mathbf{v}_0 \leftarrow 0$  (Initialize 2nd moment vector)
8:  $t \leftarrow 0$  (Initialize iteration counter)
9: while stopping criteria not reached do
10:    $t \leftarrow t + 1$  (Increment iteration counter)
11:    $\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \nabla F(\mathbf{x}_{t-1})$  (Update biased first moment estimate)
12:    $\mathbf{v}_t \leftarrow \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \nabla F(\mathbf{x}_{t-1}) \odot \nabla F(\mathbf{x}_{t-1})$  (Update biased second raw
      moment estimate)
13:    $\hat{\mathbf{m}}_t \leftarrow \mathbf{m}_t / (1 - (\beta_1)^t)$  (Compute bias-corrected first moment estimate)
14:    $\hat{\mathbf{v}}_t \leftarrow \mathbf{v}_t / (1 - (\beta_2)^t)$  (Compute bias-corrected second raw moment estimate)
15:    $\mathbf{x}_t \leftarrow \mathbf{x}_{t-1} - \alpha \hat{\mathbf{m}}_t \oslash (\sqrt{\hat{\mathbf{v}}_t} + \epsilon)$  (Update parameters)
16: end while
17: Return  $\mathbf{x}_t$  (Resulting parameters)

```

Algorithm 2 Davidon-Fletcher-Powell (DFP) optimization algorithm [33, 40]

```

1: Require:  $F(\mathbf{x})$ : Objective function with parameters  $\mathbf{x}$ 
2: Require:  $\mathbf{x}_0$ : Initial guess
3: Require:  $\mathbf{Q}_0$ : Initial inverse Hessian approximation
4:  $t \leftarrow 0$  (Initialize iteration counter)
5: while stopping criteria not reached do
6:    $\mathbf{g}_t \leftarrow \nabla F(\mathbf{x}_t)$  (Compute gradient at current point)
7:    $\mathbf{d}_t \leftarrow -\mathbf{Q}_t \mathbf{g}_t$  (Compute search direction)
8:    $\alpha_t \leftarrow \underset{\alpha_t}{\text{minimize}} F(\mathbf{x}_t + \alpha_t \mathbf{d}_t)$  (Compute the step size)
9:    $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t + \alpha_t \mathbf{d}_t$  (Update parameter vector)
10:   $\mathbf{g}_{t+1} \leftarrow \nabla F(\mathbf{x}_{t+1})$  (Compute gradient at new point)
11:   $\boldsymbol{\delta}_t \leftarrow \mathbf{x}_{t+1} - \mathbf{x}_t$  (Compute change in parameter vector)
12:   $\boldsymbol{\gamma}_t \leftarrow \mathbf{g}_{t+1} - \mathbf{g}_t$  (Compute change in gradient)
13:   $\mathbf{Q}_{t+1} \leftarrow \mathbf{Q}_t + \frac{\boldsymbol{\delta}_t \boldsymbol{\delta}_t^\top}{\boldsymbol{\delta}_t^\top \boldsymbol{\gamma}_t} - \frac{\mathbf{Q}_t \boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top \mathbf{Q}_t}{\boldsymbol{\gamma}_t^\top \mathbf{Q}_t \boldsymbol{\gamma}_t}$  (Update inverse Hessian approximation)
14:   $t \leftarrow t + 1$  (Increment iteration counter)
15: end while
16: Return  $\mathbf{x}_t$ 

```

The covariance matrix adaptation evolution strategy method [55] CMA-ES is a stochastic method inspired by natural evolution strategies. It relies on recombination, mutation and elite selection techniques. At each iteration step, the algorithm improves the parameters of a multivariate Gaussian search distribution. The change rates for the mean and the covariance of the search distribution and for the step size are updated separately. The CMA-ES is recognized to be among the leading algorithms for optimization of real-valued functions [39].

Algorithms 3 and 4 present a pseudocode for the CMA-ES optimization algorithm. A Matlab implementation of the CMA-ES algorithm can be found in [55].

Direct search method The Hooke-Jeeves algorithm [57], herein labeled DSM, performs a direct search over the search space based on evaluations at steps of magnitude α in each coordinate direction. At each iteration, for $\mathbf{x} \in \mathbb{R}^5$, $F[\mathbf{x} + (\alpha, 0, 0, 0, 0)^\top]$, $F[\mathbf{x} + (-\alpha, 0, 0, 0, 0)^\top]$, ..., $F[\mathbf{x} + (0, 0, 0, 0, \alpha)^\top]$ and $F[\mathbf{x} + (0, 0, 0, 0, -\alpha)^\top]$ are evaluated. The anchoring point \mathbf{x} thus moves to the position in which the objective function is smaller, if any improvement is found. If no improvements are verified, the

Algorithm 3 Covariance Matrix Adaptation Evolution Strategy (CMA-ES) optimization algorithm [55] (Part 1)

- 1: **Require:** $F(\mathbf{x})$: Objective function with parameters \mathbf{x}
 - 2: **Require:** \mathbf{x}_0 : Initial guess
 - 3: **Require:** $\sigma > 0$: Hyperparameter for step size
 - 4: **Require:** λ : Hyperparameter for sample size
 - 5: **Require:** μ : Hyperparameter for elite sample size
 - 6: $n \leftarrow \text{length}(\mathbf{x})$ (Problem dimension)
 - 7: $w'_i \leftarrow \ln \frac{\lambda + 1}{2} - \ln i \quad \text{for } i = 1, \dots, \lambda$
 - 8: $\mu_{\text{eff}} \leftarrow \frac{(\sum_{i=1}^{\mu} w'_i)^2}{\sum_{i=1}^{\mu} w'^2_i}$ (Variance-effective size of μ)
 - 9: $c_m \leftarrow 1$
 - 10: $c_{\sigma} \leftarrow \frac{\mu_{\text{eff}} + 2}{n + \mu_{\text{eff}} + 5}$
 - 11: $d_{\sigma} \leftarrow 1 + 2 \max \left(0, \sqrt{\frac{\mu_{\text{eff}} - 1}{n + 1}} - 1 \right) + c_{\sigma}$
 - 12: $c_{\Sigma} \leftarrow \frac{4 + \mu_{\text{eff}}/n}{n + 4 + 2\mu_{\text{eff}}/n}$
 - 13: $\alpha_{\text{cov}} \leftarrow 2$
 - 14: $c_1 \leftarrow \frac{\alpha_{\text{cov}}}{(n + 1.3)^2 + \mu_{\text{eff}}}$
 - 15: $c_{\mu} \leftarrow \min \left(1 - c_1, \alpha_{\text{cov}} \frac{1/4 + \mu_{\text{eff}} + 1/\mu_{\text{eff}} - 2}{(n + 2)^2 + \alpha_{\text{cov}}\mu_{\text{eff}}/2} \right)$
 - 16: $\alpha_{\mu}^- = 1 + c_1/c_{\mu}$
 - 17: $\alpha_{\mu_{\text{eff}}}^- = 1 + \frac{2\mu_{\text{eff}}}{\mu_{\text{eff}} + 2}$
 - 18: $\alpha_{\text{pos def}}^- = \frac{1 - c_1 - c_{\mu}}{nc_{\mu}}$
 - 19: $w_i = \begin{cases} \frac{1}{\sum_{i=j}^{\lambda} |w'_j|^+} w'_i & \text{if } w'_i \geq 0 \\ \frac{\min(\alpha_{\mu}^-, \alpha_{\mu_{\text{eff}}}^-, \alpha_{\text{pos def}}^-)}{\sum_{j=1}^{\lambda} |w'_j|^-} w'_i & \text{if } w'_i < 0 \end{cases}, i = 1, \dots, \lambda$
 - 20: $E \leftarrow \sqrt{n} \left(1 - \frac{1}{4n} + \frac{1}{21n^2} \right)$
 - 21: $\mathbf{p}_{\sigma}, \mathbf{p}_{\Sigma}, \mathbf{\Sigma} \leftarrow \mathbf{0}, \mathbf{0}, \mathbf{I}$
 - 22: $t \leftarrow 0$ (Initialize iteration counter)
 - 23: $\mathbf{m}_t \leftarrow \mathbf{x}_0$ (Initialize distribution mean)
-

Algorithm 4 Covariance Matrix Adaptation Evolution Strategy (CMA-ES) optimization algorithm [55] (Part 2)

```

24: while stopping criteria not reached do
25:   Sample  $\mathbf{x}_k \sim \mathcal{N}(\mathbf{m}, \sigma^2 \Sigma)$ ,  $k = 1, \dots, \lambda$ 
26:    $\mathbf{F} \leftarrow \text{Evaluate}(F, \{\mathbf{x}_k\}_{k=1}^\lambda)$  (Evaluate objective function)
27:    $\mathbf{idx} \leftarrow \text{SortIndices}(\mathbf{F})$  (Sort indices by objective function values)
28:    $\mathbf{y}_k \leftarrow (\mathbf{x}_k - \mathbf{m})/\sigma$ ,  $k = 1, \dots, \lambda$ 
29:    $\langle \mathbf{y} \rangle_w \leftarrow \sum_{i=1}^\lambda w_i \mathbf{y}_{\mathbf{idx}_i}$ 
30:    $\mathbf{m}_{t+1} \leftarrow \mathbf{m} + c_m \sigma \langle \mathbf{y} \rangle_w$  (Update distribution mean)
31:    $\mathbf{p}_\sigma \leftarrow (1 - c_\sigma) \mathbf{p}_\sigma + \sqrt{c_\sigma (2 - c_\sigma)} \mu_{\text{eff}} \Sigma^{-\frac{1}{2}} \langle \mathbf{y} \rangle_w$ 
32:    $\sigma \leftarrow \sigma \times \exp\left(\frac{c_\sigma}{d_\sigma} \left(\frac{\|\mathbf{p}_\sigma\|}{E} - 1\right)\right)$  (Update step size)
33:    $h_\sigma \leftarrow \begin{cases} 1 & \text{if } \frac{\|\mathbf{p}_\sigma\|}{\sqrt{1-(1-c_\sigma)^2(t+1)}} < (1.4 + \frac{2}{n+1}) E \\ 0 & \text{otherwise} \end{cases}$ 
34:    $\mathbf{p}_\Sigma \leftarrow (1 - c_\Sigma) \mathbf{p}_\Sigma + h_\sigma \sqrt{c_\Sigma (2 - c_\Sigma)} \mu_{\text{eff}} \langle \mathbf{y} \rangle_w$ 
35:    $w_i^\circ \leftarrow w_i \times \left(1 \text{ if } w_i \geq 0 \text{ else } n / \left\| \Sigma^{-\frac{1}{2}} \mathbf{y}_{\mathbf{idx}_i} \right\|^2 \right)$ ,  $i = 1, \dots, \lambda$ 
36:    $\Sigma \leftarrow \underbrace{\left( 1 + c_1(1 - h_\sigma)c_\Sigma(2 - c_\Sigma) - c_1 - c_\mu \sum_{i=1}^\lambda w_i \right)}_{\text{usually equals to 0}} \Sigma + c_1 \mathbf{p}_\Sigma \mathbf{p}_\Sigma^\top + c_\mu \sum_{i=1}^\lambda w_i^\circ \mathbf{y}_{\mathbf{idx}_i} \mathbf{y}_{\mathbf{idx}_i}^\top$ 
37:    $t \leftarrow t + 1$  (Increment iteration counter)
38: end while
39: Return  $\mathbf{x}_t$ 

```

step size decreases for a finer search.

Algorithm 5 presents a pseudocode for the Hooke-Jeeves optimization algorithm.

Algorithm 5 Hooke-Jeeves optimization algorithm

```

1: Require:  $F(\mathbf{x})$ : Objective function with parameters  $\mathbf{x}$ 
2: Require:  $\mathbf{x}_0$ : Initial guess
3: Require:  $\alpha > 0$ : Hyperparameter for initial step size
4: Require:  $\gamma > 0$ : Hyperparameter for step decay
5: Require:  $\epsilon > 0$ : Convergence tolerance
6:  $t \leftarrow 0$  (Initialize iteration counter)
7: while  $\alpha > \epsilon$  and stopping criteria not reached do
8:   improved  $\leftarrow$  False
9:    $\mathbf{x}_B \leftarrow \mathbf{x}_t$ 
10:   $y_B \leftarrow F(\mathbf{x}_t)$ 
11:  for each dimension  $i$  do
12:    for sgn in  $\{-1, 1\}$  do
13:       $\mathbf{x}_T \leftarrow \mathbf{x}_t + \text{sgn } \alpha \hat{\mathbf{e}}_i$  (Exploratory move)
14:       $y_T \leftarrow F(\mathbf{x}_T)$ 
15:      if  $y_T < y_B$  then
16:         $\mathbf{x}_B \leftarrow \mathbf{x}_T$  (Update best solution)
17:         $y_B \leftarrow y_T$ 
18:        improved  $\leftarrow$  True (Indicate improvement)
19:      end if
20:    end for
21:  end for
22:   $\mathbf{x}_{t+1} \leftarrow \mathbf{x}_B$  (Update current solution)
23:  if not improved then
24:     $\alpha \leftarrow \gamma \alpha$  (Reduce step size)
25:  end if
26:   $t \leftarrow t + 1$  (Increment iteration counter)
27: end while
28: Return  $\mathbf{x}_t$ 

```

4.2.3 Numerical implementation

We implement the Peng-Robinson EoS [96, 105] and the optimization algorithms in the Python and Julia programming languages, respectively. The Julia optimization routine calls the Python EoS routine at each evaluation of the objective function to compute the EoS predictions. Although not shown here, we validated our implementation against the commercial simulator `PVTsim`.

4.2.4 Data description

We investigate two reservoir fluid compositions. The first is a slightly volatile oil with a bubble point pressure of 181.67 bar at 104°C and solution gas-oil ratio of 134 Sm³/Sm³ from [133], denoted here as fluid A. The second composition, denoted here as fluid B, refers to a hydrocarbon reservoir fluid from the northern Campos basin offshore Brazil. It has a bubble point pressure of 186.7 bar at 76°C and solution gas-oil ratio of 48 Sm³/Sm³ [97]. Their compositions are described in Table 4.2, along with the density and molecular weight of the plus fractions.

The data used for the EoS tuning consist of measurements from a differential liberation: residual oil API, solution gas-oil ratio R_s , oil formation volume factor B_o , gas formation volume factor B_g , oil density ρ_o , gas specific gravity γ_g (air = 1), and gas compressibility factor Z . A total of 69 and 83 experimental points were used in the regression of fluids A and B, respectively.

Non-hydrocarbon-hydrocarbon BIPs are set to their tabulated values in [133]. Hydrocarbon-hydrocarbon BIPs are set to zero, except for fluid A between methane and C₇₊, which is computed according to [15]. Volume shift parameters were defined according to [133] for fluid A and computed according to [62] for fluid B.

We treat the C_{plus} fraction as a single pseudocomponent as done in [105, 109, 119]. While this approach has its detractors [4, 27, 8, 144], three main reasons justify our choice. First, our goal is to assess how the choice of an optimization routine and the initial guess affect EoS predictions of the PVT fluid behavior. Demonstrating that different combinations of values for the C_{plus} parameters can similarly represent

Table 4.2: Molar composition and plus fraction properties of reservoir fluids A [133] and B [97]

Component	Mole fraction (%)	
	Fluid A	Fluid B
CO ₂	0.91	0.02
N ₂	0.16	0.21
C ₁	36.47	41.50
C ₂	9.67	2.35
C ₃	6.95	0.78
i-C ₄	1.44	0.32
C ₄	3.93	0.40
i-C ₅	1.44	0.19
C ₅	1.41	0.19
C ₆	4.33	0.31
C ₇	33.29*	0.83
C ₈	—	2.14
C ₉	—	1.30
C ₁₀	—	1.32
C ₁₁	—	1.31
C ₁₂	—	1.44
C ₁₃	—	1.62
C ₁₄	—	1.50
C ₁₅	—	1.65
C ₁₆	—	1.41
C ₁₇	—	1.48
C ₁₈	—	1.32
C ₁₉	—	1.10
C ₂₀₊	—	35.32†

* Plus fraction. Density = 0.8515 g/cm³
at 60°F. Molecular weight = 218.

† Plus fraction. Density = 0.9811 g/cm³
at 60°F. Molecular weight = 505.

experimental data is a simple and intuitive way to meet this purpose. In particular, we use experimental data measured at the differential liberation. Second, the plus-fraction splitting and lumping procedures are empirical and ultimately the C_{plus} characterization problem has only two independent variables: the measurements of the molecular weight and the density of the plus fraction. Since the lumping scheme influences the performance of the EoS regression on volumetric data [4], we ignore this source of uncertainty. Third, the single-component approach is still seen in the oil and gas industry (e.g., [70]).

4.3 Results

4.3.1 Optimization algorithm

Figure 4.1 exhibits fluid A’s convergence history of the normalized design variables $(P_{\text{cr},C_{\text{plus}}}^*, T_{\text{cr},C_{\text{plus}}}^*, \omega_{C_{\text{plus}}}^*, c_{C_{\text{plus}}}^*, M_{C_{\text{plus}}}^*)$ for alternative optimization methods: ADAM [64], DFP [33, 40], CMA-ES [55], and DSM [57]. Figure 4.2 does the same for fluid B. These algorithms not only have distinct optimization paths, but also yield different solutions (final values of the design variables); the latter are collated in Table 4.3 for both fluids. This performance is to be expected given the high degree of nonlinearity of the optimization problem.

The non-convexity of the objective function is depicted in Figure 4.3, in which the curve was generated using 200 objective function evaluations. It shows the objective function evaluated at convex combinations between the CMA-ES and DSM solutions for fluid B. A convex combination of two points refers to the point that lies along the line segment connecting these two points. The weight θ controls the position of the point on the segment, such that $\theta = 0$ corresponds to the CMA-ES solution and $\theta = 1$ to the DSM solution; the values of θ between 0 and 1 map out the landscape of the objective function between these two solutions. All in all, the results presented in Figures 4.1 - 4.3 and Table 4.3 demonstrate that results of the EoS regression depend on the choice of an optimization algorithm.

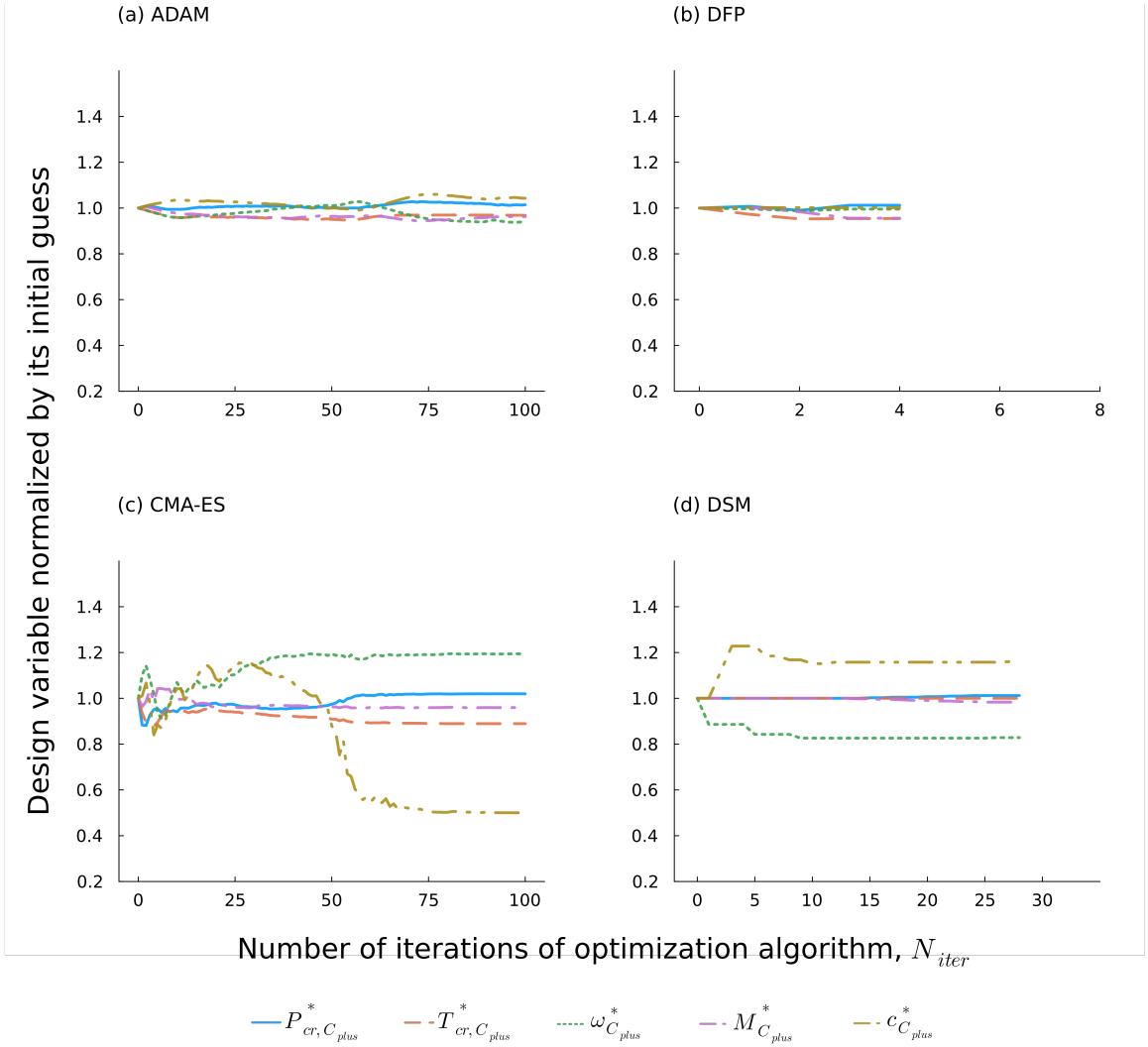


Figure 4.1: Convergence history of fluid A's normalized design variables ($P_{cr,C_{plus}}^*$, $T_{cr,C_{plus}}^*$, $\omega_{C_{plus}}^*$, $M_{C_{plus}}^*$, and $c_{C_{plus}}^*$) for alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

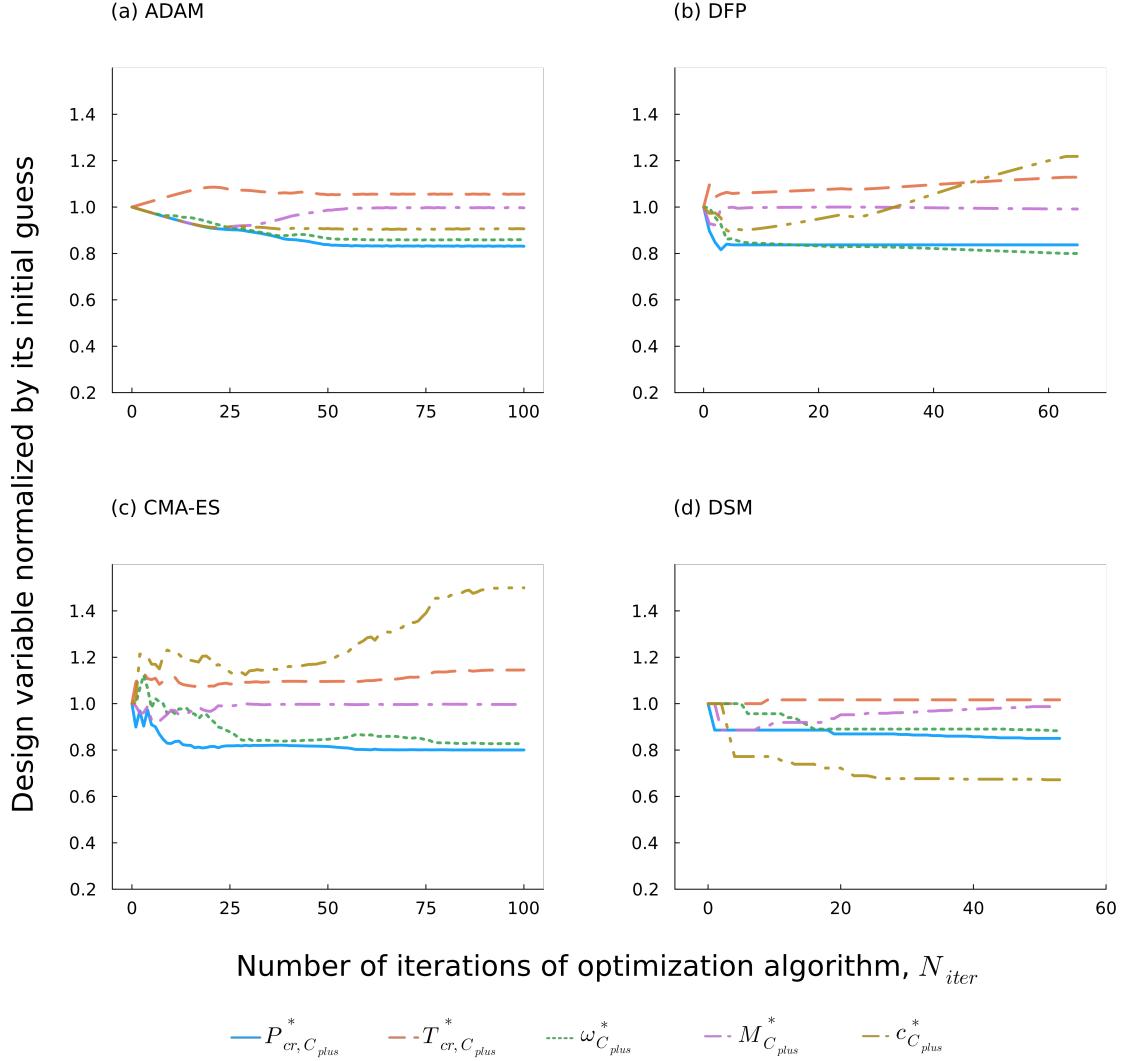


Figure 4.2: Convergence history of fluid B's normalized design variables ($P_{cr,C_{plus}}^*$, $T_{cr,C_{plus}}^*$, $\omega_{C_{plus}}^*$, $M_{C_{plus}}^*$, and $c_{C_{plus}}^*$) for alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

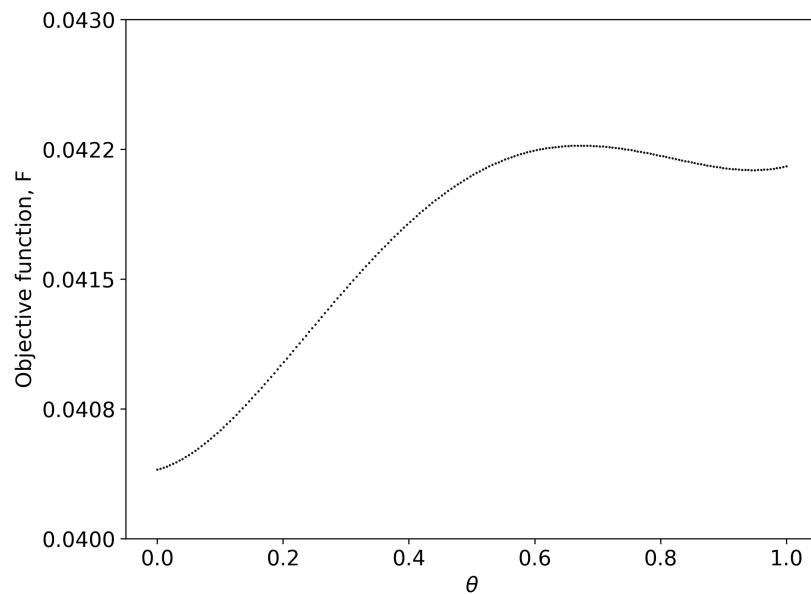


Figure 4.3: Objective function (squared relative error) evaluated for fluid B at convex combinations between the solutions obtained by the covariance matrix adaptation evolution strategy (CMA-ES) [55] ($\theta = 0$) and the direct search method (DSM) [57] ($\theta = 1$). The curve is generated using 200 objective function evaluations.

Table 4.3: EoS regression results obtained with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

Property	Initial value	ADAM	DFP	CMA-ES	DSM
Fluid A					
Critical pressure, $P_{\text{cr},C_{\text{plus}}}$ (bar)	17.39	17.64	17.60	17.73	17.60
Critical temperature, $T_{\text{cr},C_{\text{plus}}}$ (K)	772.24	747.76	736.81	686.74	772.24
Acentric factor, $\omega_{C_{\text{plus}}}$	0.7207	0.6778	0.7174	0.8609	0.5971
Molecular weight, $M_{C_{\text{plus}}}$	218	210	208	209	214
Volume shift, $c_{C_{\text{plus}}}$ (cm^3/mol)	44.2	46.1	44.3	22.1	51.4
Fluid B					
Critical pressure, $P_{\text{cr},C_{\text{plus}}}$ (bar)	13.10	10.88	10.96	10.48	11.14
Critical temperature, $T_{\text{cr},C_{\text{plus}}}$ (K)	986.40	1041.93	1113.39	1129.58	1002.72
Acentric factor, $\omega_{C_{\text{plus}}}$	1.2744	1.0944	1.0195	1.0542	1.1256
Molecular weight, $M_{C_{\text{plus}}}$	505	503	501	503	499
Volume shift, $c_{C_{\text{plus}}}$ (cm^3/mol)	133.5	121.0	162.7	200.2	90.0

Figure 4.4 shows the decay of the objective function F with the number of iterations, N_{iter} , for both fluids. It further elucidates the challenge posed by nonlinearity of the optimization problem (Equations 2.7 - 2.8). While the four optimization methods significantly reduce F from its initial value and yield essentially the same value of F at convergence, they do so for a distinctly different sets of the optimized design variables (Table 4.3). Since the objective function F is the sum of the square percentage errors, different sets of the optimized design variables match the experimental data with virtually the same level of accuracy.

Indeed, Figures 4.5 - 4.6 and Table 4.4 compare the tuned EoS predictions against experimental data. It can be observed that all of these optimization algorithms significantly improve experimental data representation compared to the original (not tuned) EoS model and yield similar predictions.

The objective function F is defined for pressure-temperature regimes in which experimental data are available. There is no guarantee that the design variables that minimize F are appropriate outside these regimes. We investigate the ability of the EoS models, tuned with the four alternative optimization techniques on “training”

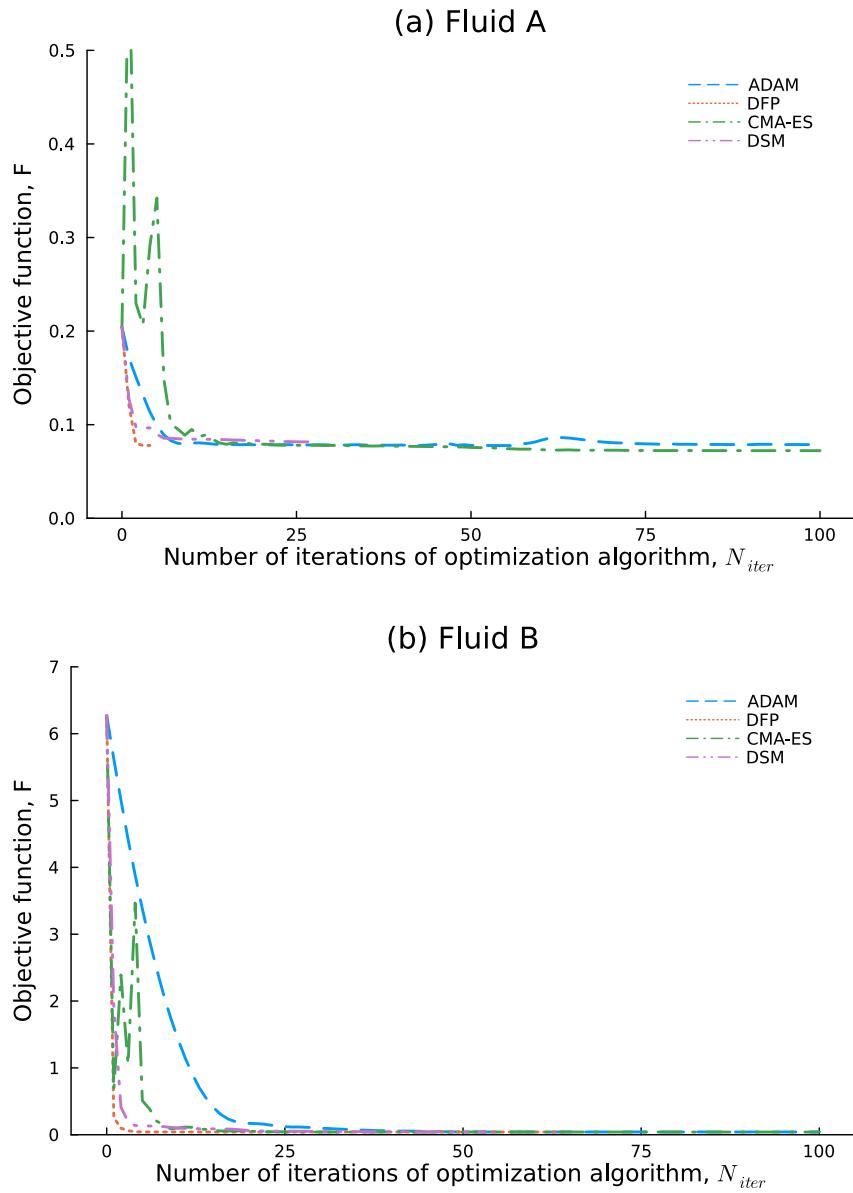


Figure 4.4: Convergence history of the objective function F for (a) fluid A and (b) fluid B, and four optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

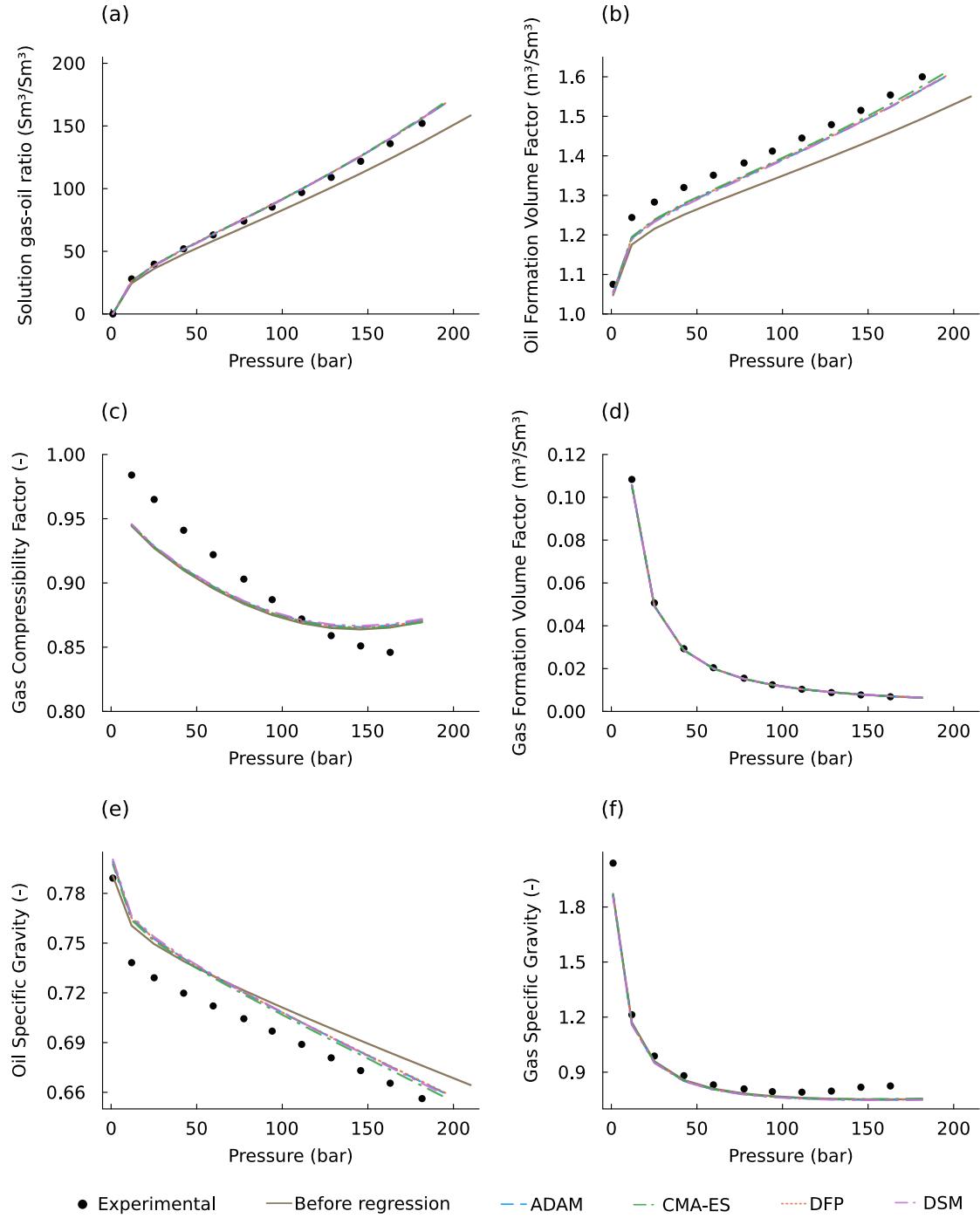


Figure 4.5: Fluid A EoS predictions and experimental data at 104°C for (a) solution gas-oil ratio, (b) oil formation volume factor, (c) gas compressibility factor, (d) gas formation volume factor, (e) oil specific gravity, and (f) gas specific gravity. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

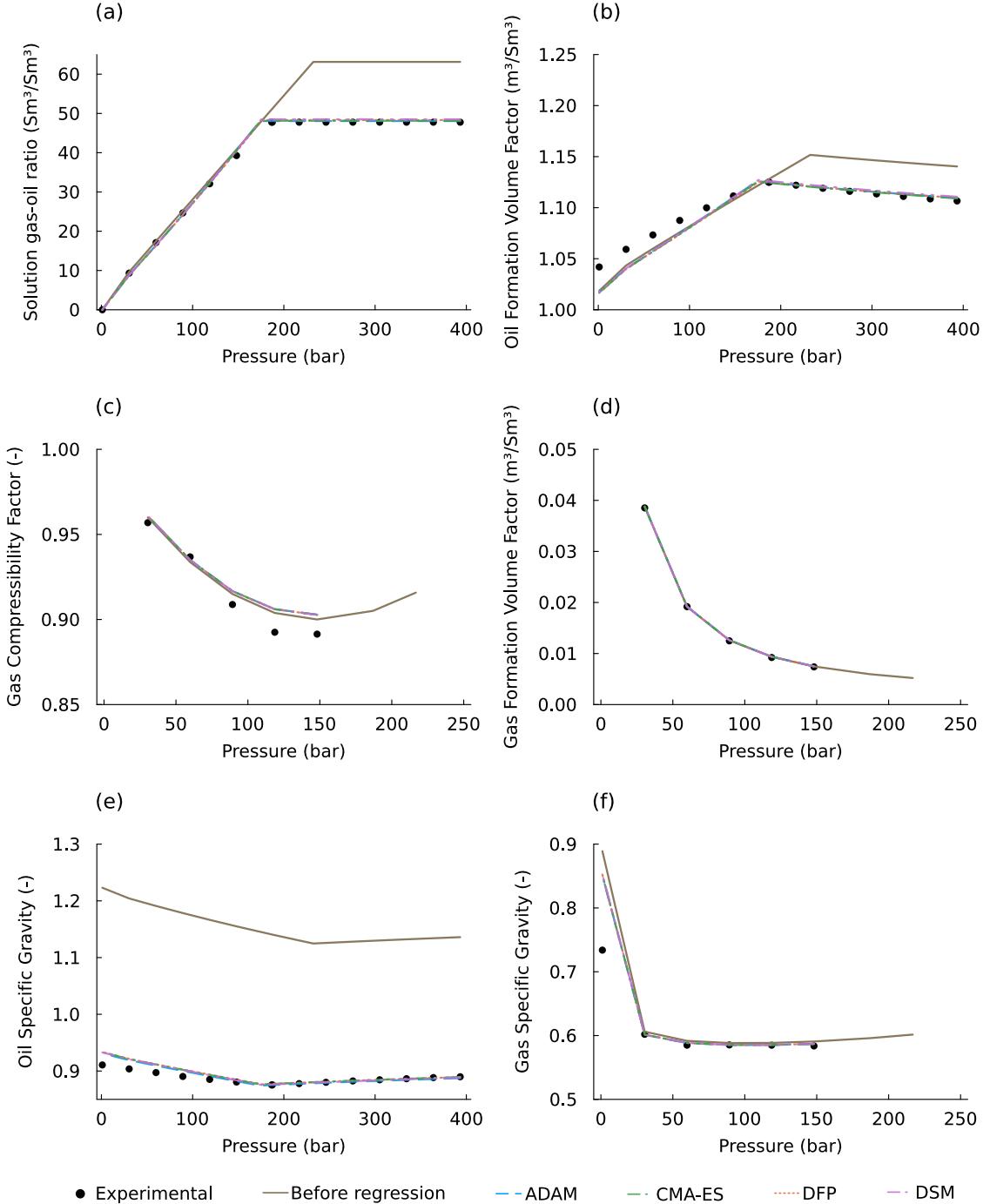


Figure 4.6: Fluid B EoS predictions and experimental data at 76°C for (a) solution gas-oil ratio, (b) oil formation volume factor, (c) gas compressibility factor, (d) gas formation volume factor, (e) oil specific gravity, and (f) gas specific gravity. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

Table 4.4: EoS predictions and experimental data for the saturation pressure and the API gravity of the residual oil from the differential liberation at 104°C and 76°C for fluids A and B, respectively. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57].

Property	Measurement	Original	ADAM	DFP	CMA-ES	DSM
Fluid A						
Saturation pressure (bar)	181.7	210.1	194.3	195.3	193.4	189.8
Residual oil API	35.1	39.4	36.7	36.6	36.6	36.5
Fluid B						
Saturation pressure (bar)	186.7	232.3	173.9	176.0	174.4	175.4
Residual oil API	17.5	17.9	17.9	17.6	17.7	17.6

data, to extrapolate from such regimes, i.e., to represent the unseen data.

Figure 4.7 depicts fluids A and B’s phase envelopes from the EoS models tuned with the four alternative optimization algorithms. Even though the phase equilibria behavior and the critical points of the regressed fluids are distinct, the lack of additional experimental data precludes any conclusion about the superiority of one model over the others. The significant discrepancy between the predicted envelopes may have severe consequences both during the design and operation of an oil and gas production system, since the tuned fluid EoS feeds into flow simulations for the reservoir and for the wellbore and pipeline.

Results validation

To validate the accuracy of the results obtained with our equation of state implementation, a comparison was made against results obtained with the commercial simulator **PVTsim**. Specifically, the results shown in Figures 4.5, 4.6, and 4.7 were reproduced in **PVTsim**. The comparison is shown in Figures 4.8-4.14. The comparison of the results obtained from both approaches illustrates that our EoS implementation is consistent with **PVTsim**, thus validating the accuracy of the results presented here.

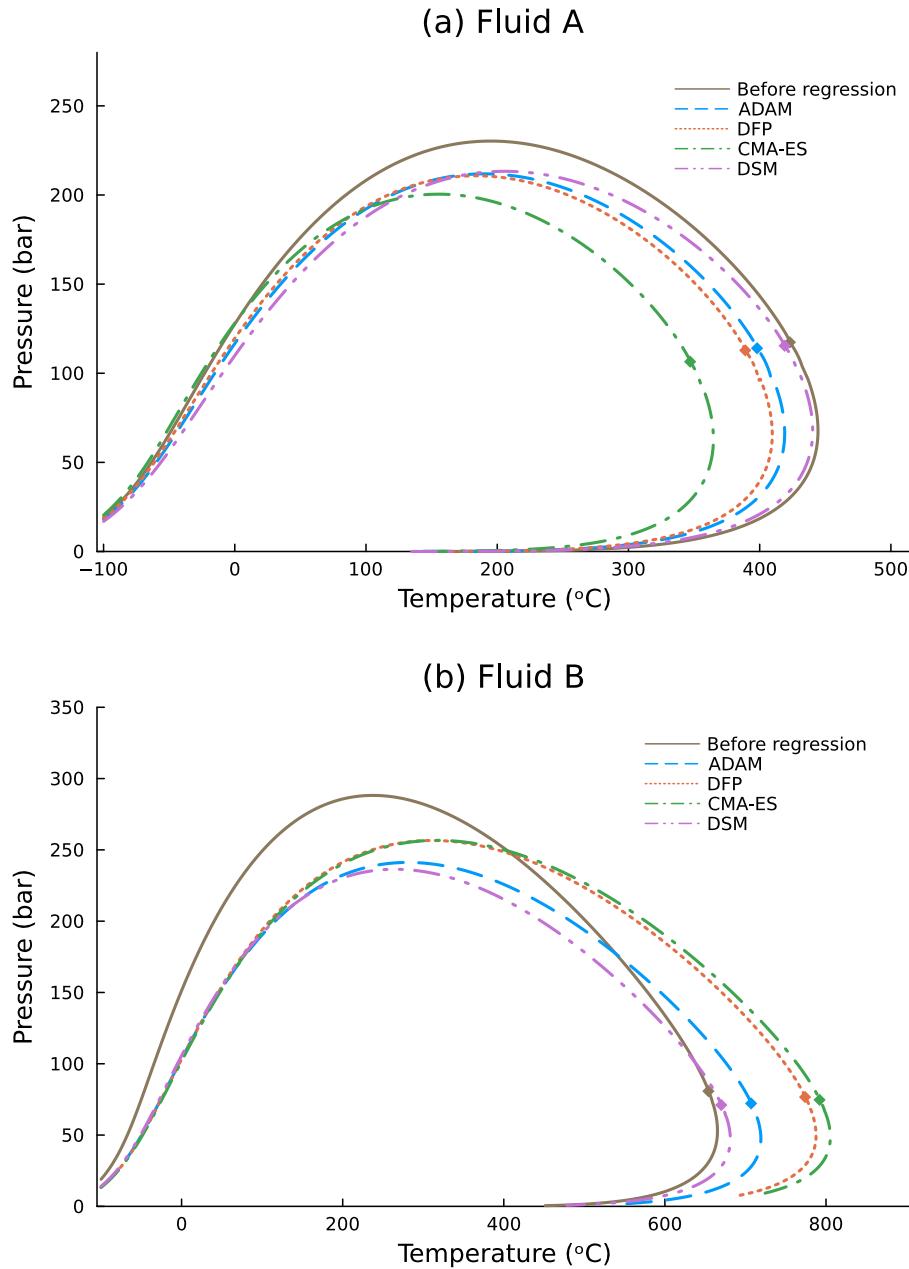


Figure 4.7: Phase envelopes (vapor liquid equilibrium) for the EoS models before and after regression of (a) fluid A and (b) fluid B. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and direct search method (DSM) [57]. The markers indicate the critical point.

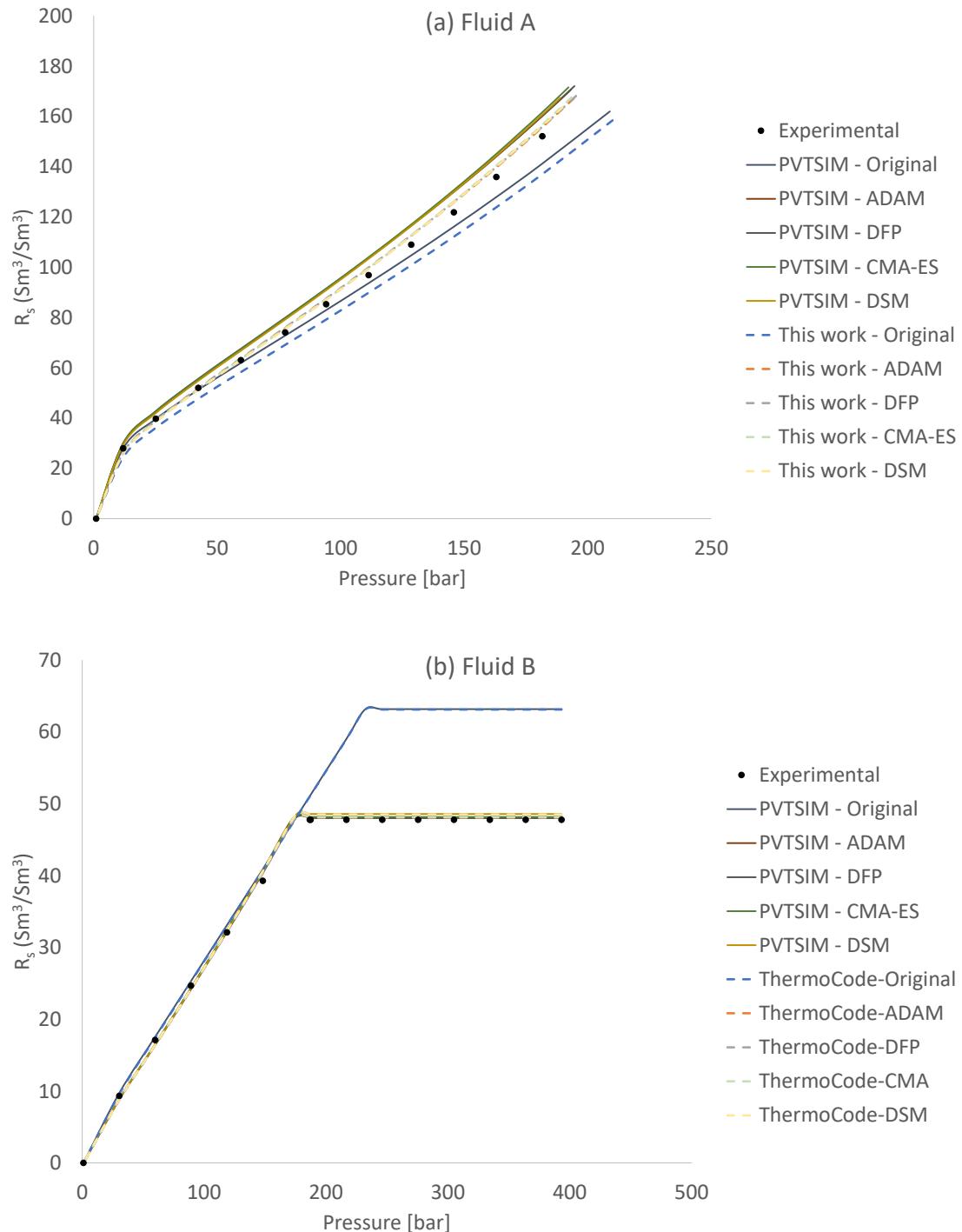


Figure 4.8: Comparison between our EoS implementation and PVTsim predictions for gas-oil solubility ratio of (a) Fluid A and (b) Fluid B.

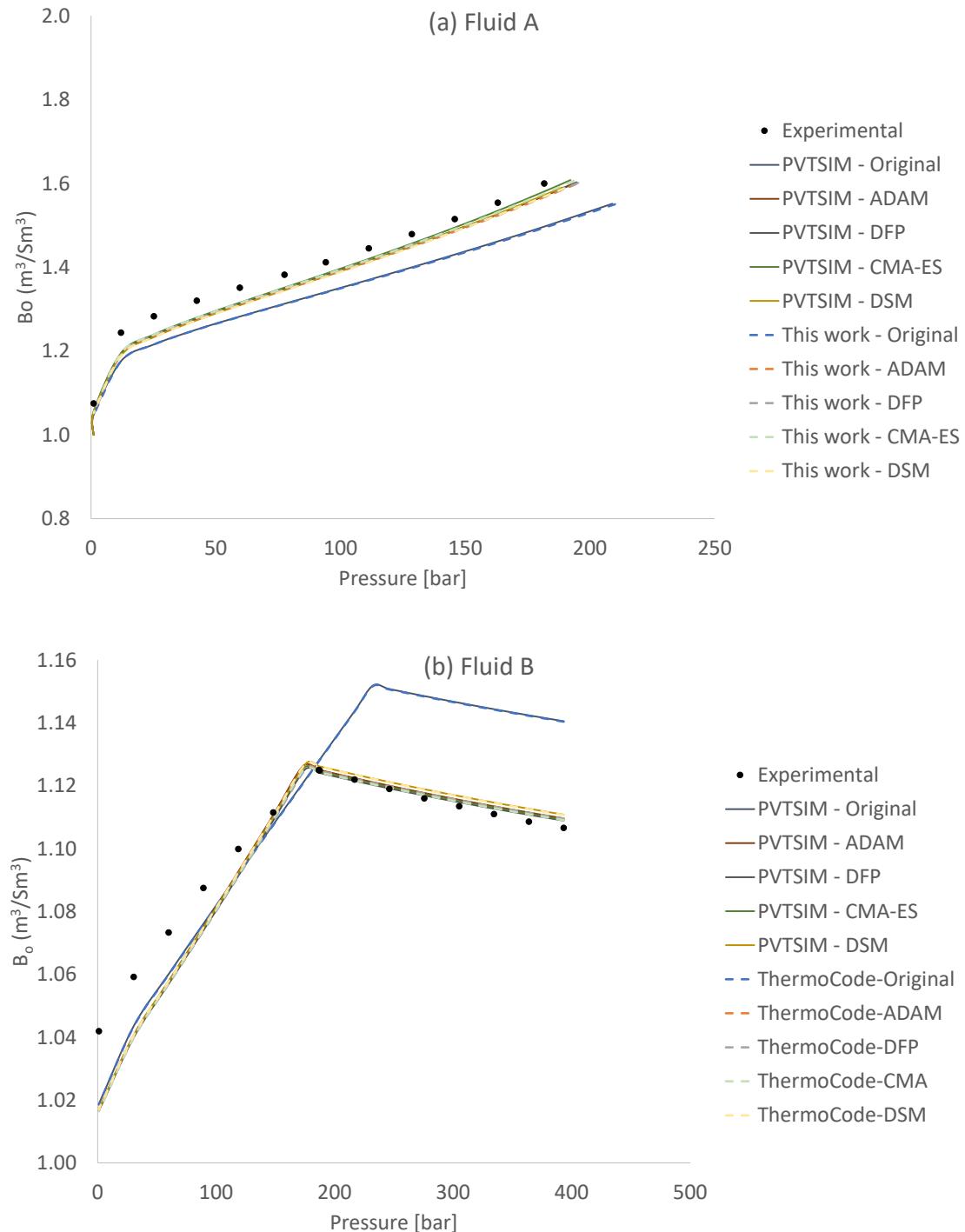


Figure 4.9: Comparison between our EoS implementation and PVTsim predictions for oil formation volume factor of (a) Fluid A and (b) Fluid B.

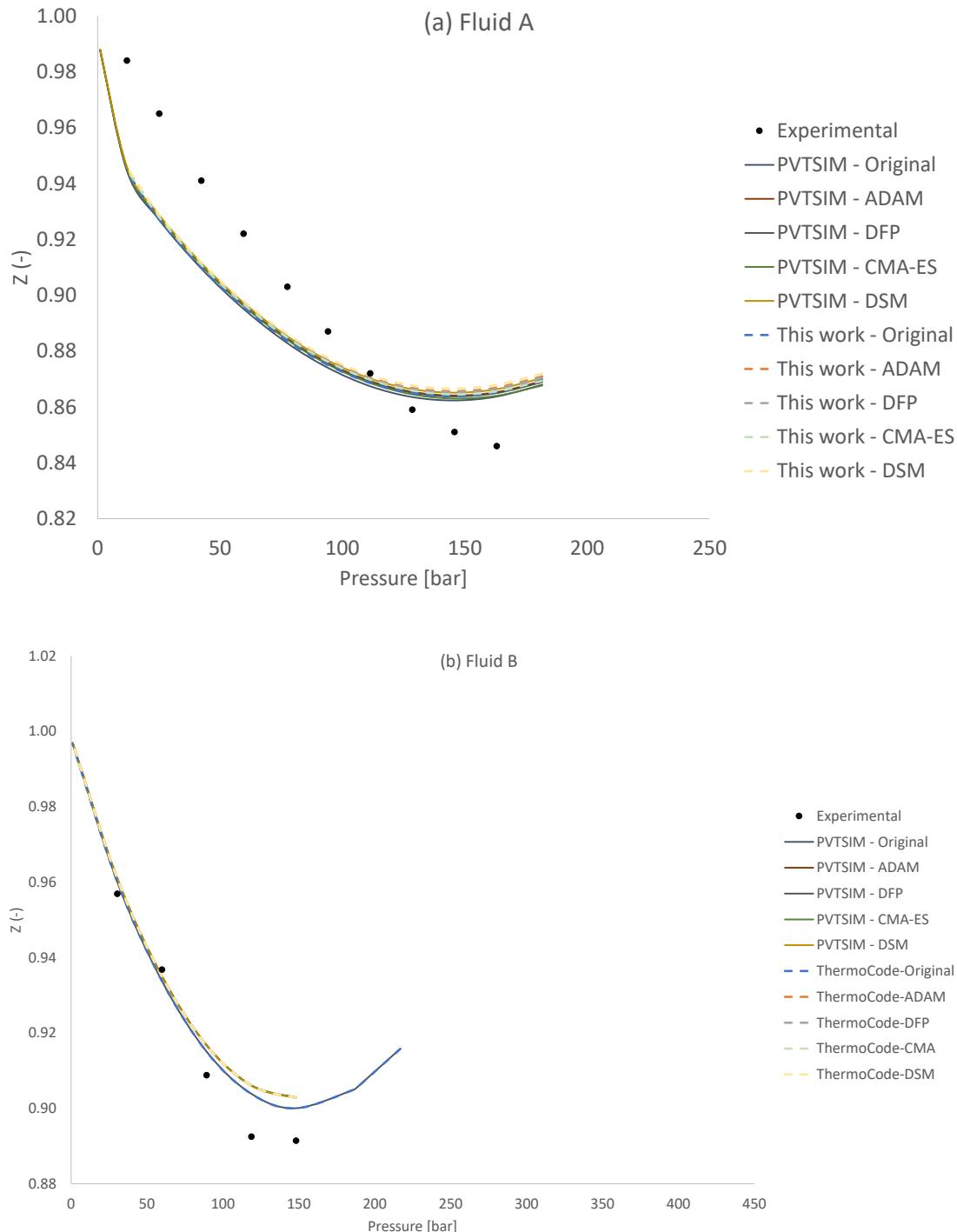


Figure 4.10: Comparison between our EoS implementation and **PVTsim** predictions for gas compressibility factor of (a) Fluid A and (b) Fluid B.

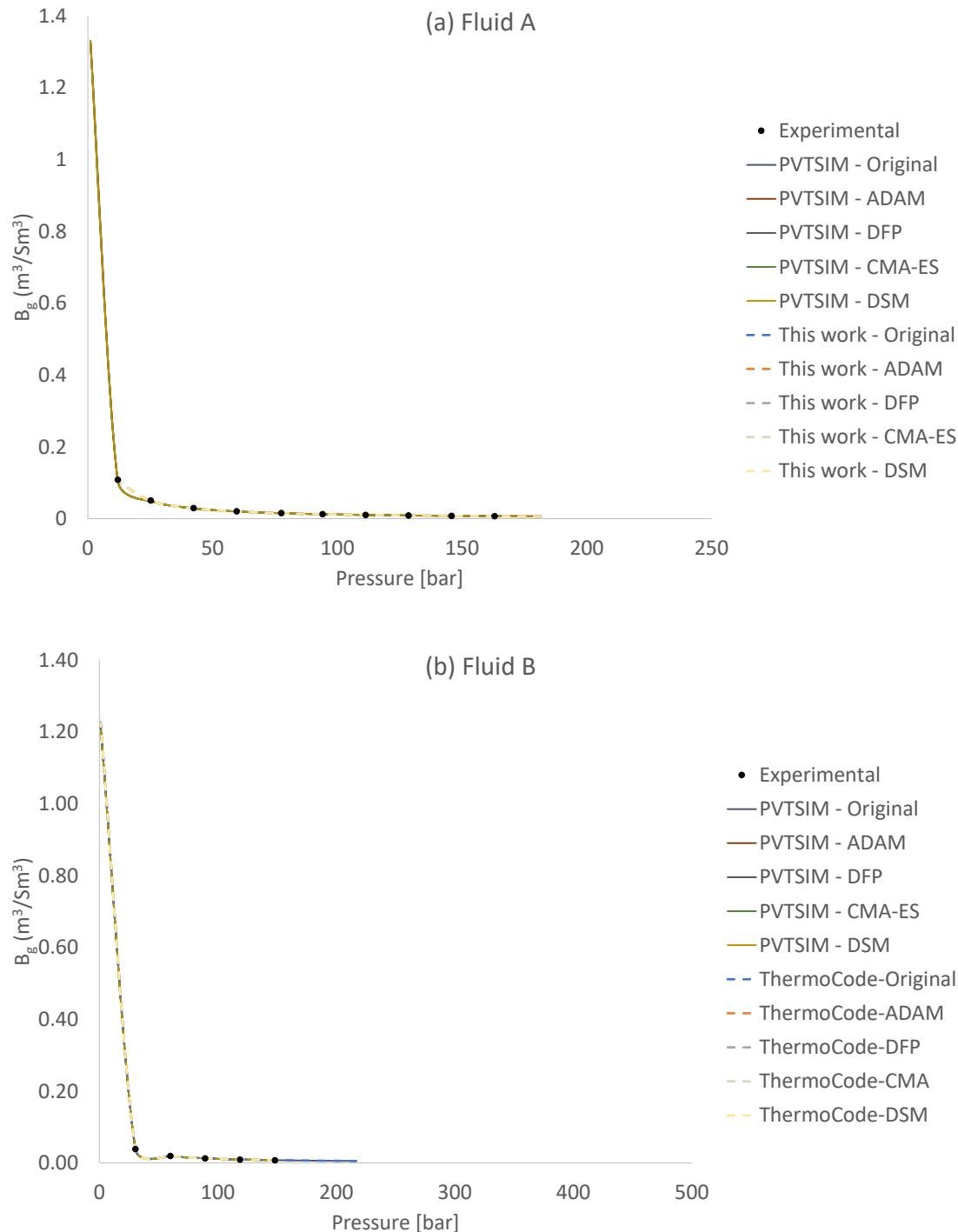


Figure 4.11: Comparison between our EoS implementation and **PVTsim** predictions for gas formation volume factor of (a) Fluid A and (b) Fluid B.

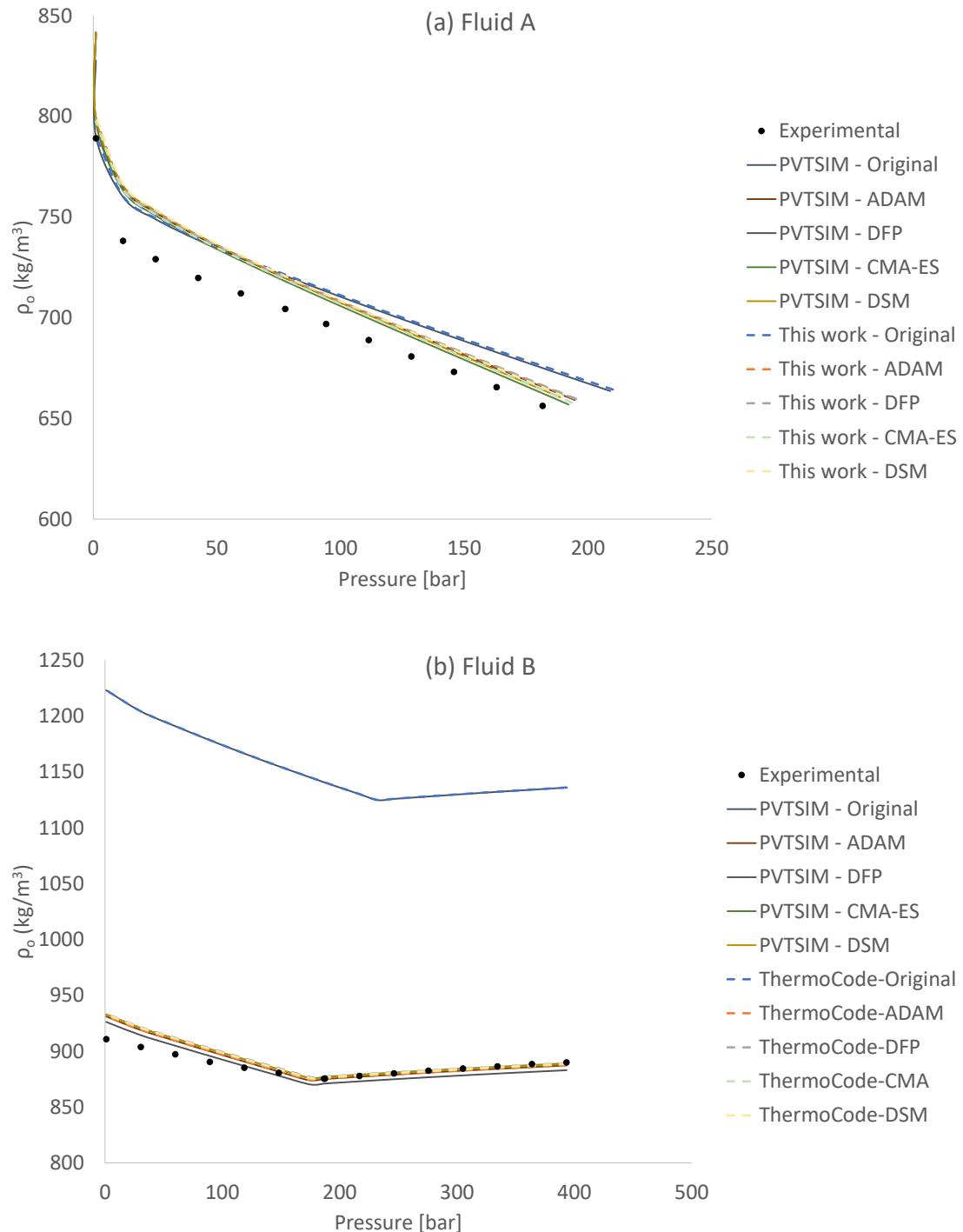


Figure 4.12: Comparison between our EoS implementation and PVTsim predictions for oil density of (a) Fluid A and (b) Fluid B.

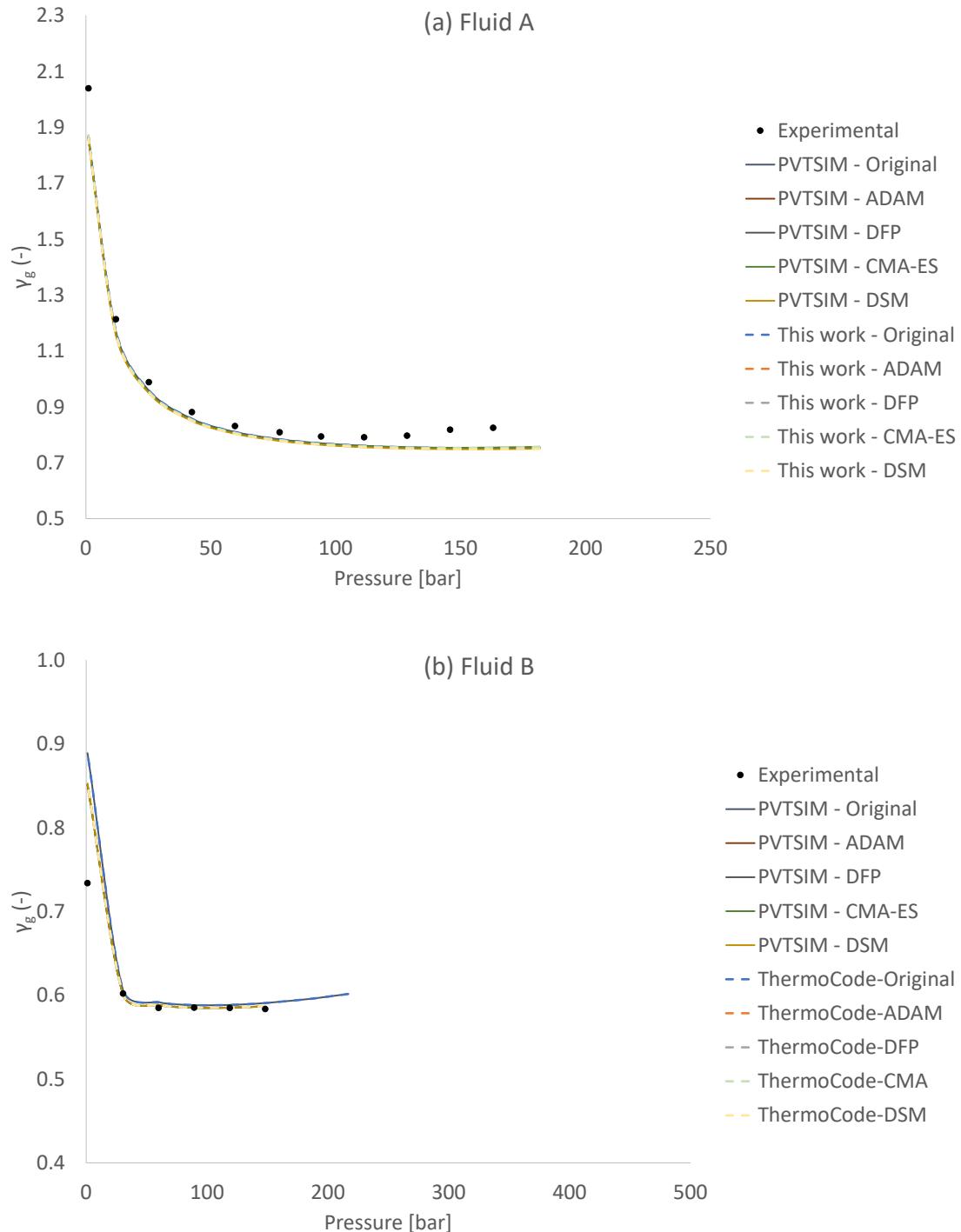


Figure 4.13: Comparison between our EoS implementation and PVTsim predictions for gas specific gravity of (a) Fluid A and (b) Fluid B.

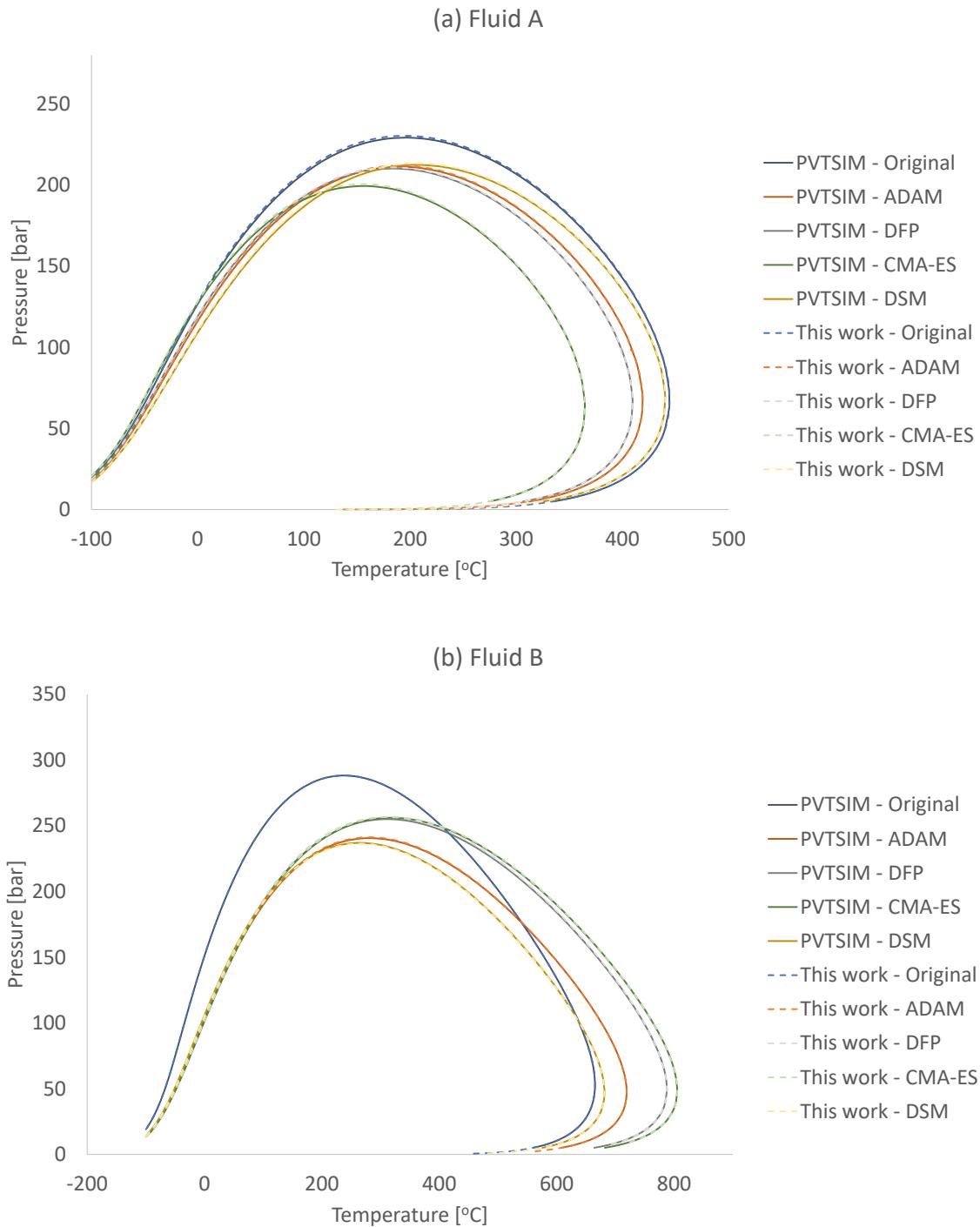


Figure 4.14: Comparison between our EoS implementation and **PVTsim** predictions for phase envelopes of (a) Fluid A and (b) Fluid B.

4.3.2 Initial guess

To ameliorate the EoS regression dependency on the optimization algorithm, we investigate a two-stage EoS-tuning procedure. The optimal values of the decision variables obtained with ADAM, CMA-ES, or DSM serve as the initial guess for the subsequent DFP minimization. The DFP is selected as the second stage optimization for being a second-order minimization strategy. We note that the commercial simulator *WinProp* v.2017 also employs a second-order algorithm, namely Agarwal et al.'s procedure [3].

The phase envelopes for thus refined EoS are plotted in Figure 4.15 for fluids A and B, demonstrating the sensitivity of DFP minimization to an initial guess. The phase envelopes optimized using the initial guesses from ADAM, CMA-ES, and DSM differ significantly. Table 4.5 reports the corresponding optima design variables. Different optimized values are reached by the same optimization algorithm with different initial guesses. This proves that the EoS regression may also depend on the design variables initialization besides of the optimization algorithm.

Table 4.5: EoS-regression results obtained via the two-step regression in which the optimal values of the decision variables obtained with ADAM [64], CMA-ES [55], or DSM [57] serve as the initial guess for the subsequent DFP minimization [33, 40].

Property	Initial guess from		
	ADAM	CMA-ES	DSM
Fluid A			
Critical pressure, $P_{\text{cr},C_{\text{plus}}}$ (bar)	17.64	17.73	17.66
Critical temperature, $T_{\text{cr},C_{\text{plus}}}$ (K)	746.46	686.75	771.37
Acentric factor, $\omega_{C_{\text{plus}}}$	0.6783	0.8609	0.5966
Molecular weight, $M_{C_{\text{plus}}}$	210	209	213
Volume shift, $c_{C_{\text{plus}}}$ (cm^3/mol)	45.3	22.1	51.4
Fluid B			
Critical pressure, $P_{\text{cr},C_{\text{plus}}}$ (bar)	10.90	10.49	11.04
Critical temperature, $T_{\text{cr},C_{\text{plus}}}$ (K)	1042.20	1131.17	1003.53
Acentric factor, $\omega_{C_{\text{plus}}}$	1.0933	1.0516	1.1228
Molecular weight, $M_{C_{\text{plus}}}$	504	504	504
Volume shift, $c_{C_{\text{plus}}}$ (cm^3/mol)	121.1	200.6	90.1

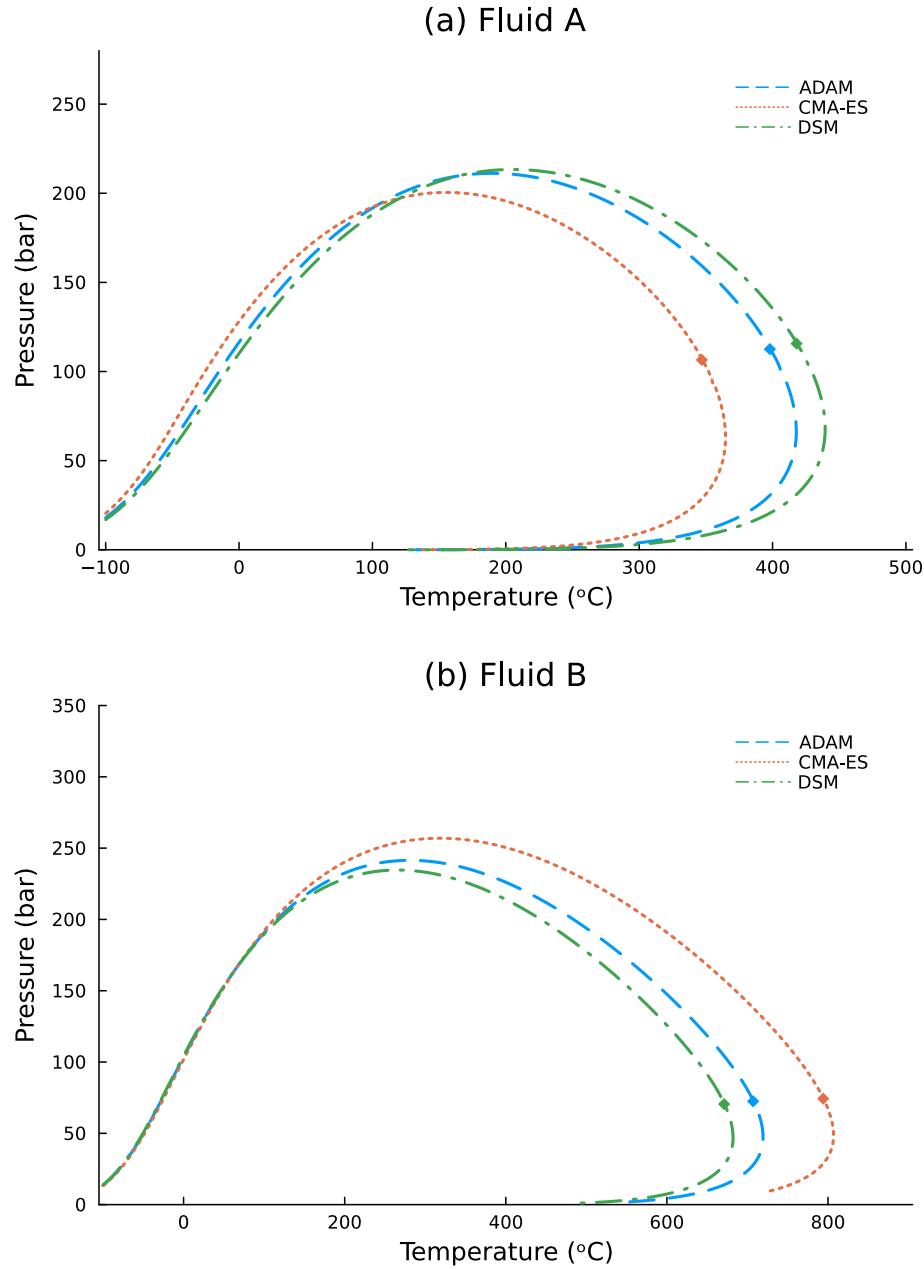


Figure 4.15: Phase envelopes (vapor liquid equilibrium) obtained for (a) fluid A and (b) fluid B via the two-step EoS regression in which the optimal values of the decision variables obtained with ADAM [64], CMA-ES [55], or DSM [57] serve as the initial guess for the subsequent GD2O minimization [3]. The markers indicate the critical point.

4.4 Discussion

The EoS models obtained using four different optimization algorithms yield significantly different predictions of phase envelopes, and there is no clear evidence that any single optimization algorithm outperforms the others during the EoS regression with generic data. The EoS predictions impact multiphase flow simulations, giving rise to predictive uncertainty in simulations of, e.g., oil and gas flow in the reservoir, wellbore, and pipelines. Since an erroneous EoS might adversely impact economic feasibility studies, cause severe flow assurance problems, and affect equipment integrity, reliance on a single optimization algorithm and a single initial guess during the EoS regression should be avoided. Yet, it is the strategy implemented in some commercial thermodynamic simulators.

Given the variability in the optimization results for a given selection of the design variables, comparative studies of different sets of the design variables and optimization procedures might have to be revisited. This is especially so since many of them, e.g., those discussed in the Introduction, do not even mention the optimization algorithm used.

Since both the initial guess and the optimization algorithm significantly impact the EoS-regression results, predictions of the regressed EoS models are fundamentally uncertain. This predictive uncertainty should be quantified, e.g., by treating the design variables probabilistically.

Probability distributions of the fraction properties, e.g., C_{plus} , can be characterized from a sample of regressed values considering different optimization algorithms and initial values. We demonstrate this approach for the design variables ($P_{\text{cr},C_{\text{plus}}}^*$, $T_{\text{cr},C_{\text{plus}}}^*$, $\omega_{C_{\text{plus}}}^*$, $M_{C_{\text{plus}}}^*$, and $c_{C_{\text{plus}}}^*$) of Fluid B, for which the optimized values are alternatively obtained by the four optimization algorithms (Section 4.2.2) and by the two-stage EoS-tuning procedure (Section 4.3.2) in which the optimized values from one algorithm are used as initial guesses for the other algorithms. From these samples containing 16 points each, we fit probability distributions to the C_{plus} properties in Figure 4.16.

The C_{plus} fraction parameters are potentially correlated, implying that a joint

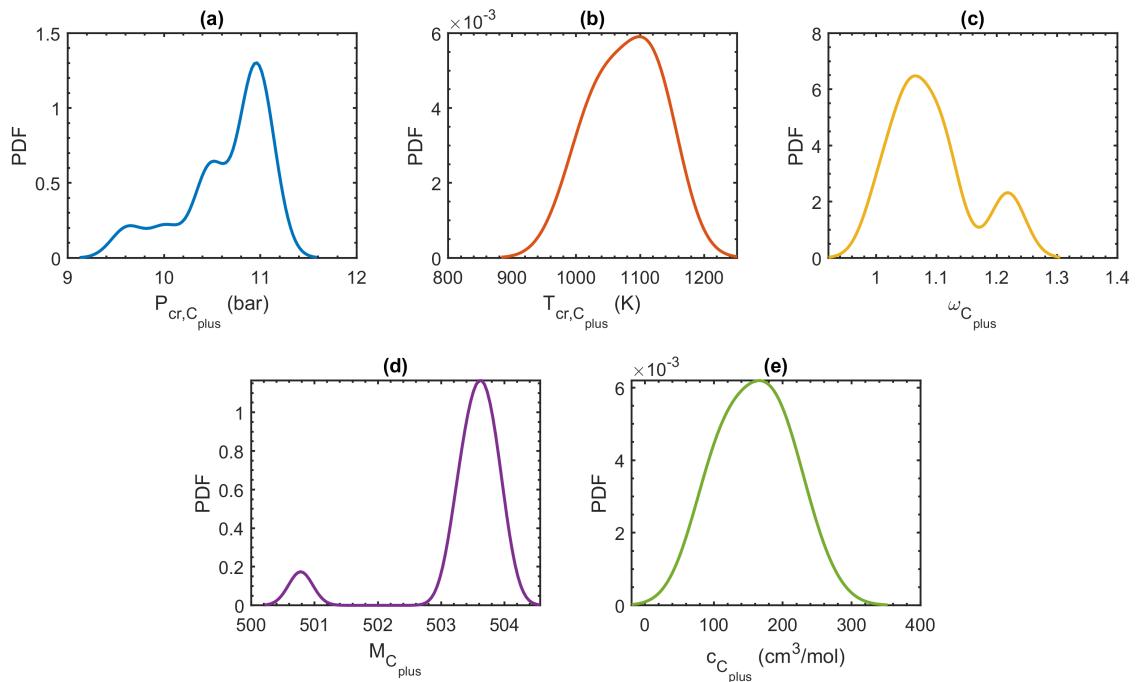


Figure 4.16: Fitted probability density function (PDF) for the C_{plus} fraction properties: (a) critical pressure, (b) critical temperature, (c) acentric factor, (d) molecular weight, and (e) volume shift parameter.

distribution would represent their uncertainty more accurately. However, undertaking such an analysis would require a more comprehensive examination and possibly additional data points, which we plan to investigate in future research.

One way to quantify the subsequent uncertainty associated with the EoS regression is through the use of Monte Carlo simulations. In these simulations, a large number of random realizations of the design variables are generated and used as inputs to the regression model. The resulting outputs can then be used to estimate the distribution of the predicted fluid properties, as well as their associated confidence intervals. This type of analysis provides a more comprehensive understanding of the uncertainty associated with the EoS regression results and can be used to guide future research efforts aimed at reducing the uncertainty of the EoS predictions in the context of partially characterized hydrocarbon mixtures. Uncertainty in the fluid-model characterization propagates through a modeling workflow, affecting the prediction uncertainty of the pipe [12, 45] and porous media flow [138, 139] models. Quantifying the EoS prediction uncertainty and its impact on the overall flow predictions is also an area of future work.

4.5 Conclusions

EoS regression on experimental data is a vital component of complex fluids modeling. While previous studies have focused mainly on the influence of EoS regression procedures and regression variables, we investigated the robustness of this nonlinear regression to the choices of an optimization algorithm and an initial guess used for its initialization. The four alternative optimization techniques considered are the adaptive moment estimation (ADAM) [64], the Davidon-Fletcher-Powell method (DFP) [33, 40], the covariance matrix adaptation evolution strategy (CMA-ES) [55], and the direct search method (DSM) [57]. These were deployed to fit the Peng-Robinson EoS [105] to experimental data for two hydrocarbon reservoir fluids. Our study leads to the following major conclusions.

- While the four optimization algorithms yield comparable errors in the tuned EoS representation of the experimental data, they lead to different EoS predictions of

the thermophysical behavior in the unsampled PVT region. This demonstrates that the choice of an optimization algorithm plays a key role in the EoS tuning.

- EoS regression is sensitive to the choice of an initial guess used to initiate the optimization procedure. This conclusion holds even for initial guesses obtained as solutions of the minimization problem obtained with different optimization algorithms.
- Our findings suggest the necessity of reporting C_{plus} fraction parameters as distributions rather than unique (expected) values. These distributions should be obtained from the optimized design variables considering different optimization algorithms and initial guesses.

Future studies could explore variations to the EoS regression process. Incorporating the C_{plus} fraction expansion and lumping procedures would introduce more design variables to the optimization problem. Including additional experimental data and assigning different weights to the experimental data would change the objective function and, therefore, the optimization process. Future studies could also investigate whether specific fluid compositions are more susceptible to disparate local minima during the EoS regression. A key aspect of future research is to conduct a comprehensive analysis of the tuned EoS prediction uncertainty and assessing its influence on the uncertainty in flow simulation results.

Chapter 5

Uncertainty in empirical fluid modeling

The content of this section follows closely the publication:

Lívia Paiva Fulchignoni, Christiano Garcia da Silva Santim, and Daniel M. Tarakovsky. Probabilistic forecasting of cumulative production of reservoir fluid with uncertain properties. *Geoenergy Science and Engineering*, 227:211819, 2023.

This chapter discusses the predictive uncertainty of simulated production flow rates arising from uncertainties in fluid characterization. We introduce a comprehensive framework for quantifying the predictive uncertainty of multiphase pipe-flow models, accounting for correlated random inputs. Through a case study, we assess the uncertainty of cumulative production for an oil reservoir with uncertain fluid properties during the exploration phase. Furthermore, we conduct a Global Sensitivity Analysis using Sobol's indices to pinpoint the model inputs that contribute most significantly to the predictive uncertainty.

5.1 Introduction

Energy companies make investment decisions based on technical and economic viability studies of exploration fields. The financial evaluation of revenues is mainly dictated by the estimated production, i.e., the oil and gas (O&G) flow rates that a reservoir can provide through its lifetime. This process is informed by simulations of multiphase flow in the reservoir, wellbore and pipelines. For this reason, such simulators are an essential tool during the design phase of a production system. Because many model parameters must be specified to run a multiphase-flow simulator, and because the values of these parameters are inherently uncertain, it is wise to evaluate production flow rates under a probabilistic approach and to quantify the uncertainty associated with their predicted values. This strategy renders the projected revenue a random variable, whose value one can calculate for the required degree of certainty, thus enabling one to make investment decisions based on risk and return.

Uncertainty quantification (UQ) gained popularity in the O&G industry in the 1960s [120] and has been evolving ever since, although at different rates among different disciplines (drilling, reservoir, production, operation, etc.) [20]. For reservoir simulations, UQ is an established practice [75, 139]. In fact, it is standard in the industry to report O&G reserve estimations under three categories (proved, probable, and possible), according to their likelihood (probability). Uncertainty in predictions of the accumulation volume is ideally tracked over time from exploration through discovery, development, and production [107]. In contrast, UQ for multiphase-flow models used in flow assurance has not yet been thoroughly investigated [65].

Section 5.1.1 discusses the sources of uncertainty in these flow models and establishes uncertain parameters (reservoir fluid properties) that are considered in this work. Section 5.1.2 introduces the Monte Carlo simulation (MCS) technique, reviews the literature for applications in multiphase pipe-flow models, and specifies this paper's goals and contributions to the topic. Section 5.2 presents the UQ methodology applied to model predictions of cumulative production, including a thorough description of the production system used as a case study. Section 5.3 discusses the results obtained in this study. Section 5.4 summarizes major conclusions.

5.1.1 Sources of uncertainty in flow simulations

Uncertainty in predictions of multiphase-flow models, which represent production from a subsurface reservoir to the land-surface facilities, arises from multiple sources. *Structural or model uncertainty* is due to inevitable approximations of “reality” introduced by a model. Various models (e.g., the drift-flux [143, 131, 59] and two-fluid [19] models), which encode the laws of conservation of mass, momentum and energy, represent the complex physics with different degrees of fidelity. These equations require empirical constitutive relations of varying validity and generality. Such relations are usually parameterized via small-scale experiments, whereas the flow models are used on a much larger scale. These experiments are commonly carried out on small acrylic pipes, with working fluids other than O&G.

Compositional and “black-oil” fluid models are a representative example of constitutive relations. They are used to predict the thermophysical behavior of a reservoir fluid by relating the properties of oil, gas, and water phases to pressure and temperature. Compositional fluid models are grounded in thermodynamics, but have a number of fitting parameters whose tuning is subjective and may yield different fluid characterizations.¹ In contrast, Black Oil fluid models are empirical but require only a few inputs (i.e., density of each phase at standard conditions, watercut, and reservoir gas-oil ratio) to characterize the fluid mixture across a wide range of pressures and temperatures. (Some black-oil formulations might include additional inputs, e.g., the CO₂ mole fraction of surface gas [41, 48].) The largely subjective choice of the constitutive fluid model yields predictions of the solution gas-oil ratio, oil formation volume factors, gas compressibility factor, oil and gas viscosities, and other derived quantities at each pressure and temperature pair in the flow simulation. This choice can significantly impact the predictions of a flow simulator; for example, different choices of the solution gas-oil ratio and gas-liquid drift models significantly influence the calculated liquid holdup profile along the production line and temperature near the outlet [111].

Other contributions to the structural uncertainty include the frequent use of a

¹For example, the choice of characterization of the heaviest hydrocarbon fraction, and splitting and lumping procedures, drastically affect the final fluid model.

well/reservoir coupling model as boundary condition [52]. In these cases, the reservoir behavior is described through an inflow performance relationship (IPR) curve that relates the reservoir and well bottom hole pressures to production flow rates. Several coupling models have been proposed [6], each of which require different uncertain inputs to characterize the reservoir behavior.

Once the governing equations and accompanying constitutive relations have been selected, they have to be parameterized, i.e., the values of the model parameters have to be specified, usually by fitting the model predictions to experimental data. Figure 5.1 collates the parameters (inputs) required by a steady-state black-oil flow model, which we use in our numerical experiments because of its simplicity and low computational cost. Data from the well perforation (drill-bit diameter, deviation profile, casing diameter and thickness, etc.), completion (tubing diameter, thickness, and roughness, gas lift, electric submersible pumps, etc.), flow lines and riser (layout, diameter, thickness, roughness and thermal insulation, etc.) and subsea equipment (separator, pump, etc.) are needed to describe fluid flow and heat transfer. Reservoir data (pressure, temperature, and productivity index) characterizes the source boundary condition. Finally, fluid data—oil, gas, and water densities, gas-oil ratio, and watercut—are inputs to the fluid models. Additional information, such as the surface boundary conditions, is necessary to run a simulation. The need to specify numerical values for all of these inputs introduces *parametric uncertainty* into the modeling process.

Model parameters differ by both their degree of uncertainty and their impact on the model’s prediction uncertainty. For instance, frequent measurement of reservoir pressure through pressure buildup tests is expensive; infrequent measurements, the industry standard, increase the uncertainty in this important boundary condition. Conversely, the pipeline diameter is directly measurable and, thus, uncertainty in its value is comparatively low, arising primarily from measurement errors. Within the probabilistic framework, this statement is equivalent to saying that the normalized standard deviation of reservoir pressure, σ_p , exceeds that of pipe diameter, σ_d . Model nonlinearity implies that the condition $\sigma_p \gg \sigma_d$ does not automatically mean that uncertainty in reservoir pressure has larger impact on prediction uncertainty of the

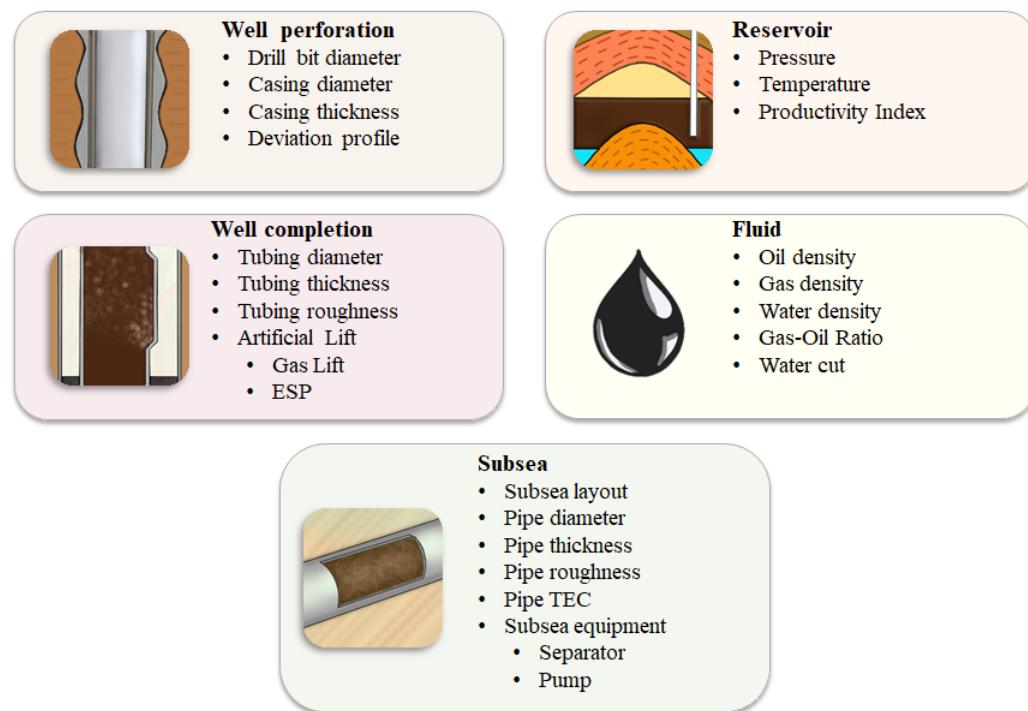


Figure 5.1: Physical properties that provide input for a steady-state black-oil flow simulation.

pipe-flow simulations than uncertainty in pipe diameter does. Addressing the latter issue falls under the purview of global sensitivity analysis, which is discussed in Section 5.2.3.

Our study focuses on parametric uncertainty associated with fluid properties, which is prevalent during the exploration phase, when little site-specific information is available. Fluid properties data are collected from a few exploration/appraisal wells. In the absence of such wells, the reservoir fluid is modeled based on information from a basin model whose parameterization relies on data from analogous geological areas. Even when samples are available, their analysis is carried out in laboratory conditions that can be quite different from the reservoir conditions [94]. For these reasons, we treat as uncertain the following fluid-model parameters: oil and gas specific gravities at standard conditions (γ_o and γ_g , respectively), reservoir fluid gas-oil ratio, and watercut. We assume the water density at standard condition to be known with certainty, even though the water properties depend on water salinity, which can be uncertain. Furthermore, density and viscosity of each phase throughout the flow (at varying pressure and temperature conditions) are computed through the black-oil model, whose uncertainties are out of the scope of this study. Our primary goal is to quantify the influence of parametric uncertainty in the fluid model on prediction uncertainty of the cumulative production.

5.1.2 Uncertainty quantification for flow simulations

While many computationally efficient alternatives to Monte Carlo simulations (MCS) have been used to evaluate prediction uncertainty of models of multiphase flow in heterogeneous porous media [108, 123, 139], comparative UQ efforts for multiphase flow in O&G pipes are scarce and mostly limited to MCS. Examples of the latter include quantification of uncertainty in predictions of a two-phase pipe-flow simulator (pressure drop and liquid holdup in an experimental pipe) with uncertain/random inputs (flow rates, viscosities, and densities of both phases; diameter, inclination angle, and length of the pipe; surface tension; and parameters of a non-Newtonian fluid model) [99].

In real production scenarios, MCS were deployed to quantify uncertainty in model predictions of time-varying oil-flow rates from multiple producer wells during the optimization of a field layout design, with the initial flow rates and the reservoir's decline rate acting as uncertain inputs [110]. They were also used to evaluate the impact of the watercut uncertainty on model predictions of the oil production flow rate [79], and to analyze the influence of uncertainty in the values of three input parameters (outlet pressure, ambient temperature, and wall roughness) on predictions of the pressure drop and liquid holdup [65].

We use an accelerated version of MCS, which relies on Latin hypercube sampling [124], to investigate the impact of parametric uncertainty in the fluid model on model predictions of the time-dependent production flow rate during the exploration phase, in the context of flow assurance. The analysis is performed for an offshore production scenario, based on a real well located in Campos Basin, Brazil. A result of our analysis is the probabilistic forecast of cumulative oil production and, consequently, the project's revenue.

5.1.3 Global sensitivity analysis

Global sensitivity analysis (GSA) is a distinct and complementary facet of uncertainty management and risk assessment [28]. It seeks to rank the random inputs by their relative contribution to uncertainty in predictions of a quantity of interest (QoI). Depending on whether prediction uncertainty is represented in terms of the QoI's variance or full distribution, GSA can be classified as variance- or distribution-based, respectively; the latter is applicable to a wider range of inputs, while the former is easier to compute [29]. Variance-based GSA ranks the input parameters by their Sobol' indices, whose definition derives from the multivariate analysis of variance (ANOVA) [135]. These indices can be computed via either MCS [99] or polynomial chaos expansions [121].

Undergirding the variance-based GSA is the requirement that random inputs are mutually uncorrelated. This is a questionable assumption in the context of fluid modeling, where many of the model parameters are inter-related. To account for this

complication, we use the Rosenblatt transform to map the correlated model inputs onto a set of uncorrelated random variables [29], which are then used to perform a variance-based GSA. This enables us to identify the most influential parameters to the production flow rate output at different stages of production lifetime.

5.2 Methodology

5.2.1 Probabilistic characterization of uncertain fluid properties

We focus on uncertainty in four fluid properties $\mathbf{X} = (X_1, \dots, X_4)^\top \in \mathbb{R}^4$: the oil and gas specific gravities ($X_1 = \gamma_o$ and $X_2 = \gamma_g$), the gas-oil ratio ($X_3 = \text{GOR}$), and watercut ($X_4 = \text{WC}$). Statistical properties of these parameters, e.g., their means and variances, are estimated during the exploration phase, when little information about the reservoir field and its fluid is available. For the sake of generalization and interpretability of the results, and in the absence of evidence to the contrary, we use a multivariate Gaussian probability density function (PDF),

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{2\pi|\Sigma|^{1/2}} \exp \left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^\top \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right], \quad (5.1)$$

to characterize these random inputs. Here, the vector $\boldsymbol{\mu} \in \mathbb{R}^4$ comprises the means of the four fluid properties, and the positive-definite matrix $\Sigma \in \mathbb{R}^{4 \times 4}$ is composed of the covariances between these properties.

The mean values in the vector $\boldsymbol{\mu}$ are typically estimated by experts from their knowledge of similar geological areas. As an example, we use the mean oil specific gravity μ_{γ_o} , the mean gas specific gravity μ_{γ_g} , and the mean gas-oil ratio μ_{GOR} (with units Sm^3/Sm^3), which are the values measured in a PVT analysis of a reservoir fluid that is deemed representative. Following the treatment of Well 03 in [79], we assume the mean watercut $\mu_{\text{WC}}(t)$ (in percentage) to vary linearly with production time t (in months), increasing at a rate of 0.667%/month. In our experiments, we use the following numbers:

$$\boldsymbol{\mu}(t) = (\mu_{\gamma_o}, \mu_{\gamma_g}, \mu_{GOR}, \mu_{WC}(t))^\top = (0.944, 0.610, 163.0, 0.667t)^\top. \quad (5.2)$$

The covariance matrix $\boldsymbol{\Sigma}$ encapsulates the degree of uncertainty in the values of the model parameters and their correlations. Such uncertainty in the fluid description arises from potential heterogeneity of fluid properties across the field. It can vary greatly from field to field, as it is influenced by factors such as data acquisition, production experience, and geological complexity. In the exploration phase, field experience suggests the standard deviation of oil density and gas-oil ratio to be up to 2% and 20% of their mean values, respectively [75]. We assume gas density to be as uncertain as the oil density and the standard deviation of watercut to be 10% of its mean value, approximately as in [79]. The watercut is independent of the reservoir fluid composition. Our correlation analysis of 220 reservoir fluid samples from Brazilian oil fields yields the correlation matrix

$$\begin{aligned} \boldsymbol{\Sigma}(t) &= \begin{pmatrix} \sigma_{X_1}^2 & C_{X_1 X_2} & C_{X_1 X_3} & C_{X_1 X_4} \\ C_{X_2 X_1} & \sigma_{X_2}^2 & C_{X_2 X_3} & C_{X_2 X_4} \\ C_{X_3 X_1} & C_{X_3 X_2} & \sigma_{X_3}^2 & C_{X_3 X_4} \\ C_{X_4 X_1} & C_{X_4 X_2} & C_{X_4 X_3} & \sigma_{X_4}^2(t) \end{pmatrix} \\ &= \begin{pmatrix} 3.6 \cdot 10^{-4} & -9.4 \cdot 10^{-5} & -1.8 \cdot 10^{-1} & 0 \\ -9.4 \cdot 10^{-5} & 1.5 \cdot 10^{-4} & 2.7 \cdot 10^{-1} & 0 \\ -1.8 \cdot 10^{-1} & 2.7 \cdot 10^{-1} & 1.1 \cdot 10^3 & 0 \\ 0 & 0 & 0 & 4.5 \cdot 10^{-3} t^2 \end{pmatrix}. \end{aligned} \quad (5.3)$$

This analysis indicates the strong correlation between the inputs, especially between the gas specific gravity and the gas-oil ratio.

5.2.2 Monte Carlo simulations

MCS is an ensemble-based computation consisting of three steps. First, N realizations, $\mathbf{x}_1, \dots, \mathbf{x}_N$, of the random input vector \mathbf{X} are drawn from the joint PDF in (5.1). Second, for each realization of the inputs, \mathbf{x}_n ($n = 1, \dots, N$), the flow model is solved to obtain the corresponding realizations of the model output and QoIs. Third, the sample statistics of the QoIs, including their PDFs, are computed.

The root mean squared error (RMSE) of an MC estimate of the QoI's mean decays as $1/\sqrt{N}$ [85]. This slow convergence rate is due to the purely random sampling in which most of the samples come from the center of a PDF and relatively few sample from the PDF's tails. To accelerate convergence within a given tolerance, we deploy the Latin hypercube sampling, in which the domain of definition of PDF $f_{\mathbf{X}}(\mathbf{x})$ is subdivided into equal intervals ("strata") and random realizations \mathbf{x}_n are drawn from each interval.

5.2.3 Global sensitivity analysis

The application of variance-based GSA is limited to uncorrelated random inputs. We use the Rosenblatt transform [106] to map \mathbf{X} onto a random vector $\mathbf{U} = (U_1, \dots, U_4)^{\top} \in \mathbb{R}^4$ uniformly and independently distributed over the unit hypercube $[0, 1]^4$. Let $F_{X_i|X_1, \dots, X_{i-1}}(x_i|x_1, \dots, x_{i-1})$ be the conditional cumulative distribution function of X_i given X_1, \dots, X_{i-1} . The Rosenblatt transform of $\mathbf{X} \in \mathbb{R}^4$ is defined as $T(\mathbf{X}) = \mathbf{U}$ such that

$$u_1 = F_{X_1}(x_1), \quad (5.4)$$

$$u_2 = F_{X_2|X_1}(x_2|x_1), \quad (5.5)$$

$$u_3 = F_{X_3|X_1, X_2}(x_3|x_1, x_2), \quad (5.6)$$

$$u_4 = F_{X_4|X_1, X_2, X_3}(x_4|x_1, x_2, x_3). \quad (5.7)$$

Because the Rosenblatt transform is bijective, we define

$$g(\mathbf{U}) = h \circ T^{-1}(\mathbf{U}) = h(\mathbf{X}) = \text{QoI}.$$

In our example, QoI is the simulated oil flow rate at each time step. Thus, we perform the GSA on $g(\mathbf{U})$. The first-order, S_{X_i} , and total, T_{X_i} , Sobol's indices are thus defined as

$$S_{X_i} = \frac{\text{Var}[\mathbb{E}[g(\mathbf{U})|U_i]]}{\text{Var}[g(\mathbf{U})]} = \frac{\text{Var}[\mathbb{E}[\text{QoI}|X_i]]}{\text{Var}[\text{QoI}]} \quad (5.8)$$

$$T_{X_i} = \frac{\mathbb{E}[\text{Var}[g(\mathbf{U})|U_{\sim i}]]}{\text{Var}[g(\mathbf{U})]} = \frac{\mathbb{E}[\text{Var}[\text{QoI}|X_{\sim i}]]}{\text{Var}[\text{QoI}]} \quad (5.9)$$

where $U_{\sim i}$ denotes all parameters but U_i ; and $\mathbb{E}[\cdot]$ and $\text{Var}[\cdot]$ are the mean and variance operator, respectively. We rely on the R package `sensobol` [101] to compute S_{X_i} and T_{X_i} .

5.2.4 Case study

We demonstrate our approach on a simplified model of a real offshore production system located in Campos Basin, Brazil, denoted as “Well B” in [81]; the well is assumed to produce by natural lift. Figure 5.2 shows a schematic representation of the simplified production well in the multiphase flow simulator, along with some general characteristics. The flow simulation boundary conditions are the separator pressure of 16.6 bar; the linear IPR model is used for the reservoir, with the reservoir pressure of 324.1 bar and the productivity index of $53.0 \text{ m}^3/(\text{d.bar})$.

Table 5.1 fully describes the simplified production system represented in the multiphase flow simulation implemented in this work. For the fluid parameters treated as random variables, the mean values of their respective probability distributions at the beginning of the production life are reported. Table 5.2 shows the flow and fluid correlations considered in the simulation. Fig. 5.3 presents the simulated flow behavior at the beginning, middle and end of the production lifetime, considering the mean watercut values of 0%, 40%, and 80%, respectively. The pressure, temperature,

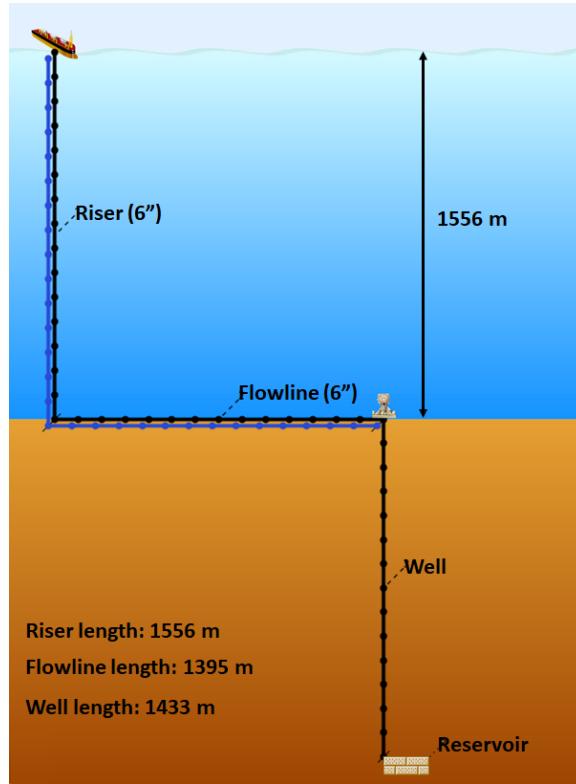


Figure 5.2: Schematic of a simplified production well implemented in the flow simulator.

and liquid holdup profiles along the wellbore and pipelines are reported.

Flow simulations are performed with the Petrobras in-house simulator MARLIM II[®] (Multiphase flow and ARtificial LIft Modeling), which outputs the production flow rates. The production lifetime is simulated in time steps of one month. The reservoir pressure and temperature are assumed to be constant throughout the production lifetime due to the waterflooding process, a secondary recovery technique.

The cumulative oil production is computed via numerical integration over time of the discrete flow rates. For that, we conduct additional MCS, in which flow rates serve as the inputs whose realizations are drawn from the flow-rate PDFs obtained at each time step.

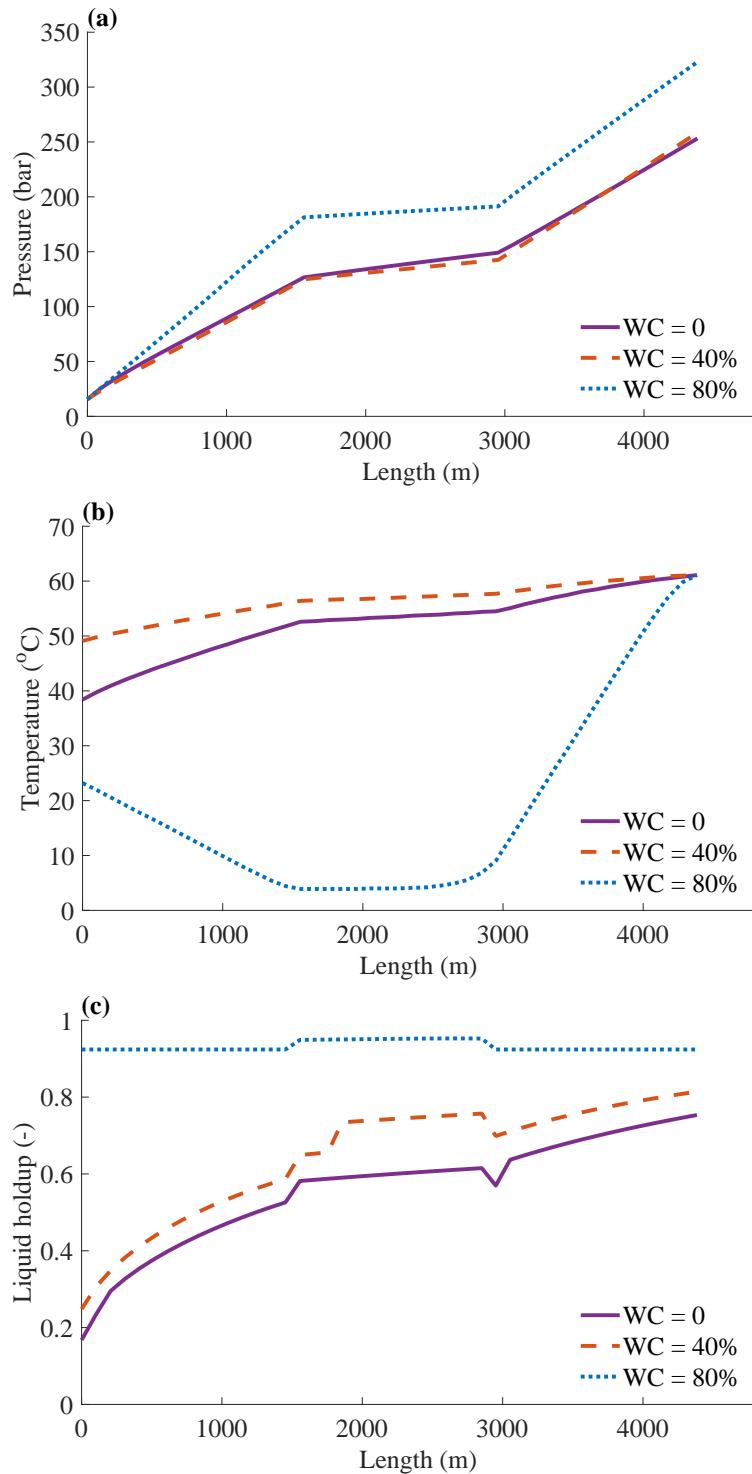


Figure 5.3: (a) Pressure, (b) temperature, and (c) liquid holdup behaviors at the steady-state regime in the wellbore and production line for watercut values of 0%, 40%, and 80%, representing the beginning, middle and end of the production lifetime, respectively.

Table 5.1: Parameters and their numerical values used in the multiphase flow simulations.

	Parameter name	Parameter value
Reservoir	Reservoir pressure	324.1 bar
	Reservoir temperature	61.1°C
	Productivity index	53.0 m ³ /(d.bar)
	Tubing length	1433 m
	Tubing internal diameter	5.92 in
Vertical well	Tubing thickness	0.0089 m
	Tubing roughness	0.000178 m
	Casing length	1433 m
	Casing internal diameter	9.66 in
	Casing thickness	0.0138 m
	Cement thickness	0.03 m
Horizontal flowline	Overall heat transfer coefficient	7.1 J/(s.K.m ²)
	Length	1395 m
	Internal diameter	6 in
	Thickness	0.0783 m
	Roughness	0.00061 m
	Overall heat transfer coefficient	5.2 J/(s.K.m ²)
Vertical riser	Length	1556 m
	Internal diameter	6 in
	Thickness	0.0488 m
	Roughness	0.00061 m
	Overall heat transfer coefficient	8.4 J/(s.K.m ²)
Ambient and outlet conditions	Soil temperature at reservoir	61.1°C
	Soil temperature at wellhead	3.9°C
	Soil temperature gradient	Linear
	Seawater temperature at seabed	3.9°C
	Seawater temperature at surface	25°C
	Seawater temperature gradient	Linear
	Seawater current velocity	0.1 m/s
Black-oil fluid	Pressure at the outlet	16.6 bar
	Stock-tank oil specific gravity (γ_o)	0.944
	Stock-tank gas specific gravity (γ_g)	0.610
	Water specific gravity (γ_w)	1.030
	Reservoir fluid gas-oil ratio (GOR)	163.0 Sm ³ /Sm ³
	Watercut (WC)	0%

Table 5.2: Flow and black-oil constitutive relations used in the multiphase-flow simulations.

	Constitutive relation
Black-oil fluid property	Gas-oil solubility ratio
	Oil formation volume factor
	Gas compressibility factor
	Undersaturated oil viscosity
	Live oil viscosity
	Dead oil viscosity
	Gas viscosity
	Water viscosity
	Emulsion relative viscosity
Flow	Standing [118]
	Standing [118]
	Dranchuk et al. [38]
	Vazquez and Beggs [128]
	Beggs and Robinson [18]
	Beggs and Robinson [18]
	Lee et al. [69]
	van Wingen [127]
	Woelflin tight emulsion [136] (inversion watercut of 90%)
	Beggs, Brill and Palmer [17, 86]

5.3 Results

5.3.1 Monte Carlo simulations

Figure 5.4 exhibits the results of our convergence study of MCS with Latin hypercube sampling, with the flow rate at the end of the well’s lifetime (when the uncertainty on the inputs is higher) playing the role of QoI. The sample mean, μ , and standard deviation, σ , of this QoI converge after $N \approx 5000$ MC realizations (Figure 5.4a). This result demonstrates the limited value of the QoI statistics obtained via MCS with $N = 100$ Latin hypercube samples, as done in [79] for a similar setting. For instance, the mean oil flow rate estimated with $N = 100$ MC realizations is $\mu = 218.4 \text{ Sm}^3/\text{d}$, which is 5% higher than $\mu = 207.7 \text{ Sm}^3/\text{d}$ estimated from $N = 10^5$ realizations. This highlights the importance of a convergence analysis of MCS, including the establishment of a convergence criteria that represents the specific goal of a study.

We define such a criteria in terms of the absolute difference between two sample standard deviations, $|\sigma(N) - \sigma(N-100)| \leq \mathcal{E}$, computed from N and $N-100$ Monte Carlo realizations. For the tolerance level $\mathcal{E} = 1 \text{ Sm}^3/\text{d}$, the MCS convergence is attained after $N \approx 7300$ realizations (Figure 5.4b). This is the number of realizations used to obtain the results below. For this N , $\mu = 207.8 \text{ Sm}^3/\text{d}$.

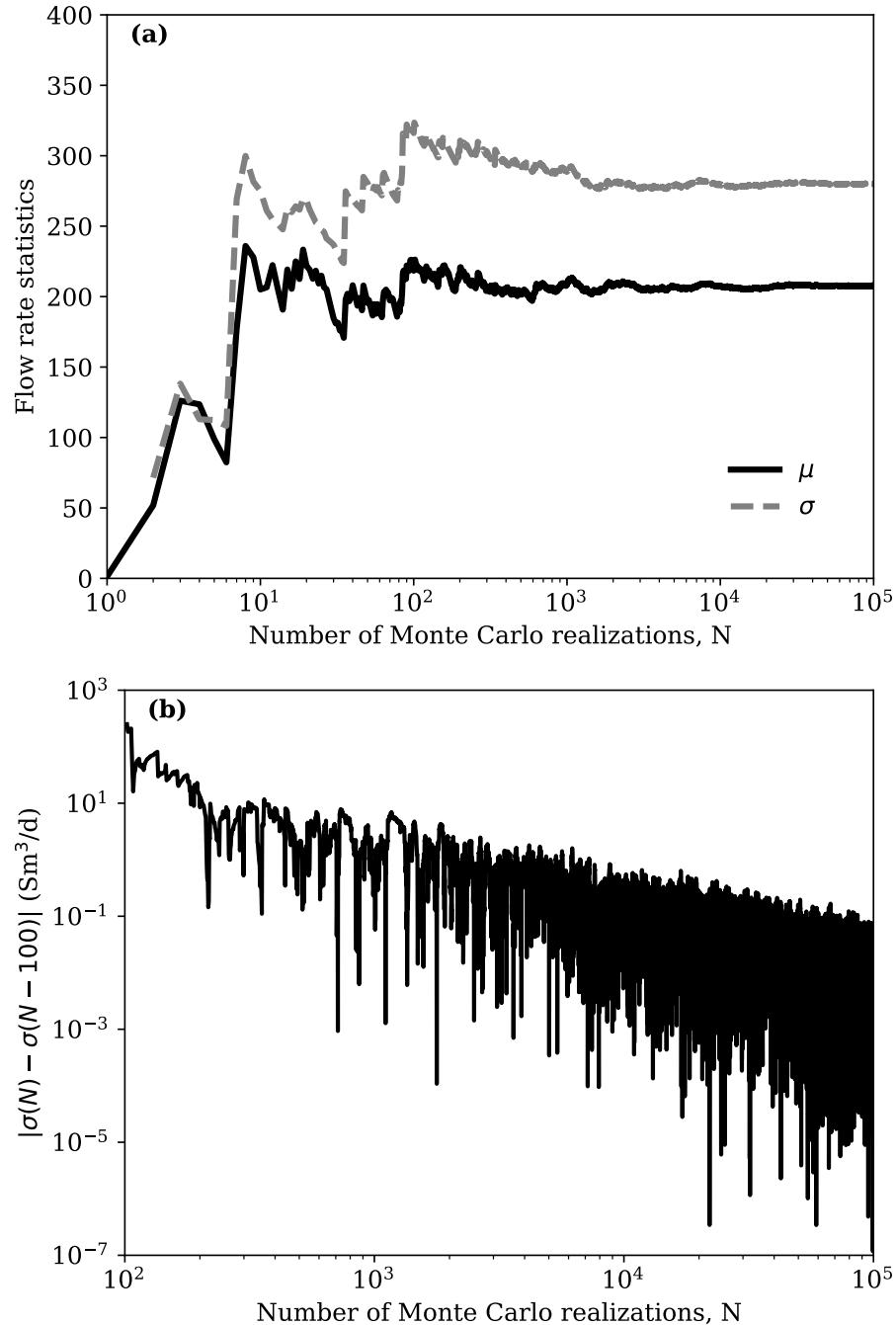


Figure 5.4: Convergence analysis of MCS after 10 years of production: (a) sample mean, μ , and standard deviation, σ , of the cumulative production at the end of the well's lifetime as function of the number of MC realizations, N ; (b) the absolute difference between two sample standard deviations, $|\sigma(N) - \sigma(N - 100)|$, computed from N and $N - 100$ MC realizations.

Table 5.3: Descriptive statistics of the oil flow rate throughout production life.

Production time (years)	Oil flow rate (Sm^3/d)						
	μ	σ	min	P25	P50	P75	max
0	3683.3	115.9	2655.4	3614.9	3691.4	3760.1	3994.4
1	3372.9	108.9	2625.2	3309.6	3380.4	3445.8	3715.0
2	3051.5	118.0	2312.6	2981.2	3059.2	3131.3	3450.4
3	2718.9	140.4	1877.6	2631.7	2725.2	2814.6	3218.9
4	2377.7	169.7	1348.6	2267.3	2382.2	2496.7	3121.2
5	2020.6	203.6	791.4	1883.3	2024.0	2162.4	2729.9
6	1670.4	237.2	0.0	1516.6	1667.3	1826.4	2630.4
7	1302.4	294.5	0.0	1155.0	1326.8	1489.0	2378.6
8	899.6	378.7	0.0	625.6	955.7	1179.2	2095.1
9	488.1	389.1	0.0	171.8	435.4	778.5	1927.1
10	207.8	280.3	0.0	0.0	97.6	333.0	1720.0

Table 5.3 reports MC estimates of the descriptive statistics of the oil flow throughout the well’s production life. As expected, the mean oil flow rate, μ , decreases as the mean watercut increases throughout the production lifetime, in accordance with Eq. 5.2). The standard deviation, σ , increases due to higher uncertainty in the watercut, except at Year 10 when more realizations predict zero flow rate. While the PDF of the oil flow rate is symmetric at earlier years, it becomes skewed to the right towards the end of production lifetime (see, also, Figure 5.5). The positive skewness of the PDFs at the end of production life reflects the large number of MC realizations predicting no production, due to insufficient reservoir pressure to naturally lift fluid from the formation to surface.

It is instructive to compare the probabilistic forecast of the oil flow rate with its deterministic counterpart that uses the expected value of each input in the flow simulation. The latter computation yields the oil flow rates of $3704.4 \text{ Sm}^3/\text{d}$ at production start ($WC = 0$), $2057.4 \text{ Sm}^3/\text{d}$ after 5 years of production ($WC = 40\%$), and $2.0 \text{ Sm}^3/\text{d}$ after 10 years of production ($WC = 80\%$). (Pressure, temperature, and liquid holdup profiles of these simulations are reported in Fig. 5.3.) These deterministic predictions differ significantly from the expected flow rates, especially at the end of production lifetime (Table 5.3). Furthermore, the deterministic approach lacks any uncertainty quantification associated with its predictions, the information

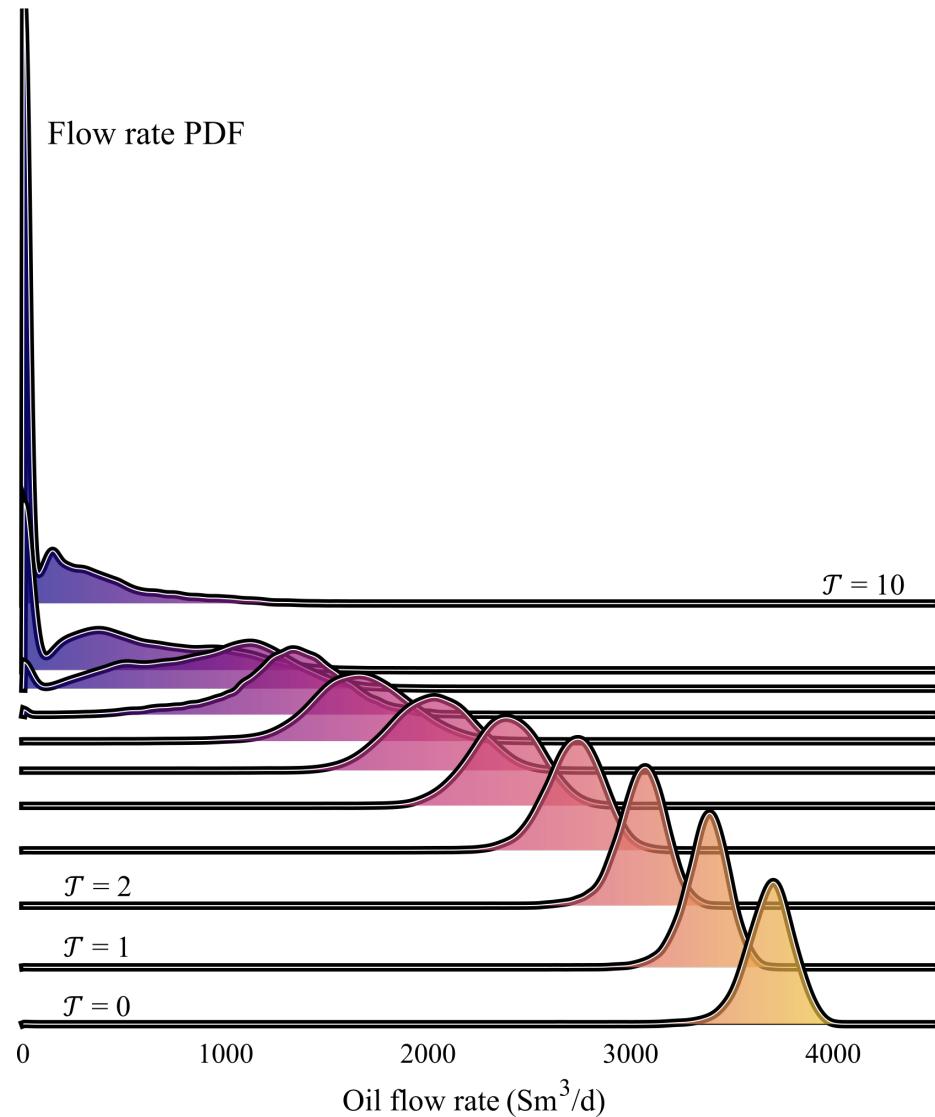


Figure 5.5: Probability density function of the oil production flow rate throughout the well's production lifetime, at several times \mathcal{T} .

that is just as important as the prediction itself.

Mean cumulative production is computed via numerical integration of the mean flow rates. Its statistics—mean and two confidence intervals—are shown in Figure 5.6, together with its deterministic estimate. When the cumulative production is computed from the monthly flow rates (Figure 5.6a), its expected value at the end of the well’s lifetime is 7.20 Mm^3 and the standard deviation is 0.08 Mm^3 , which is 1.1% of the mean. Given that uncertainty in the inputs is significantly higher (20% for gas-oil ratio and 10% for watercut), these results reveal that uncertainty in the values of the reservoir fluid properties is attenuated in the process of computing the cumulative production. The 95th percentile of the total cumulative production is 7.33 Mm^3 , which is 0.27 Mm^3 (3.8%) higher than the 5th percentile.

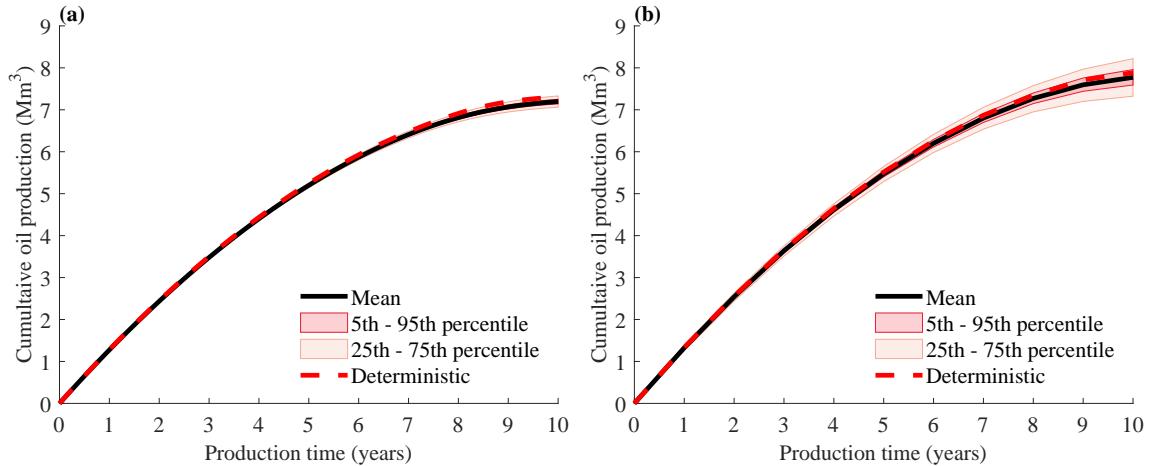


Figure 5.6: Cumulative oil production through the well’s production lifetime considering as time step for the numerical integration of the flow rates: (a) one month and (b) one year.

At the well’s production lifetime of 10 years, the deterministic computation of the cumulative production (7.30 Mm^3) overpredicts the the mean cumulative production (7.20 Mm^3). While this difference might appear to be small, such estimates serve as input to financial evaluations of new production systems projects, in order to quantify the expected revenue and its associated uncertainty. In monetary terms, considering the oil price of the “North Sea Brent” (BRENT) closing value of US\$92.36 on 09/01/2022 as the benchmark price, and applying no discount rate to account for

the various production times, the deterministic framework overestimates the project's revenues by US\$58.6M.

Both expected value and standard deviation of the final cumulative production depend on the time-step size considered for the numerical integration (Figure 5.6). Smaller time steps yield smaller standard deviations, i.e., tighter confidence intervals. On the other hand, smaller time steps increase the computational time, which can be prohibitive depending on the application. For example, the use of the yearly flow rates instead of monthly ones reduces the computational cost of the numerical integration of flow rates over 10 years by 92%. It results in the mean cumulative oil production of 7.77 Mm³ and standard deviation of 0.27 Mm³ (Figure 5.6b). In this case, the deterministic prediction overestimates the project's revenue by US\$63.8M. The 5th and 95th percentiles of the total cumulative production are 7.32 Mm³ and 8.22 Mm³, respectively. Therefore, the 92% reduction in computational time corresponds to the 0.19 Mm³ (US\$17.7M) increase in prediction uncertainty (standard deviation) of cumulative production.

The aforementioned difference between the 5th and 95th percentiles of the cumulative production is equivalent to US\$24.6M and US\$82.8M using the monthly and yearly time steps, respectively. When such large sums of money are involved, which is typical for complex engineering systems, a proper evaluation of the associated risks is essential for making informed business decisions and, ultimately, for the company's success. Hence, it is important that O&G production estimates are performed through the probabilistic approach, where uncertainties are accounted for not only in reservoir flow simulations but also in wellbore and pipeline flow simulations. Attempts to reduce the prediction uncertainty should be considered in view of the cost of information, which may influence arguments in favor or against new data acquisition.

5.3.2 Sensitivity Analysis

Figure 5.7 presents Sobol's indices for the flow simulation model at the beginning, middle, and end of production lifetime, i.e., at the well opening, and after five and ten

years of production, respectively. The mean watercut at each snapshot is 0%, 40%, and 80%, respectively. The sensitivity of the oil flow rate to the input parameters depends on the production scenario. Here, as production time and, consequently, the expected watercut increase, so does their influence on the predicted oil flow rate. In the early production stages, the oil flow rate is more sensitive to the oil specific gravity, while at the end of the production lifetime it is more sensitive to the watercut. The fact that the well produces an emulsion, whose viscosity significantly increases with watercut (according to Woelflin's [136] tight emulsion model), contributes to the great influence of watercut at later production stages. This result suggests that the use of subsea demulsifying strategies may enhance oil production [81].

Throughout production lifetime, the variable with lowest effect on the predictive uncertainty of the oil flow rate is the gas specific gravity, which means that the multiphase model is less sensitive to variations in this input than to changes in oil density, gas-oil ratio or watercut. Since first-order and total-order indices are similar, the model has no significant interaction effects between γ_o , γ_g , gas-oil ratio, and watercut.

5.4 Conclusion

Uncertainty quantification plays a key role in decision making for O&G projects, especially in risk assessment of its heavy investments. While uncertainty quantification has been an established practice in reservoir simulation, it has not been thoroughly investigated for multiphase flow models used in flow assurance. This paper quantifies uncertainty in predictions of the production flow rate over the life time of a simplified real production well and, consequently, of its cumulative production. The sources of uncertainty are oil and gas specific gravities, reservoir fluid gas-oil ratio, and watercut. Uncertainty of such inputs is modeled through a multivariate Gaussian probability density function, while uncertainty in the flow simulation output (i.e., the production flow rate) is quantified via Monte Carlo simulations with Latin hypercube sampling. A global sensitivity analysis is performed to identify which of the four fluid parameters affect most the overall predictive uncertainty. Major conclusions are summarized

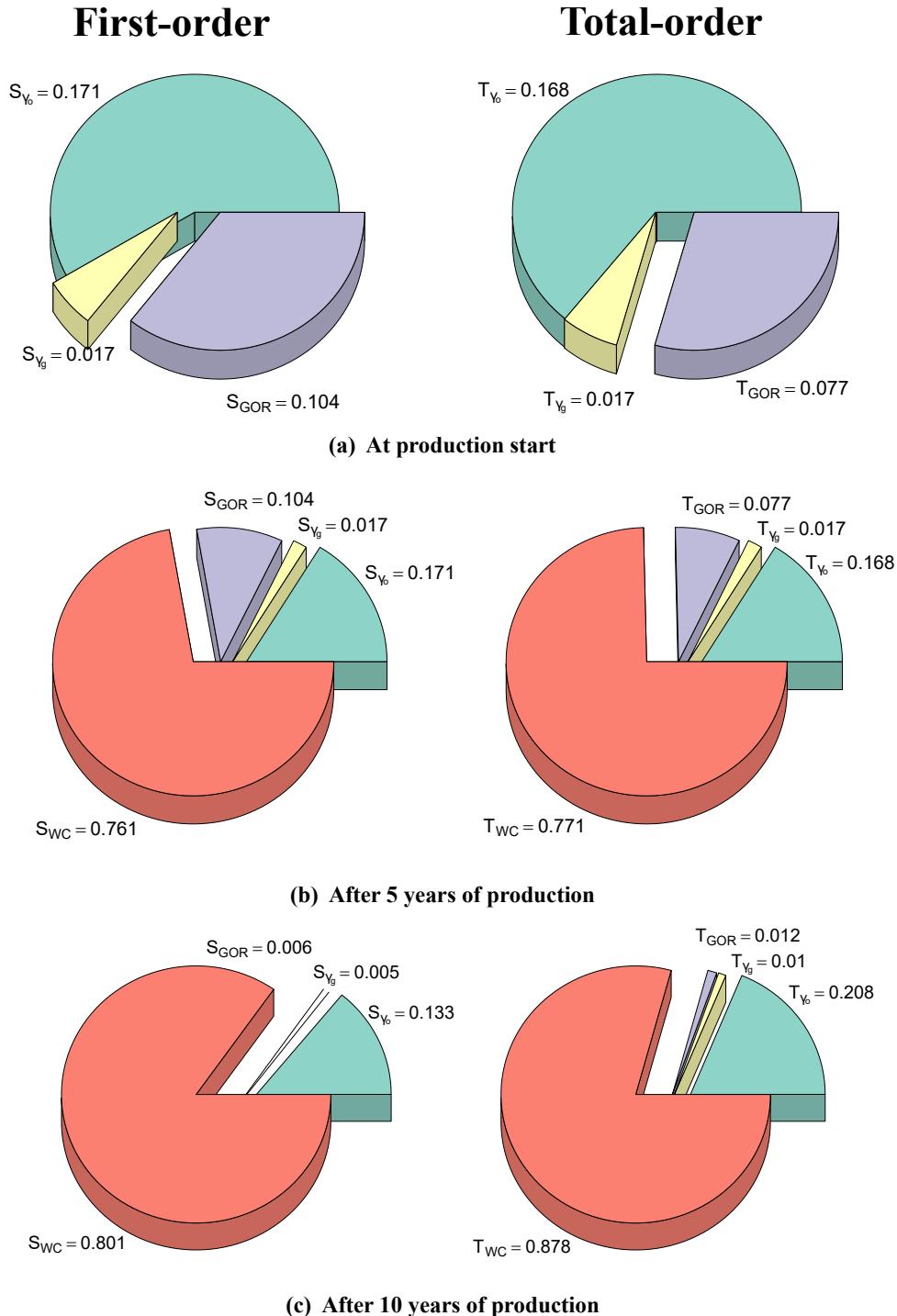


Figure 5.7: First-order and total-order Sobol's indices for the input parameters (γ_o , γ_g , GOR, WC) of flow simulation model (a) at the well opening, (b) after five years of production, and (c) after ten years of production.

below:

- The sensitivity of the simulated oil flow rate to input parameters depends on the production stage. As production time and, consequently, the expected watercut increase, so does their influence on the flow rate. By the end of the well's lifetime, watercut dominates the flow rate predictive uncertainty, followed by oil density. Conversely, gas density is the input variable with lowest effect on the predictive uncertainty of the oil flow rate throughout the whole production lifetime.
- Convergence properties of Monte Carlo simulations were discussed. The sample size depends on the goal of the study, i.e., on the acceptable error tolerance in predictions of a quantity of interest.
- Flow rate probability distributions change over production time, not only in their parameters but also in shape. While they are symmetric at early production stages, they are skewed to the right at later stages due to the higher number of realizations that result in a zero flow rate. This asymmetry of the output distribution happens despite of the symmetry of the inputs distribution. Also, the expected oil flow rate decreases with production time due to watercut increase.
- Cumulative production estimates depend on the time step used in the numerical integration of computed flow rates. Its expected value and standard deviation at the end of the well's lifetime are 7.20 Mm^3 and 0.08 Mm^3 , respectively, considering a monthly time step. For a yearly time step, these statistics are 7.77 Mm^3 and 0.27 Mm^3 , respectively. These results indicate that uncertainties associated with fluid properties are attenuated when computing the cumulative production.
- The deterministic framework overestimates production when compared to the probabilistic framework while suggesting certainty. This fact highlights the importance of the probabilistic assessment of estimates from pipe flow simulation models.

- Uncertainty quantification of reservoir fluid properties is particularly important during the design phase of a reservoir development project, when field data of fluid properties or pressure and temperature in the wellbore and pipeline are not available. Once field data is available, they can be used to restrict the uncertainty on the simulation inputs parameters through data assimilation techniques.

The analyses presented here are based on a simplified case study. We did not consider a minimum production rate for the well to be operating, i.e., a minimum revenue that covers operational costs and prevents the well from being shut down, which would be a likely assumption in a real case scenario. The presence of flow instabilities (such as severe slugging) that could reduce production was not evaluated. Changes in fluid composition, i.e., in oil and gas densities and gas-oil ratio, may affect the reservoir's relative permeability [140] and consequently the reservoir's productivity index. This and other-second order interactions were also not considered in this work. They are expected to increase the uncertainty in predictions of the quantity of interest even further, and should be a focus of future research.

Chapter 6

Uncertainty in flow modeling

Previous chapters have investigated uncertainties associated with BO and compositional fluid models and their impact on flow model predictions. In the oil and gas industry, it is widely acknowledged that flow simulation results inherently contain uncertainties, even when the specific sources and magnitudes of these uncertainties are not well-defined. Consequently, flow predictions are not expected to perfectly align with field data. When field data is available, data assimilation techniques can be employed to enhance the predictive capacity of flow models. Nonetheless, the adoption of these techniques in the realm of flow assurance has not yet become widespread.

In practical applications, pipe-flow models representing the flow within pipelines and wells are often calibrated to field data through tuning factors. This chapter examines the implementation of tuning factors in flow models and proposes a heuristic method for optimizing tuning factors applied to the calculated pressure and temperature gradients.

6.1 Introduction

Multiphase pipe flow simulations play a key role in carbon capture, utilization and storage (CCUS) applications and in the oil and gas (O&G) industry. The accurate calculation of pressure and temperature profiles from/to the surface facility to/from

the well bottom hole is essential to the design, operation, and optimization of injection/production systems. The enormous amount of research and development work that has been invested in multiphase flow modeling over the past 60 years [114] evidences the importance of the topic.

As a result of such work, several flow models historically applied mostly to petroleum industry problems have been proposed in the literature. Until the 1980's empirical correlations for predicting the pressure gradient were most common; then mechanistic models that better describe the flow physical phenomena gained traction [21]. More recently, machine learning techniques have been applied to characterize the multiphase flow.

Nonetheless, even the best models are not always able to match experimental/field data since their predictions are fundamentally uncertain. Such uncertainties can be structural, due to inevitable approximations of "reality" introduced by a model, or parametric, due to the need to specify numerical values for the model's inputs [45]. Almost every model and simulator has been developed with a limited applicability for some flow pattern and some pipe inclination [26], and most researchers agree that no single correlation is applicable over all ranges of variables with suitable accuracy [61].

For instance, different flow models may show an extreme variability in performance when compared to real field data from oil and gas producing fields, with no model giving the best results for all fields [100]. For a given real production system, governing equations that most accurately predicts pressure, temperature and liquid holdup can vary with operational conditions [111]. This uncertainty holds even for smaller-scale experimental settings. An example is the phase slip and frictional pressure drop models uncertainty for upward and downward two-phase flow of CO₂ [53]. Regarding the fluid modeling, subjective choices of optimization algorithm and initial guess used in the equation of state regression may affect the final fluid characterization [46], and consequently flow predictions. Also, uncertainties in the fluid model selection may significantly impact flow simulation results [42, 41].

A common practice to increase the accuracy and predictability of flow models is to apply tuning factors on calculated parameters in order to better represent field

measurements available. Such tuning factors are essentially multiplicative correction constants for one or more inputs and/or calculated variables. This is the approach used by traditional simulators. For example, PIPESIM® v.2017 applies tuning factors on computed friction factors, holdups and overall heat transfer coefficients (U-values). Petrobras' in-house steady-state multiphase simulator, Marlim II, applies them on computed pressure and temperature gradients. The use of tuning factors in pipe flow models is analogous to the well established practice of History Matching (HM) in reservoir engineering. Nonetheless, while HM techniques focused on adjusting reservoir models have been extensively studied [82], they have not yet been thoroughly investigated for multiphase flow models used in flow assurance. This disparity may be mainly due to the fact that reservoir simulations can take hours or even days for more complex cases, whereas typical pipe flow simulations take seconds or minutes.

Finding optimal tuning factors to a particular flow model consists primarily of an optimization problem. The tuning factors to be optimized are the design variables. The constraints are the upper and lower bounds in which the tuning factors are allowed to vary. The objective function to be minimized evaluates the flow model prediction errors against measured data. In O&G applications, field data are usually obtained during production tests given that production wells are required to undergo periodic testing for regulatory compliance. In such tests, production flow is diverted to a test separator, which is used to separate and measure the oil, gas, and water flowrates.

Previous studies have applied traditional optimization algorithms to determine the optimal tuning factors in pipe flow simulations. Monteiro et al. [78] and Chaves et al. [24] adjust pressure gradients of well flow models using the bisection method. Although this approach is efficient for adjusting solely pressure gradients, its performance may deteriorate when incorporating additional measured variables, such as temperature gradients, into the error function. Seman et al. [113] implement the OrthoMADS direct-search algorithm [2] to compute tuning factors for pressure and temperature gradients estimated by well flow models. The algorithm takes 464 s (in a computer with two Intel Core Xeon E5-2630 v4 2.20 GHz processors and 64 GB of RAM) and over 100 iterations to converge in a case study.

We propose a heuristic strategy for optimizing tuning factors applied on pressure and temperature gradients computed by pipe flow models. Heuristic techniques offer satisfactory or near-optimal solutions to optimization problems more rapidly and efficiently than traditional (generic) methods. These solutions can serve as informed initial estimates to expedite the optimization process when identifying global minima is crucial. The proposed method has been effectively employed to calibrate well flow models at Petrobras for nearly a decade, demonstrating its computational efficiency and efficacy.

To evaluate the performance of the proposed heuristic, we conduct a comparative analysis of the tuning factor results obtained during the calibration of a real well model, considering four alternative optimization methods. These alternative methods, which employ distinct search space exploration strategies, include the adaptive moment estimation (ADAM), Davidon-Fletcher-Powell (DFP), covariance matrix adaptation evolution strategy (CMA-ES), and Hooke-Jeeves (HJ) algorithms. Our heuristic yields highly competitive results at a substantially reduced computational cost. Moreover, we illustrate a two-step optimization approach, wherein the heuristic method provides the initial estimates for the generic algorithms.

6.2 Methodology

Our methodology is flow model-independent, i.e. it is applicable to any set of governing equations and closure relations available in the literature (e.g. empirical, Two-Fluid, and Drift-Flux approaches). We consider a general multiphase flow model denoted by function f , with inputs denoted in vector form by \mathbf{x} and outputs (i.e. pressure and temperature values predicted throughout the domain of interest) denoted by $\hat{\mathbf{y}}$, so that $\hat{\mathbf{y}} = f(\mathbf{x})$. The inputs vector contains the boundary conditions to the flow simulation: outlet pressure, well bottom hole temperature and water, gas, and oil production flow rates.

We assume that the system is equipped with two subsea sensors, known as permanent downhole gauge (PDG) and TPT, which are located in the tubing and at the wellhead, respectively. Both sensors measure pressure and temperature. Moreover, a

surface sensor measures pressure and temperature at the outlet. Since it is possible to calculate measured pressure and temperature differences in two segments (well and pipe), we assume four tuning factors among the inputs:

- ϕ_P^{well} and ϕ_P^{pipe} are multipliers for the calculated pressure gradients in the tubing and flowline, respectively;
- ϕ_T^{well} and ϕ_T^{pipe} are multipliers for the calculated temperature gradient in the tubing and flowline, respectively.

Therefore, keeping the fluid and flow characteristics constant, i.e. for a particular production/injection system, $\hat{\mathbf{y}} = f(\boldsymbol{\phi}) = f((\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}})^\top)$. In particular, the model i th outputs are obtained by imposing boundary conditions of the i th production test, so that $\hat{\mathbf{y}}_i = f_i(\boldsymbol{\phi})$.

It is worth noticing that pressure and temperature gradients tuning factors must represent flow segments for which measurements are available, for a well-formulated problem. Because the numerical solver for the flow model associates both tuning factors to each element of the spatial discretization (mesh), the methodology can be easily adapted to include more tuning factors if additional pressure and/or temperature measurements are available. Similarly, it can be easily adapted to reduce the number of tuning factors in case no pressure and/or temperature measurements from the TPT sensor are available.

6.2.1 Optimization problem formulation

The objective function F to be minimized is the mean absolute percentage error (MAPE) of N_{meas} pressure and temperature measurements from different sensors obtained during N_{tests} production tests,

$$\begin{aligned} \text{MAPE} &= F(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_{N_{\text{tests}}}) \\ &= F(f(\boldsymbol{\phi})) \end{aligned}$$

$$= \frac{1}{N_{\text{tests}} + N_{\text{meas}}} \sum_{i=1}^{N_{\text{tests}}} \sum_{j=1}^{N_{\text{meas}}} \left| \frac{y_i^j - \hat{y}_i^j(\phi)}{y_i^j} \right| \quad (6.1)$$

where $\hat{y}_i^j(\phi)$ is the j th measurement of pressure or temperature predicted by the i th production test flow model with design variables ϕ .

It is worth mentioning that PIPESIM® v.2017 uses the root mean square error (RMSE) of the N_{meas} measurements as the optimization problem's objective function [112], which can be misleading due to unit inconsistencies between the variables being measured.

The tuning factors are initialized as 1. The upper and lower bounds for each design variable are defined as a percentage of their initial guesses. A maximum variation of 20% is permitted for the pressure and temperature gradient multipliers; higher values would potentially indicate that the flow model is not able to represent the underlying physics of the real system. Representing these limits by the vector $\Delta\phi$, the constrained optimization problem is

$$\underset{\phi}{\text{minimize}} \quad F(\phi), \quad \text{subject to} \quad 1 - \Delta\phi_k \leq \phi_k \leq 1 + \Delta\phi_k \quad \text{for all } k. \quad (6.2)$$

The stopping criteria for the optimization process is either the convergence of the objective function F within a tolerance of 10^{-3} for the change in $|F|$ between two consecutive iterations or a maximum number of 20 iterations.

6.2.2 Alternative optimization strategies

Besides of the proposed optimization algorithm (described in Section 6.3), four alternative algorithms were considered in this paper for comparison purposes. They were selected both due to their popularity in the field and conceptual dissimilarity from each other. They are, namely: adaptive moment estimation method (ADAM), Davidon-Fletcher-Powell method (DFP), covariance matrix adaptation evolution strategy method (CMA-ES), and Hooke-Jeeves algorithm (HJ); and are summarized below. All of their hyperparameters were optimized for this particular application.

The adaptive moment estimation method [64] ADAM is a first-order method that computes individual adaptive learning rates for each parameter from estimates of the first and second moments of the gradients of the objective function. At each iteration, the intermediate variables

$$\mathbf{m}_t \leftarrow \beta_1 \mathbf{m}_{t-1} + (1 - \beta_1) \nabla F(\boldsymbol{\phi}_{t-1}), \quad (6.3a)$$

$$\hat{\mathbf{m}}_t \leftarrow \frac{\mathbf{m}_t}{(1 - \beta_1^t)}, \quad (6.3b)$$

$$\mathbf{v}_t \leftarrow \beta_2 \mathbf{v}_{t-1} + (1 - \beta_2) \nabla F(\boldsymbol{\phi}_{t-1}) \odot \nabla F(\boldsymbol{\phi}_{t-1}), \quad (6.3c)$$

$$\hat{\mathbf{v}}_t \leftarrow \frac{\hat{\mathbf{v}}_t}{(1 - \beta_2^t)} \quad (6.3d)$$

are computed in order to update the design point

$$\boldsymbol{\phi}_t \leftarrow \boldsymbol{\phi}_{t-1} - \alpha \hat{\mathbf{m}}_t \oslash (\sqrt{\hat{\mathbf{v}}_t} + 10^{-8}). \quad (6.3e)$$

The symbols \odot and \oslash refer to the element-wise vector product and division, respectively.

The Davidon-Fletcher-Powell method [33, 40] DFP is a second-order gradient descent method. It rests on the Newton's method,

$$\mathbf{x}_t \leftarrow \mathbf{x}_{t-1} - (\mathbf{H}_{t-1})^{-1} \nabla F(\mathbf{x}_{t-1}), \quad (6.4)$$

where the inverse of the Hessian matrix \mathbf{H} is approximated by a symmetric and positive definite matrix \mathbf{Q} . Because the second-order information is approximated, it is called a quasi-Newton method. At each iteration t , the design point is updated according to

$$\mathbf{x}_t \leftarrow \mathbf{x}_{t-1} - \alpha_{t-1} \mathbf{Q}_{t-1} \nabla F(\mathbf{x}_{t-1}), \quad (6.5)$$

where α_{t-1} is a scalar step factor.

The covariance matrix adaptation evolution strategy method [55] CMA-ES is a stochastic method inspired by natural evolution strategies such as recombination, mutation and elite selection techniques. It is recognized to be among the leading algorithms for optimization of real-valued functions [39]. At each iteration step, the algorithm improves the mean and the covariance of a multivariate Gaussian search distribution. Change rates for the search distribution parameters and for the step size are updated separately.

Hooke-Jeeves algorithm [57] HJ performs a direct search over the search space based on function evaluations in each coordinate direction, with steps of magnitude α . At each iteration, for $\phi \in \mathbb{R}^4$, $F[\phi + (\alpha, 0, 0, 0)^\top]$, $F[\phi + (-\alpha, 0, 0, 0)^\top]$, $F[\phi + (0, \alpha, 0, 0)^\top]$, $F[\phi + (0, -\alpha, 0, 0)^\top]$, ..., $F[\phi + (0, 0, 0, \alpha)^\top]$, and $F[\phi + (0, 0, 0, -\alpha)^\top]$ are evaluated. If any improvement is found, the anchoring point ϕ thus moves to the position in which the objective function is smaller. On the other hand, if no improvements are verified, the step size decreases for a finer search.

6.2.3 Case study

We demonstrate the proposed heuristic algorithm by applying it to the calibration of a real offshore oil production well located in Campos Basin, Brazil. This well uses the gas lift method for artificial lifting. A schematic representation of the well, along with its key characteristics, can be found in Figure 6.1. Flow simulations are performed with Petrobras' in-house simulator MARLIM II® (Multiphase flow and ARTificial LIft Modeling), which outputs pressure and temperature profiles.

The boundary conditions required for these flow simulations include the outlet pressure (P_{out}), reservoir temperature, produced liquid flow rate (Q_{liq}), watercut (WC), gas-oil ratio (GOR), as well as the gas lift injection flow rate (Q_{gl}) and temperature. Our analysis incorporates nine production tests conducted over a 2.5-year period, , with the aim of adjusting tuning factors to accurately represent the production history. Table 6.1 reports of the boundary conditions associated with each production test. Throughout these tests, reservoir and gas injection temperatures remain constant at 61°C and 40°C, respectively, with the GOR held at 69 Sm³/Sm³.

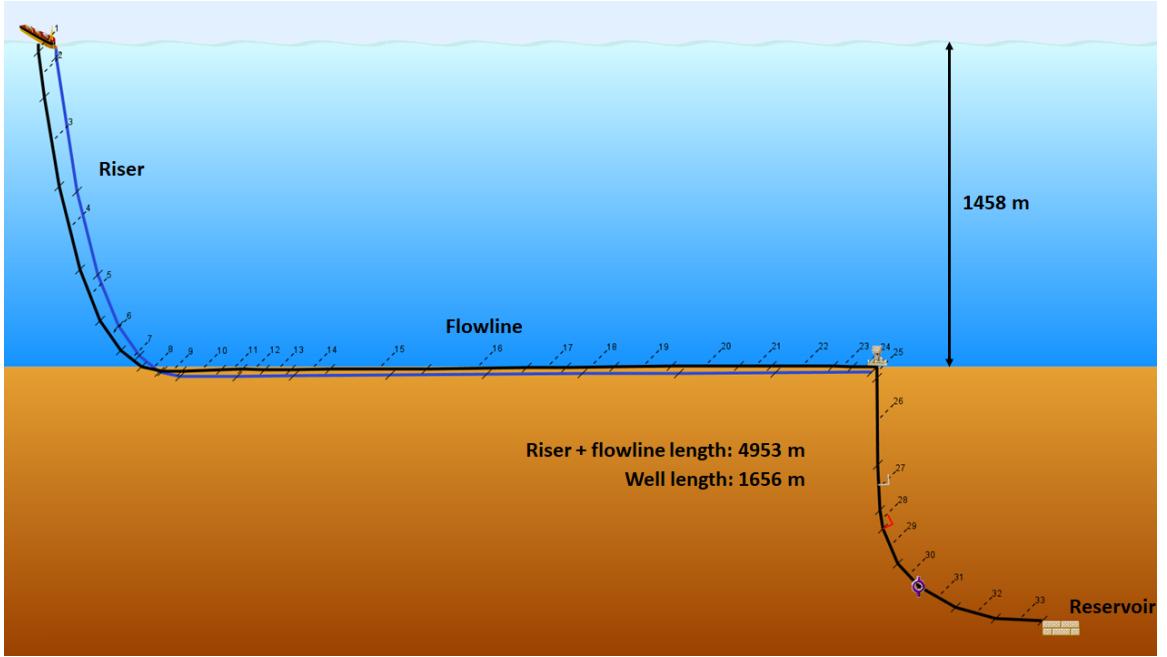


Figure 6.1: Schematic of a simplified production well implemented in the flow simulator.

Pressure and temperature values measured at the PDG (P^{PDG} and T^{PDG}), TPT (P^{TPT} and T^{TPT}), and outlet (T^{out}) during the production tests are reported in Table 6.2.

6.3 Proposed optimization algorithm

The optimization starts from $\phi = (\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}})^{\top} = (1, 1, 1, 1)^{\top}$. With these tuning factor values, flow simulations are performed for each i th production test, $i = 1, \dots, N_{\text{tests}}$, i.e. with its specific boundary conditions. Each production test simulation outputs a pressure and temperature profile.

We represent pressure predictions as vectors $\hat{\mathbf{P}}^{PDG}$, $\hat{\mathbf{P}}^{TPT}$, $\hat{\mathbf{P}}^{out} \in \mathbb{R}^{N_{\text{tests}}}$, with \hat{P}_i^{PDG} , \hat{P}_i^{TPT} , and \hat{P}_i^{out} being the calculated pressures at the PDG sensor, TPT sensor, and outlet, respectively, for the i th production test. For temperature predictions, $\hat{\mathbf{T}}_i^{PDG}$, $\hat{\mathbf{T}}_i^{TPT}$, $\hat{\mathbf{T}}_i^{out} \in \mathbb{R}^{N_{\text{tests}}}$, such that \hat{T}_i^{PDG} , \hat{T}_i^{TPT} , and \hat{T}_i^{out} are the calculated temperatures at the PDG sensor, TPT sensor, and outlet, respectively. $\Delta \hat{\mathbf{P}}^{\text{well}}$,

Production test ID	P_{out} (barg)	Q_{liq} (Sm^3/d)	WC (%)	Q_{gl} (Sm^3/d)
1	27.1	4146	0.6	268522
2	20.0	4144	12.7	216059
3	13.9	4347	19.2	256944
4	13.9	4357	31.9	261032
5	12.5	4303	38.1	254150
6	12.7	4182	40.4	260798
7	12.6	4155	46.4	258019
8	11.8	4193	44.0	257558
9	12.0	4157	45.5	263600

Table 6.1: Boundary conditions from nine production tests.

Production test ID	P^{PDG} (barg)	T^{PDG} ($^{\circ}\text{C}$)	P^{TPT} (barg)	T^{TPT} ($^{\circ}\text{C}$)	T^{out} ($^{\circ}\text{C}$)
1	215.0	60.3	132.3	51.4	35.3
2	212.4	60.4	126.9	53.2	38.8
3	212.3	60.6	125.4	53.3	39.3
4	213.6	60.6	125.7	53.7	41.4
5	215.6	60.8	127.4	54.3	42.2
6	216.8	60.8	126.3	54.8	43.9
7	216.8	60.9	126.2	54.9	43.3
8	217.7	60.9	126.7	55.0	44.0
9	218.1	61.0	127.4	55.0	44.0

Table 6.2: Pressure and temperature measurements at the PDG, TPT, and outlet during nine production tests.

$\Delta \hat{\mathbf{P}}^{\text{pipe}}$, $\Delta \hat{\mathbf{T}}^{\text{well}}$, $\Delta \hat{\mathbf{T}}^{\text{pipe}} \in \mathbb{R}^{N_{\text{tests}}}$ are thus obtained:

$$\Delta \hat{\mathbf{P}}^{\text{well}} = \hat{\mathbf{P}}^{\text{PDG}} - \hat{\mathbf{P}}^{\text{TPT}} \quad (6.6)$$

$$\Delta \hat{\mathbf{P}}^{\text{pipe}} = \hat{\mathbf{P}}^{\text{TPT}} - \hat{\mathbf{P}}^{\text{out}} \quad (6.7)$$

$$\Delta \hat{\mathbf{T}}^{\text{well}} = \hat{\mathbf{T}}^{\text{PDG}} - \hat{\mathbf{T}}^{\text{TPT}} \quad (6.8)$$

$$\Delta \hat{\mathbf{T}}^{\text{pipe}} = \hat{\mathbf{T}}^{\text{TPT}} - \hat{\mathbf{T}}^{\text{out}} \quad (6.9)$$

Analogously, we represent pressure measurements as \mathbf{P}^{PDG} , \mathbf{P}^{TPT} , $\mathbf{P}^{\text{out}} \in \mathbb{R}^{N_{\text{tests}}}$, with P_i^{PDG} , P_i^{TPT} , and P_i^{out} denoting the measured pressures at the PDG sensor, TPT sensor, and outlet, respectively, during the i th production test. For temperature measurements, \mathbf{T}^{PDG} , \mathbf{T}^{TPT} , $\mathbf{T}^{\text{out}} \in \mathbb{R}^{N_{\text{tests}}}$, where T_i^{PDG} , T_i^{TPT} , and T_i^{out} denotes the measured temperatures at the PDG sensor, TPT sensor, and outlet, respectively, during the i th production test. $\Delta \mathbf{P}^{\text{well}}$, $\Delta \mathbf{P}^{\text{pipe}}$, $\Delta \mathbf{T}^{\text{well}}$, $\Delta \mathbf{T}^{\text{pipe}} \in \mathbb{R}^{N_{\text{tests}}}$ are thus calculated:

$$\Delta \mathbf{P}^{\text{well}} = \mathbf{P}^{\text{PDG}} - \mathbf{P}^{\text{TPT}} \quad (6.10)$$

$$\Delta \mathbf{P}^{\text{pipe}} = \mathbf{P}^{\text{TPT}} - \mathbf{P}^{\text{out}} \quad (6.11)$$

$$\Delta \mathbf{T}^{\text{well}} = \mathbf{T}^{\text{PDG}} - \mathbf{T}^{\text{TPT}} \quad (6.12)$$

$$\Delta \mathbf{T}^{\text{pipe}} = \mathbf{T}^{\text{TPT}} - \mathbf{T}^{\text{out}} \quad (6.13)$$

The tuning factors are updated at each iteration according to:

$$\phi_P^{\text{well}} \leftarrow \|\Delta \mathbf{P}^{\text{well}}\|_2^2 (\Delta \mathbf{P}^{\text{well}} \cdot \widehat{\Delta \mathbf{P}}^{\text{well}})^{-1} \phi_P^{\text{well}} \quad (6.14)$$

$$\phi_P^{\text{pipe}} \leftarrow \|\Delta \mathbf{P}^{\text{pipe}}\|_2^2 (\Delta \mathbf{P}^{\text{pipe}} \cdot \widehat{\Delta \mathbf{P}}^{\text{pipe}})^{-1} \phi_P^{\text{pipe}} \quad (6.15)$$

$$\phi_T^{\text{well}} \leftarrow \|\Delta \mathbf{T}^{\text{well}}\|_2^2 (\Delta \mathbf{T}^{\text{well}} \cdot \widehat{\Delta \mathbf{T}}^{\text{well}})^{-1} \phi_T^{\text{well}} \quad (6.16)$$

$$\phi_T^{\text{pipe}} \leftarrow \|\Delta \mathbf{T}^{\text{pipe}}\|_2^2 (\Delta \mathbf{T}^{\text{pipe}} \cdot \widehat{\Delta \mathbf{T}}^{\text{pipe}})^{-1} \phi_T^{\text{pipe}} \quad (6.17)$$

where \cdot denotes the dot product.

Constraints are imposed by limiting the tuning factors to the upper and lower bounds at each iteration. For instance, considering ϕ_P^{well} and a maximum variation of 20% allowed, imposing the constraints is implemented as:

$$\phi_P^{\text{well}} \leftarrow \max(\phi_P^{\text{well}}, 0.8) \quad (6.18)$$

$$\phi_P^{\text{well}} \leftarrow \min(\phi_P^{\text{well}}, 1.2) \quad (6.19)$$

The updated tuning factor values are utilized as input for the flow simulations of production tests in the following iteration. This iterative process continues until convergence is reached or a predefined termination criterion is satisfied. In this paper, the convergence criteria is that the absolute difference between two successive objective function evaluations must be less than $\epsilon = 1 \times 10^{-3}$, with a maximum number of iterations set at $N_{\text{max}} = 20$. The method involves solving a series of convex optimization problems, which are local and approximate representations of the original problem.

Algorithms 6-7 presents a pseudocode for the heuristic method implementation and its auxiliary functions.

6.4 Results

Figure 6.2 presents the convergence history of the tuning parameters for the proposed heuristic method along with the four alternative optimization methods. As these algorithms employ distinct strategies for exploring the search space, both their convergence history and final optimized points differ. Notably, the ADAM and Hooke-Jeeves algorithms halt upon reaching the maximum number of iterations, set to 20 in this study. Interestingly, it has been observed that the heuristic method typically converges in fewer than 10 iterations for a wide variety of models.

Figure 6.3 displays the Mean Absolute Percentage Error (MAPE) at each iteration, representing the value of the objective function. All algorithms significantly reduce the error of the original (untuned) model. Due to its stochastic nature, the initial iterations of the CMA-ES method exhibit a high error, which is subsequently rectified

Algorithm 6 Part 1. Proposed heuristic method for the constrained optimization of tuning factors applied on pressure and temperature gradients calculated in flow pipe simulations.

```

1: procedure FITTUNINGFACTORS( $\Delta\mathbf{P}^{\text{well}}$ ,  $\Delta\mathbf{P}^{\text{pipe}}$ ,  $\Delta\mathbf{T}^{\text{well}}$ ,  $\Delta\mathbf{T}^{\text{pipe}}$ )
2:    $\epsilon \leftarrow 10^{-3}$                                  $\triangleright$  Specify threshold for convergence criteria
3:    $N_{\text{max}} \leftarrow 20$                              $\triangleright$  Specify maximum number of iterations
4:    $N_{\text{iter}} \leftarrow 0$                              $\triangleright$  Initialize number of iterations
5:    $\Delta F \leftarrow 10$                                  $\triangleright$  Initialize  $\Delta F > \epsilon$ 
6:    $(\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}}) = (1, 1, 1, 1)$        $\triangleright$  Initialize tuning factors
7:    $F, \Delta\hat{\mathbf{P}}^{\text{well}}, \Delta\hat{\mathbf{P}}^{\text{pipe}}, \Delta\hat{\mathbf{T}}^{\text{well}}, \Delta\hat{\mathbf{T}}^{\text{pipe}} \leftarrow \text{COMPUTEOBJECTIVEFUNCTION}(\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}})$ 
8:   while  $\Delta F > \epsilon$  &  $N_{\text{iter}} < N_{\text{max}}$  do
9:     for all  $k \in \text{well, pipe}$  do
10:     $\phi_P^k \leftarrow \|\Delta\mathbf{P}^k\|_2^2 (\Delta\mathbf{P}^k \cdot \Delta\hat{\mathbf{P}}^k)^{-1} \phi_P^k$ 
11:     $\phi_T^k \leftarrow \|\Delta\mathbf{T}^k\|_2^2 (\Delta\mathbf{T}^k \cdot \Delta\hat{\mathbf{T}}^k)^{-1} \phi_T^k$ 
12:     $\triangleright$  Impose specified constraints on tuning factors:
13:     $\phi_P^k \leftarrow \max(\min(\phi_P^k, 1.2), 0.8)$ 
14:     $\phi_T^k \leftarrow \max(\min(\phi_T^k, 1.2), 0.8)$ 
15:   end for
16:    $F_0 \leftarrow F$ 
17:    $F, \Delta\hat{\mathbf{P}}^{\text{well}}, \Delta\hat{\mathbf{P}}^{\text{pipe}}, \Delta\hat{\mathbf{T}}^{\text{well}}, \Delta\hat{\mathbf{T}}^{\text{pipe}} \leftarrow \text{COMPUTEOBJECTIVEFUNCTION}(\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}})$ 
18:    $\Delta F \leftarrow |F - F_0|$ 
19:    $N_{\text{iter}} \leftarrow N_{\text{iter}} + 1$ 
20: end while
21: end procedure

```

Algorithm 7 Part 2. Auxiliary functions.

```

22: function COMPUTEOBJECTIVEFUNCTION( $\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}}$ )
23:   for  $i = 1$  to  $N_{\text{tests}}$  do
24:      $\hat{\mathbf{y}}_i \leftarrow \text{RUNSIMULATION}(i, (\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}}, \phi_T^{\text{pipe}})^\top)$ 
25:      $\Delta\hat{P}_i^{\text{well}}, \Delta\hat{P}_i^{\text{pipe}}, \Delta\hat{T}_i^{\text{well}}, \Delta\hat{T}_i^{\text{pipe}} \leftarrow \text{GET}\Delta\mathbf{S}(\hat{\mathbf{y}}_i)$ 
26:   end for
27:    $F \leftarrow \text{COMPUTEEERROR}(\hat{\mathbf{y}}_1, \hat{\mathbf{y}}_2, \dots, \hat{\mathbf{y}}_{N_{\text{tests}}})$             $\triangleright$  According to Eq. (6.1), for
      example
28:   return  $F, \Delta\hat{P}^{\text{well}}, \Delta\hat{P}^{\text{pipe}}, \Delta\hat{T}^{\text{well}}, \Delta\hat{T}^{\text{pipe}}$ 
29: end function

30: function RUNSIMULATION( $i, \phi$ )
31:    $\hat{\mathbf{y}} \leftarrow f_i(\phi)$             $\triangleright f_i$  runs the flow simulation of the  $i$ th production test
32:   return  $\hat{\mathbf{y}}$ 
33: end function

34: function GET $\Delta\mathbf{S}(\hat{\mathbf{y}})$ 
35:    $(\dots, \hat{P}_{PDG}, \dots, \hat{P}_{TPT}, \dots, \hat{P}_{out}, \dots, \hat{T}_{PDG}, \dots, \hat{T}_{TPT}, \dots, \hat{T}_{out}) \leftarrow \hat{\mathbf{y}}$ 
36:    $\Delta\hat{P}^{\text{well}} \leftarrow \hat{P}_{PDG} - \hat{P}_{TPT}$ 
37:    $\Delta\hat{P}^{\text{pipe}} \leftarrow \hat{P}_{TPT} - \hat{P}_{out}$ 
38:    $\Delta\hat{T}^{\text{well}} \leftarrow \hat{T}_{PDG} - \hat{T}_{TPT}$ 
39:    $\Delta\hat{T}^{\text{pipe}} \leftarrow \hat{T}_{TPT} - \hat{T}_{out}$ 
40:   return  $\Delta\hat{P}^{\text{well}}, \Delta\hat{P}^{\text{pipe}}, \Delta\hat{T}^{\text{well}}, \Delta\hat{T}^{\text{pipe}}$ 
41: end function

```

in later iterations. Despite their varying convergence histories and final outcomes, all algorithms attain a comparable final MAPE.

Table 6.3 compares the running time of the five different algorithms for this specific multiphase flow model. Our implementation was written in Julia version 1.6.3 and run on a computer system with an Intel Core i7 processor running at 3.6 GHz, 16 GB of RAM. Our proposed heuristic method is approximately 40 times faster than the alternative algorithms. The elapsed time is measured in Julia using the @btime macro of the BenchmarkTools package [25].

It is important to note that the cost analysis presented here does not include the computational time required for training the hyperparameters, which can be considerable. In this regard, an additional benefit of the proposed algorithm is its

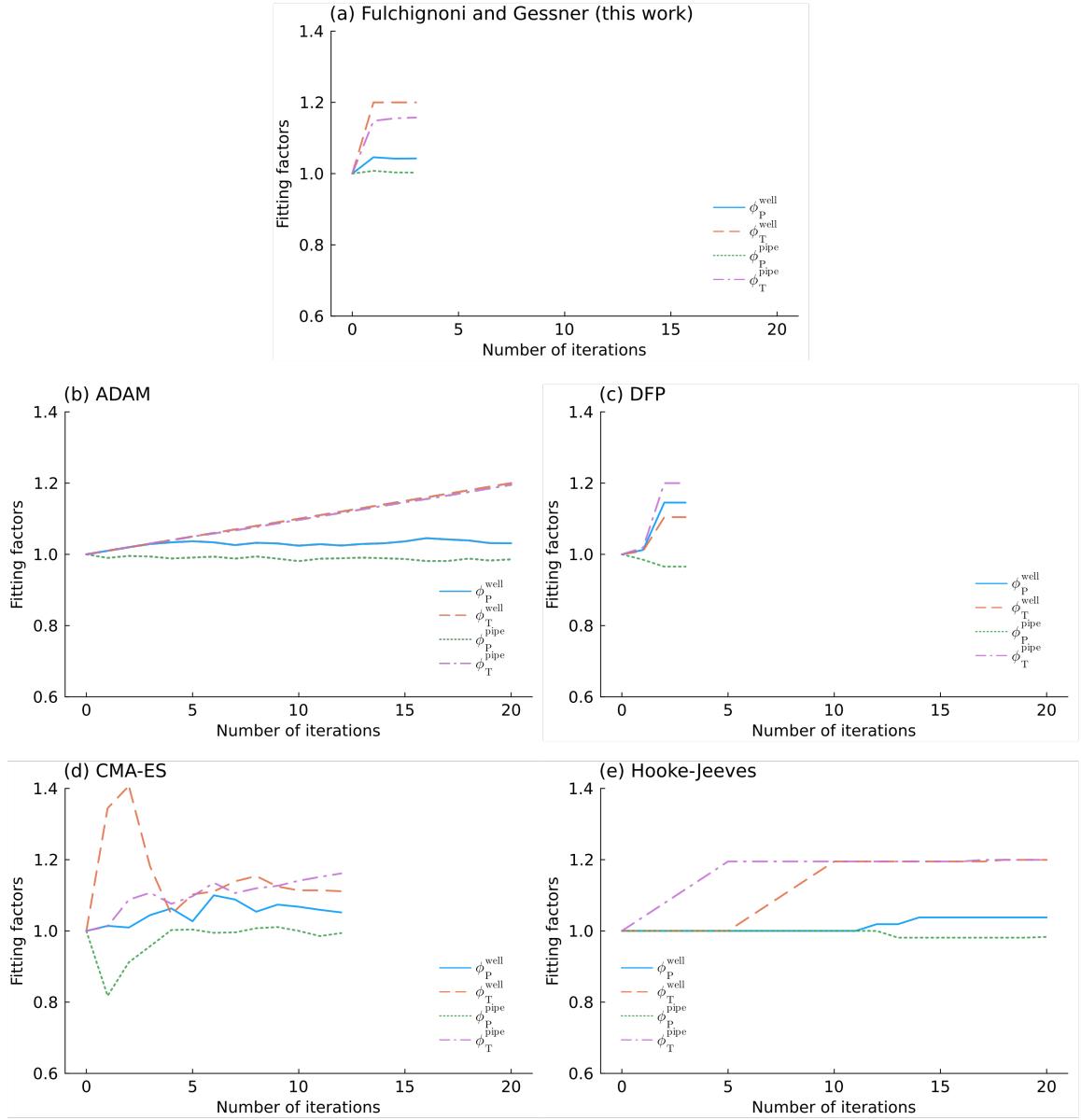


Figure 6.2: Convergence history of the tuning factors ($\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}},$ and ϕ_T^{pipe}) using: (a) proposed heuristic method (this work), (b) adaptive moment estimation (ADAM) [64], (c) Davidon-Fletcher-Powell method (DFP) [33, 40], (d) covariance matrix adaptation evolution strategy (CMA-ES) [55], and (e) Hooke-Jeeves (HJ) [57].

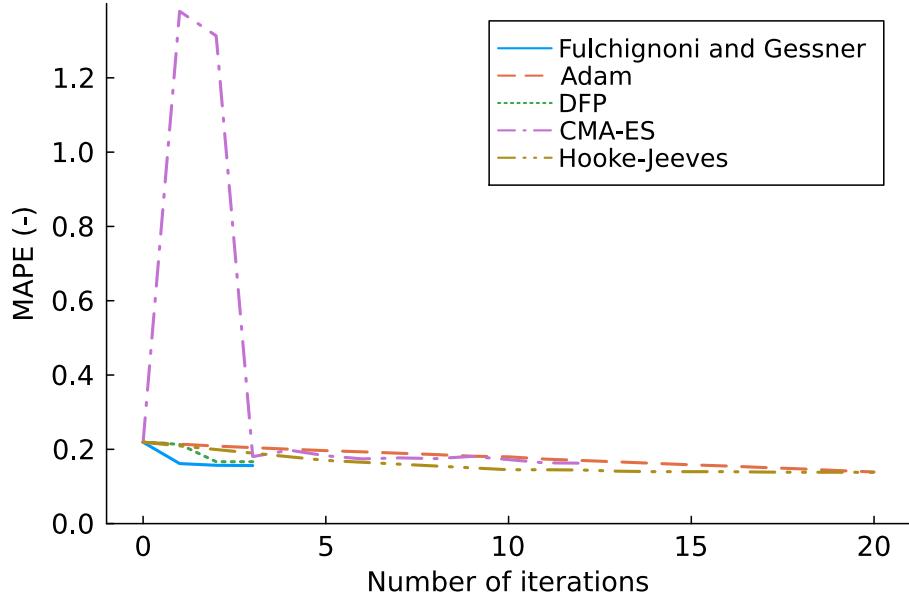


Figure 6.3: Convergence history of the Mean Absolute Percentage Error (MAPE) using the proposed heuristic method (this work), adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and Hooke-Jeeves (HJ) [57].

Table 6.3: Comparison of computational time (normalized) for the tuning factors regression performed with different optimization algorithms: ADAM, DFP, CMA-ES, HJ, and proposed heuristic algorithm.

	Optimization algorithm				
	Proposed heuristic (this work)		ADAM	DFP	CMA-ES
	Normalized computational time	1	43.8	43.2	37.8
					39.3

lack of hyperparameters, further simplifying the optimization process.

In summary, the heuristic demonstrates a competitive prediction MAPE in comparison to alternative optimization algorithms while being an order of magnitude faster. When contrasted with Hooke-Jeeves (the alternative algorithm yielding the smallest error), the heuristic method is 37.8 times faster and attains a MAPE that is 13% higher.

We further explore a two-step optimization approach, wherein the heuristic solution serves as an informed initial guess for the generic optimization algorithms. This strategy can be employed when a more refined optimization is necessary. Figure 6.4 illustrates the convergence history of the ADAM, DFP, CMA-ES, and HJ algorithms when utilizing the heuristic solution as the initial guess. Notably, the DFP algorithm is unable to further decrease the MAPE, resulting in the optimal point being the initial guess itself. Figure 6.5 displays the MAPE at each iteration. Although the HJ algorithm manages to further reduce the MAPE, the achieved reduction is not substantial.

6.5 Conclusions

We propose an efficient, derivative-free heuristic algorithm specifically tailored for adjusting tuning factors applied to pressure and temperature gradients predicted by multiphase pipe-flow simulators. As the algorithm is designed for this particular problem, it provides a competitive approximate solution at a significantly reduced computational cost compared to generic optimization methods. By applying our heuristic algorithm to the calibration of a real offshore oil production well, we demonstrated its effectiveness and efficiency when compared to four alternative optimization algorithms with different optimization strategies, namely: adaptive moment estimation method, Davidon-Fletcher-Powell method, covariance matrix adaptation evolution strategy method, and Hooke-Jeeves algorithm.

The proposed heuristic method not only achieves a highly competitive optimized

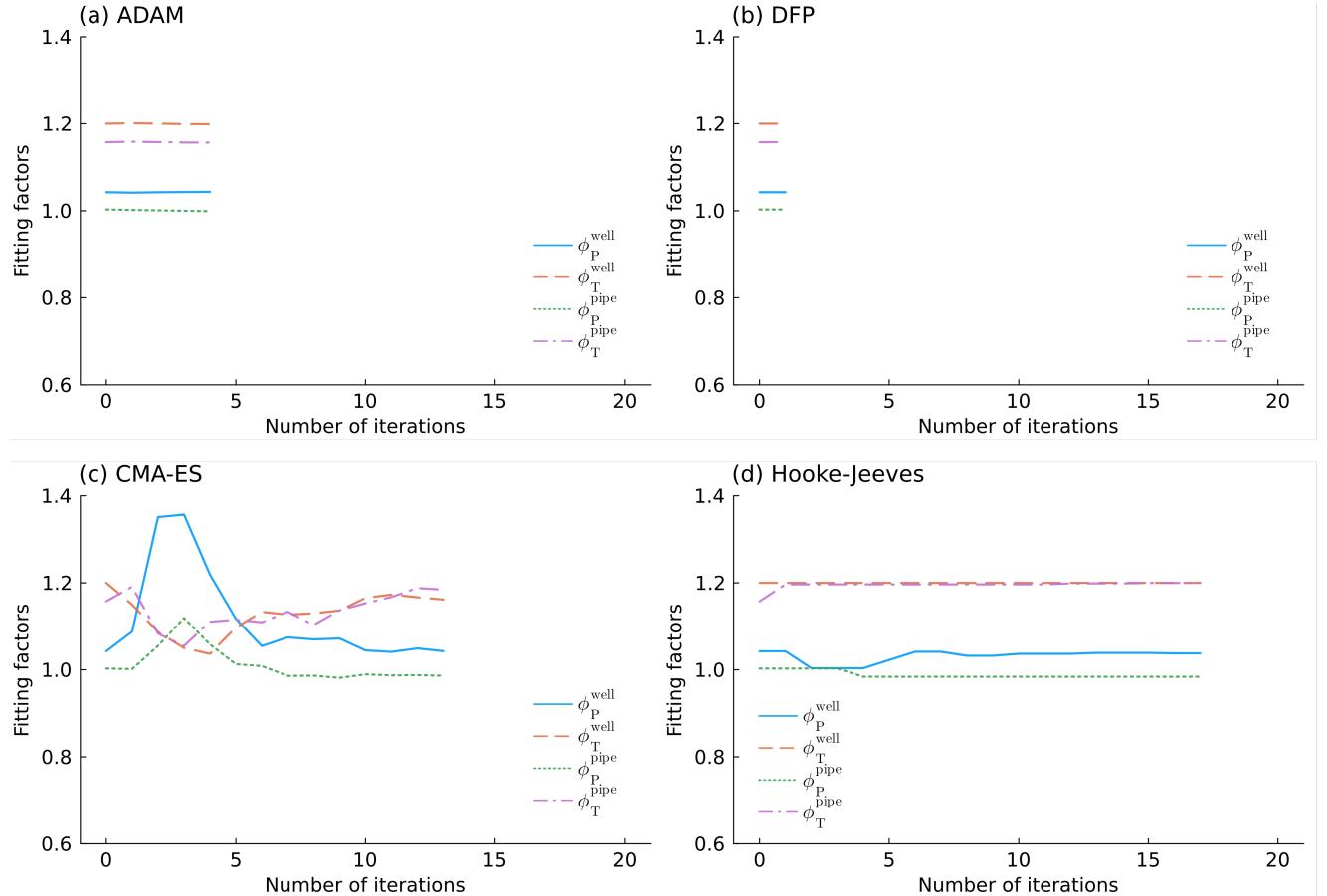


Figure 6.4: Convergence history of the tuning factors ($\phi_P^{\text{well}}, \phi_P^{\text{pipe}}, \phi_T^{\text{well}},$ and ϕ_T^{pipe}) using the proposed heuristic for the initial guess of four optimization algorithms: (a) adaptive moment estimation (ADAM) [64], (b) Davidon-Fletcher-Powell method (DFP) [33, 40], (c) covariance matrix adaptation evolution strategy (CMA-ES) [55], and (d) Hooke-Jeeves (HJ) [57].

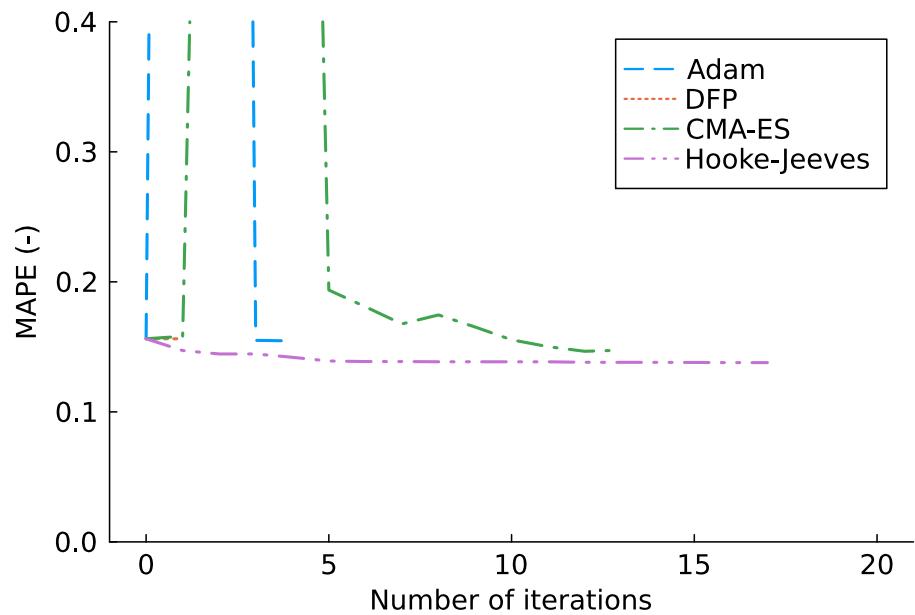


Figure 6.5: Convergence history of the Mean Absolute Percentage Error (MAPE) using the proposed heuristic for the initial guess of four optimization algorithms: adaptive moment estimation (ADAM) [64], Davidon-Fletcher-Powell method (DFP) [33, 40], covariance matrix adaptation evolution strategy (CMA-ES) [55], and Hooke-Jeeves (HJ) [57].

outcome but also accomplishes this task approximately 40 times faster than the alternative algorithms. Moreover, the algorithm does not rely on derivatives or hyperparameters, which simplifies the optimization process and further enhances its computational efficiency. We also explore a two-step optimization approach, wherein the heuristic solution serves as an informed initial guess for the generic optimization algorithms.

Accurate prediction of pipe flow behavior across the diverse conditions encountered in oil and gas fields is a complex task. Often, tuning factors must be applied to flow models to improve their accuracy and predictability by aligning them with available measurements. By incorporating a history of production tests during model calibration, the refined model is expected to better represent the underlying physics of the flow problem. The proposed heuristic algorithm offers a practical and computationally efficient means to achieve these improvements, enabling enhanced design, operation, and optimization of production/injection systems in the energy industry.

Chapter 7

Overall Conclusions and Future Work

7.1 Conclusions and Discussion

This dissertation examined uncertainties associated with complex fluid models through various perspectives. In particular, we scrutinized uncertainties arising from model selection, uncertain model inputs (i.e., parametric uncertainty), and uncertain models (i.e., structural uncertainty).

In Chapter 3, we explored the uncertainty in fluid model selection, focusing on empirical models for the solution gas-oil ratio. We established that existing models for this property are insufficient in accurately predicting specific reservoir fluid compositions. Consequently, we proposed a generalized model for the gas-oil solution ratio of hydrocarbon reservoir fluid that exhibits superior performance across our database.

In Chapter 4, we showed the influence of subjective choices, such as optimization algorithms and initial guesses used in the equation of state regression, on resulting predictions of the reservoir fluid's thermophysical behavior. We suggested a probabilistic approach to treating the heaviest hydrocarbon properties to quantify the predictive uncertainty of the resulting fluid models.

In Chapter 5, we examined the consequences of uncertainties in input parameters

to fluid models on crucial flow predictions, specifically focusing on cumulative production. We introduced a comprehensive framework for quantifying the predictive uncertainty of multiphase pipe-flow models, accounting for correlated random inputs.

In Chapter 6, we addressed pipe-flow model calibration to field data using tuning factors, a common technique in the oil and gas industry to enhance the accuracy and predictability of such models. We presented an affordable, efficient, and derivative-free heuristic algorithm specifically tailored for adjusting tuning factors applied to pressure and temperature gradients predicted by multiphase pipe-flow simulators.

This research underscores the significance of quantifying predictive uncertainties in fluid models. We hope that our findings will contribute to an improved understanding and management of uncertainties in reservoir fluid modeling. A thorough characterization and understanding of predictive uncertainty in fluid models hold significant implications for enhancing the accuracy and reliability of reservoir and flow assurance engineering simulations, ultimately fostering efficient and sustainable management of production and injection systems.

7.2 Future work

While specific avenues for future work have been outlined in each chapter, there are several general directions that can further enhance the probabilistic characterization of flow and fluid model predictions. Potential future work includes:

- Considering the variability in predictions offered by different empirical fluid models, as observed in Chapter 3, the exploration of Bayesian Model Averaging (BMA) and Bayesian Model Selection (BMS) methods could provide valuable insights.
- Further investigating the impact of uncertainties in fluid model selection on reservoir and pipe-flow simulations can lead to a better understanding of their implications.
- Evaluating the proposed empirical model for the solution gas-oil ratio property using an extended database will help in validating its effectiveness.

- Exploring variations in the equation of state regression process, such as incorporating the C_{plus} fraction expansion and lumping procedures, and examining their impact on the fluid model's predictive uncertainty.
- Developing a framework to characterize the joint distribution of unknown fluid model parameters, specifically tailored for the production system under investigation, can improve prediction accuracy.
- For characterizing the uncertainty in cumulative production, incorporating a wider range of unknown inputs, from reservoir descriptions to the fluid models selection, can offer a more comprehensive understanding.
- Integrating reservoir and pipe-flow models into the analysis of probabilistic forecasting for cumulative production of reservoir fluid with uncertain properties can provide more holistic insights into the system's behavior.
- Assessing the influence of equation of state predictive uncertainty on the uncertainty in porous media and pipe-flow simulation outcomes.

Appendix A

Implemented R_s BO models

The nine R_s models [7, 9, 11, 48, 68, 98, 118, 128, 129] implemented in Chapter 3 are reported below. In the upcoming equations, following this thesis notation unless stated otherwise, R_s is the solution gas-oil ratio in scf/STB, R_{sb} is the reservoir fluid gas-oil ratio in scf/STB, γ_g is the gas specific gravity, γ_{API} is the oil API gravity, P is the pressure in psia, and T is the temperature in °F.

Al-Marhoun [7]:

$$R_s = \frac{P^{1.398441} \gamma_g^{2.626048} (\gamma_{API} + 131.5)^{4.396279}}{1.914510 \times 10^6 (T + 460)^{1.855130}} \quad (\text{A.1})$$

Al-Shammai [9]:

$$R_s = \frac{P^{1.275972} (\gamma_{API} + 131.5)^{7.052572}}{1.473546 \times 10^{15} \gamma_g (T + 460)} \exp\left(\frac{332.4663 \gamma_g}{\gamma_{API} + 131.5}\right) \quad (\text{A.2})$$

Alakbari et al. [11]: R_s is obtained by solving the nonlinear system of equations (Eqs. A.3 to A.8).

$$\begin{aligned} A_1 = & 4.36 \times 10^3 - 88.1\gamma_{\text{API}} - 5.1 \times 10^3\gamma_g + 8.37R_s + 71.5\gamma_g\gamma_{\text{API}} \\ & - 0.0857R_s\gamma_{\text{API}} - 1.74R_s\gamma_g + 0.359\gamma_{\text{API}}^2 + 969\gamma_g^2 - 0.000533R_s^2 \end{aligned} \quad (\text{A.3})$$

$$\begin{aligned} A_2 = & 933 + 1.2T - 43.4\gamma_{\text{API}} + 5.83R_s + 0.113\gamma_{\text{API}}T + 0.00387R_sT \\ & - 0.0706R_s\gamma_{\text{API}} - 0.0112T^2 + 0.207\gamma_{\text{API}}^2 - 0.000792R_s^2 \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} A_3 = & 929 - 34.6\gamma_{\text{API}} + 6.55R_s - 0.0719R_s\gamma_{\text{API}} + 0.371\gamma_{\text{API}}^2 \\ & - 0.000648R_s^2 \end{aligned} \quad (\text{A.5})$$

$$\begin{aligned} A_4 = & -385 + 0.581A_1 + 5.21T - 0.0217R_s - 0.00132TA_1 - 0.00173R_sA_1 \\ & + 0.00848R_sT + 0.000405A_1^2 - 0.00989T^2 + 0.0016R_s^2 \end{aligned} \quad (\text{A.6})$$

$$\begin{aligned} A_5 = & -51.4 - 0.253A_3 + 0.773A_2 + 0.526A_1 + 0.000949A_2A_3 \\ & - 9.93 \times 10^{-5}A_1A_3 - 0.000614A_1A_2 - 0.000625A_3^2 \\ & - 8.53 \times 10^{-5}A_2^2 + 0.000489A_1^2 \end{aligned} \quad (\text{A.7})$$

$$\begin{aligned} P = & -66.1 + 0.183A_5 + 0.82A_4 + 0.721T - 0.00287A_4A_5 - 0.00307TA_5 \\ & + 0.00334TA_4 + 0.00157A_5^2 + 0.00129A_4^2 - 0.00326T^2 + 14.7 \end{aligned} \quad (\text{A.8})$$

Glasø [48]:

$$R_s = \gamma_g \left(\frac{p^* \gamma_{\text{API}}^{0.989}}{T^{0.172}} \right)^{1.2255} \quad (\text{A.9})$$

where p^* is given by Eq. (A.10), in which p_{corr} (Eq. A.11) is the corrected pressure through the correction factors f_{CO_2} , f_{N_2} , $f_{\text{H}_2\text{S}}$ (Eqs. A.12-A.14) to account for the presence of CO₂, N₂ and H₂S, respectively. The CO₂, N₂ and H₂S molar fractions in

the surface gas are y_{CO_2} , y_{N_2} , and $y_{\text{H}_2\text{S}}$, respectively.

$$p^* = 10^{2.8869 - 1.6546 \sqrt{5.1797 - 1.2087 \log(p_{corr})}} \quad (\text{A.10})$$

$$p_{corr} = \frac{p}{f_{\text{CO}_2} f_{\text{N}_2} f_{\text{H}_2\text{S}}} \quad (\text{A.11})$$

$$f_{\text{CO}_2} = 1.0 - 693.8 y_{\text{CO}_2} T^{-1.553} \quad (\text{A.12})$$

$$\begin{aligned} f_{\text{N}_2} = & 1.0 + [(-2.65 \times 10^{-4} \gamma_{\text{API}} + 5.5 \times 10^{-3})T + 0.0931 \gamma_{\text{API}} - 0.8295] y_{\text{N}_2} \\ & + (1.954 \times 10^{-11} \gamma_{\text{API}}^{4.699} T + 0.027 \gamma_{\text{API}} - 2.366) y_{\text{N}_2}^2 \end{aligned} \quad (\text{A.13})$$

$$f_{\text{H}_2\text{S}} = 1.0 - (0.9035 + 0.0015 \gamma_{\text{API}}) y_{\text{H}_2\text{S}} + 0.019(45 - \gamma_{\text{API}}) y_{\text{H}_2\text{S}}^2 \quad (\text{A.14})$$

Glaso proposes correcting the flash γ_{API} for paraffinicity. However, this correction worsened the results in 91% of the cases. Therefore, it is not included here.

Lasater [68]:

$$R_s = \frac{1.878\,483\,25 \times 10^7 y_g}{(1 - y_g)(\gamma_{\text{API}} + 131.5) M_o} \quad (\text{A.15})$$

$$p_f = \frac{P \gamma_g}{T + 459.6} \quad (\text{A.16})$$

where y_g is obtained from p_f (Eq. A.16) and the chart presented in Fig. 2 of Lasater's [68] original paper. We use Whitson and Brûlé's [133] curve fitting (Eq. A.17) for the original chart.

$$p_f = \begin{cases} 0.83918 \times 10^{1.17664 y_g} y_g^{0.57246} & y_g \leq 0.6 \\ 0.83918 \times 10^{1.08000 y_g} y_g^{0.31109} & y_g > 0.6 \end{cases} \quad (\text{A.17})$$

The original relationship between γ_{API} and M_o is also graphical and is shown in Figure A.1 (black curve). For the sake of completion, we show three alternative curve fittings of Lasater's original chart in Figure A.1. The fittings proposed by Almeida [14] and Cragoe [31] are given in Eqs. (A.18) and (A.19) respectively. The last is recommended by Whitson and Brûlé [133]. The commercial flow simulator PIPESIM® version 2015 calculates M_o through Eq (A.20).

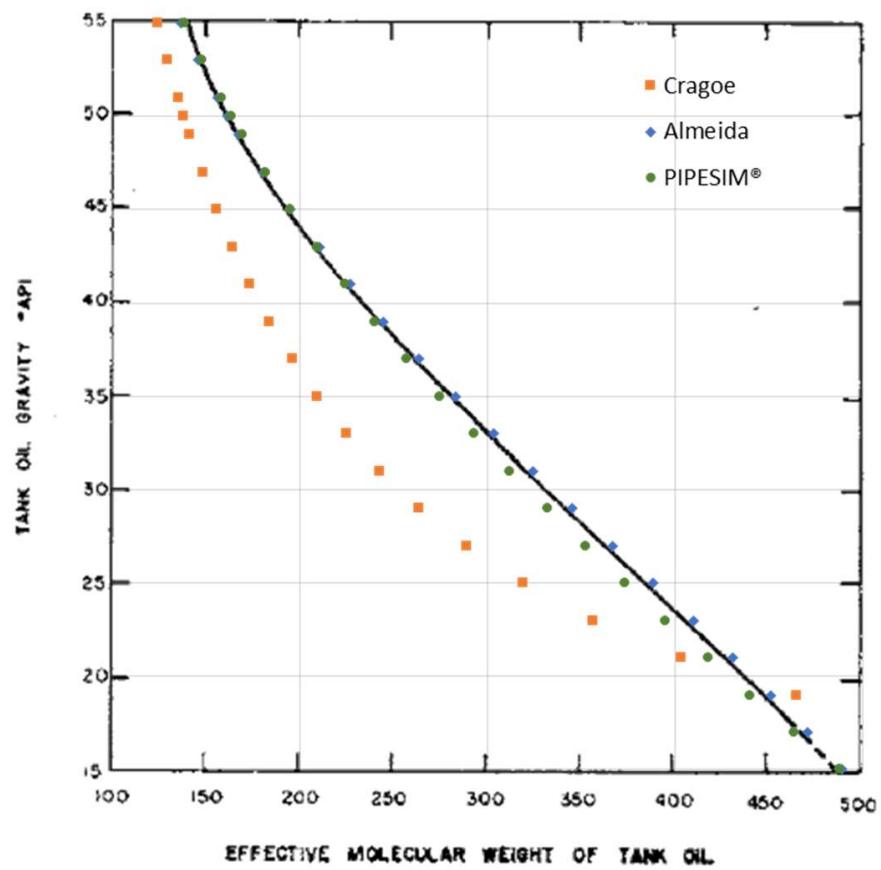


Figure A.1: Adaptation of Lasater's [68] original chart for the relationship between the stock-tank oil gravity ($^{\circ}$ API) and the effective molecular weight (M_o) superposed by the curve fittings proposed by Cragoe [31], Almeida [14], and Pipesim[®] v.2015

$$M_o = 558.16 + 1.1025\gamma_{\text{API}} - 0.50033\gamma_{\text{API}}^2 + 8.3259 \times 10^{-3}\gamma_{\text{API}}^3 - 3.8688 \times 10^{-5}\gamma_{\text{API}}^4 \quad (\text{A.18})$$

$$M_o = \frac{6084}{\gamma_{\text{API}} - 5.9} \quad (\text{A.19})$$

$$M_o = 677.3893 - 13.2161 \circ API + 0.024775 \circ API^2 + 0.00067851 \circ API^3 \quad (\text{A.20})$$

It is clear that Almeida gives a better representation of the original Lasater's chart. Moreover, Cragoes's equation gives unrealistic M_o values for low values of γ_{API} . Al-Shammasi [9] also presents fitting equations, but they are less precise than Almeida's since they involve 2nd degree polynomials.

In this dissertation, M_o follows Eq. (A.18), and y_g is obtained by solving Eq. (A.17) [133], in which p_f is given by Eq. (A.16).

Petrosky and Farshad [98]:

$$R_s = \left[\left(\frac{P}{112.727} + 12.340 \right) \gamma_g^{0.8439} 10^{7.916 \times 10^{-4} \gamma_{\text{API}}^{1.5410} - 4.561 \times 10^{-5} T^{1.3911}} \right]^{1.73184} \quad (\text{A.21})$$

Standing [118, 117]:

$$R_s = \gamma_g \left[\left(\frac{P}{18.2} + 1.4 \right) 10^{0.0125 \gamma_{\text{API}} - 0.00091 T} \right]^{1.2048} \quad (\text{A.22})$$

Vazquez and Beggs [128]:

$$R_s = C_1 d_{gs} P^{C_2} \exp \left(C_3 \frac{\gamma_{\text{API}}}{T + 460} \right) \quad (\text{A.23})$$

where C_1 , C_2 , and C_3 are constants, respectively equal to 0.0362, 1.0937, 25.7240 for $\text{API} \leq 30$; and to 0.0178, 1.1870, 23.9310 for $\text{API} > 30$. The term d_{gs} is the corrected gas gravity for separator conditions of 100 psig, given by Eq. (A.24) with $T_{ref} = 60^\circ\text{F}$ and $P_{ref} = 14.7$ psia.

$$d_{gs} = \gamma_g \left[1 + 5.912 \times 10^{-5} \gamma_{\text{API}} T_{ref} \log \left(\frac{p_{ref}}{114.7} \right) \right] \quad (\text{A.24})$$

Velarde et al. [129]:

$$R_s = \left[a_1 \left(\frac{P - 14.7}{P_b - 14.7} \right)^{a_2} + (1 - a_1) \left(\frac{P - 14.7}{P_b - 14.7} \right)^{a_3} \right] \text{GOR} \quad (\text{A.25})$$

$$a_1 = A_0 \gamma_g^{A_1} \gamma_{\text{API}}^{A_2} T^{A_3} (P_b - 14.7)^{A_4} \quad (\text{A.26})$$

$$a_2 = B_0 \gamma_g^{B_1} \gamma_{\text{API}}^{B_2} T^{B_3} (P_b - 14.7)^{B_4} \quad (\text{A.27})$$

$$a_3 = C_0 \gamma_g^{C_1} \gamma_{\text{API}}^{C_2} T^{C_3} (P_b - 14.7)^{C_4} \quad (\text{A.28})$$

where the coefficients ($A_0, \dots, A_4, B_0, \dots, B_4, C_0, \dots, C_4$) are given below and P_b is the bubble point pressure in psia. In the lack of measured data, P_b is estimated through Eq. (A.29), in which γ_{gs} is the gas gravity at (unspecified) separator conditions and considered in this study to be equal to γ_g .

$$\begin{array}{lll} A_0 = 9.73 \times 10^{-7} & B_0 = 0.022339 & C_0 = 0.725167 \\ A_1 = 1.672608 & B_1 = -1.004750 & C_1 = -1.485480 \\ A_2 = 0.929870 & B_2 = 0.337711 & C_2 = -0.164741 \\ A_3 = 0.247235 & B_3 = 0.132795 & C_3 = -0.091330 \\ A_4 = 1.056052 & B_4 = 0.302065 & C_4 = 0.047094 \end{array}$$

$$P_b = 1091.47 \left[R_{sb}^{0.081465} \gamma_{\text{gs}}^{-0.161488} 10^{0.013098T^{0.282372} - 8.2 \times 10^{-6}\gamma_{\text{API}}^{2.176124}} - 0.740152 \right]^{5.354891} \quad (\text{A.29})$$

Bibliography

- [1] Al-Gathe Abedelrigaab, Ahmed El-Banbi, Kh A Abdel Fattah, and KA El-Metwally. An Artificial Intelligent Approach for Black Oil PVT Properties. In *Journal of Physics: Conference Series*, volume 1962, page 012026. IOP Publishing, 2021.
- [2] Mark A Abramson, Charles Audet, John E Dennis Jr, and Sébastien Le Digabel. OrthoMADS: a deterministic MADS instance with orthogonal directions. *SIAM Journal on Optimization*, 20(2):948–966, 2009.
- [3] R K Agarwal, Y K Li, and L Nghiem. A regression technique with dynamic parameter selection for phase-behavior matching. *SPE Reservoir Engineering*, 1:115–120, 1990.
- [4] Rafael A Aguilar Zurita and William D McCain. An efficient tuning strategy to calibrate cubic EOS for compositional simulation. In *SPE Annual Technical Conference and Exhibition*, pages SPE-77382-MS. OnePetro, 2002.
- [5] Tarek Ahmed. *Equations of state and PVT analysis*. Elsevier, 2013.
- [6] Tarek Ahmed. *Reservoir engineering handbook*. Gulf professional publishing, 2018.
- [7] Muhammad Ali Al-Marhoun. PVT correlations for Middle East crude oils. *Journal of Petroleum Technology*, 40(05):650–666, 1988.
- [8] Ali Abdallah Al-Meshari. *New strategic method to tune equation-of-state to*

- match experimental data for compositional simulation.* Texas A&M University, 2004.
- [9] AA Al-Shammasi. A review of bubblepoint pressure and oil formation volume factor correlations. *SPE Reservoir Evaluation & Engineering*, 4(02):146–160, 2001.
 - [10] Fahd Saeed Alakbari, Mysara Eissa Mohyaldinn, Mohammed Abdalla Ayoub, and Ali Samer Muhsan. Deep learning approach for robust prediction of reservoir bubble point pressure. *ACS Omega*, 6(33):21499–21513, 2021.
 - [11] Fahd Saeed Alakbari, Mysara Eissa Mohyaldinn, Mohammed Abdalla Ayoub, Ali Samer Muhsan, and Ibnelwaleed A Hussein. An Accurate Reservoir’s Bubble Point Pressure Correlation. *ACS omega*, 2022.
 - [12] A. Alawadhi, F. Boso, and D. M. Tartakovsky. Method of distributions for water-hammer equations with uncertain parameters. *Water Resources Research*, 54(11):9398–9411, 2018.
 - [13] Mahmoud T Ali and Ahmed H El-Banbi. EOS tuning - comparison between several valid approaches and new recommendations. In *SPE North Africa Technical Conference and Exhibition*, pages SPE-175877-MS. OnePetro, 2015.
 - [14] Alcino R Almeida. Escoamento e transferência de calor em produção de petróleo. Rio de Janeiro, RJ: Petrobras, 2013.
 - [15] Sepehr Arbabi and Abbas Firoozabadi. Near-critical phase behavior of reservoir fluids using equations of state. *SPE Advanced Technology Series*, 3(01):139–145, 1995.
 - [16] Khalid Aziz and Antonin Settari. Petroleum reservoir simulation. 1979.
 - [17] Dale H Beggs and James P Brill. A study of two-phase flow in inclined pipes. *Journal of Petroleum technology*, 25(05):607–617, 1973.

- [18] H Dale Beggs and John R Robinson. Estimating the viscosity of crude oil systems. *Journal of Petroleum Technology*, 27(09):1140–1141, 1975.
- [19] Kjell H Bendiksen, Dag Maines, Randi Moe, and Sven Nuland. The dynamic two-fluid model OLGA: Theory and application. *SPE production engineering*, 6(02):171–180, 1991.
- [20] J Eric Bickel and Reidar B Bratvold. From uncertainty quantification to decision making in the oil and gas industry. *Energy Exploration & Exploitation*, 26(5):311–325, 2008.
- [21] J P Brill and S J Arirachakaran. State of the art in multiphase flow. *Journal of Petroleum Technology*, 44(05):538–541, 1992.
- [22] Mai Bui, Claire S Adjiman, André Bardow, Edward J Anthony, Andy Boston, Solomon Brown, Paul S Fennell, Sabine Fuss, Amparo Galindo, Leigh A Hackett, et al. Carbon capture and storage (CCS): the way forward. *Energy & Environmental Science*, 11(5):1062–1176, 2018.
- [23] B Carrier, M Rogalski, and A Peneloux. Choice of pseudocomponents for flash calculations of petroleum fluids. Technical report, New York, NY; American Institute of Chemical Engineers, 1988.
- [24] Gabriela Chaves, Danielle Monteiro, and Virgilio José Martins Ferreira. A Back Allocation Methodology to Estimate the Real-Time Flow and Assist Production Monitoring. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2021.
- [25] Jiahao Chen and Jarrett Revels. Robust benchmarking in noisy environments, 2016.
- [26] Jinho Choi, Eduardo Pereyra, Cem Sarica, Hoyoung Lee, Il Sik Jang, and JooMyoung Kang. Development of a fast transient simulator for gas–liquid two-phase flow in pipes. *Journal of Petroleum science and engineering*, 102:27–35, 2013.

- [27] PL Christensen. Regression to experimental PVT data. *Journal of Canadian Petroleum Technology*, 38(13), 1999.
- [28] V. Ciriello, I. Lauriola, S. Bonvicini, V. Cozzani, V. Di Federico, and D. M. Tartakovsky. Impact of hydrogeological uncertainty on estimation of environmental risks posed by hydrocarbon transportation networks. *Water Resour. Res.*, 53(11):8686–8697, 2017.
- [29] V. Ciriello, I. Lauriola, and D. M. Tartakovsky. Distribution-based global sensitivity analysis in hydrology. *Water Resour. Res.*, 55(11):8708–8720, 2019.
- [30] K H Coats and G T Smart. Application of a regression-based EOS PVT program to laboratory data. *SPE Reservoir Engineering*, 1(3):277–299, 1986.
- [31] Carl Susan Cragoe. *Thermal Properties of Petroleum Products*. Number 97. US Government Printing Office, 1929.
- [32] Yaghoub Dastkhan, Alireza Kazemi, and Rashid S Al-Maamari. Well and reservoir characterization using combined temperature and pressure transient analysis: A case study. *Results in Engineering*, 15:100488, 2022.
- [33] W C Davidon. Variable metric method for minimization. Technical report, Argonne National Laboratory, 1959.
- [34] Giambattista De Ghetto and Marco Villa. Reliability analysis on pvt correlations. In *European Petroleum Conference*. OnePetro, 1994.
- [35] Deepstar. Flow Assurance Design Guideline, 2001. Deepstar IV Project, DSIV CTR 4203b–1.
- [36] DeepStar. DeepStar IV Project Report, 2011.
- [37] Mahmoud E Dokla and Mohammad E Osman. Correlation of PVT properties for UAE crudes. *SPE formation evaluation*, 7(01):41–46, 1992.

- [38] P.M. Dranchuk, R.A. Purvis, and D.B. Robinson. Computer Calculation Of Natural Gas Compressibility Factors Using The Standing And Katz Correlation. *PETSOC Annual Technical Meeting*, 1973.
- [39] Agoston E Eiben, James E Smith, et al. *Introduction to evolutionary computing*, volume 53. Springer, 2003.
- [40] Roger Fletcher and Michael JD Powell. A rapidly convergent descent method for minimization. *The Computer Journal*, 6(2):163–168, 1963.
- [41] Lívia Paiva Fulchignoni, Alcino R Almeida, and Roberto da Fonseca Jr. A novel model for the solution gas-oil ratio suitable for CO₂-rich reservoir fluids. *Results in Engineering*, 16(100681), 2022.
- [42] Lívia Paiva Fulchignoni, Alcino R Almeida, and Roberto Fonseca Jr. Evaluation of the solution gas-oil ratio correlation proposed by Lasater applied to oils with high CO₂ and high GOR. *Brazilian Journal of Petroleum and Gas*, 15(3), 2021.
- [43] Lívia Paiva Fulchignoni and Roberto Fonseca Jr. Influence of the solution gas-oil Ratio (R_s) black oil modeling on the multiphase flow modeling for an Oil with high CO₂ content. 10th International Conference on Multiphase Flow, 2019.
- [44] Lívia Paiva Fulchignoni, Christiano Garcia da Silva Santim, and Daniel M. Tartakovsky. Probabilistic forecasting of cumulative production of reservoir fluid with uncertain properties. *Geoenergy Science and Engineering*, 227:211819, 2023.
- [45] Lívia Paiva Fulchignoni, Christiano Garcia da Silva Santim, and Daniel M. Tartakovsky. Probabilistic forecasting of cumulative production of reservoir fluid with uncertain properties. *Geoenergy Science and Engineering*, page 211819, 2023.
- [46] Lívia Paiva Fulchignoni and Daniel M Tartakovsky. Uncertain characterization of reservoir fluids due to brittleness of equation of state regression. *Geoenergy Science and Engineering Journal*, 2023.

- [47] Hamzeh Ghorbani, David A Wood, Abouzar Choubineh, Nima Mohamadian, Afshin Tatar, Hamed Farhangian, and Ali Nikooey. Performance comparison of bubble point pressure from oil PVT data: Several neurocomputing techniques compared. *Experimental and Computational Multiphase Flow*, 2(4):225–246, 2020.
- [48] Oistein Glasø. Generalized pressure-volume-temperature correlations. *Journal of Petroleum Technology*, 32(05):785–795, 1980.
- [49] S Gomaa. New bubble point pressure correlation for middle east crude oils. *Int. Adv. Res. J. Sci. Eng. Tech*, 3(12):1–9, 2016.
- [50] S. Gourari, Fateh Mebarek-Oudina, Oluwole Daniel Makinde, and M. Rabhi. Numerical Investigation of Gas-Liquid Two-Phase Flows in a Cylindrical Channel. *Defect and Diffusion Forum*, 409:39–48, 6 2021.
- [51] Miroslav Grmela. Multiscale equilibrium and nonequilibrium thermodynamics in chemical engineering. In *Advances in Chemical Engineering*, volume 39, pages 75–129. Elsevier, 2010.
- [52] Teklu Hadgu, Robert W Zimmerman, and Gudmundur S Bodvarsson. Coupling of a reservoir simulator and a wellbore simulator for geothermal applications. 1993.
- [53] Morten Hammer, Han Deng, Lan Liu, Morten Langsholt, and Svend Tollak Munkejord. Upward and downward two-phase flow of CO₂ in a pipe: Comparison between experimental data and model predictions. *International Journal of Multiphase Flow*, 138:103590, 2021.
- [54] HH Hanafy, SM Macary, YM ElNady, AA Bayomi, and MH El Batanony. Empirical PVT correlations applied to Egyptian crude oils exemplify significance of using regional correlations. In *International Symposium on Oilfield Chemistry*. OnePetro, 1997.

- [55] Nikolaus Hansen. The CMA evolution strategy: a comparing review. *Towards a New Evolutionary Computation*, pages 75–102, 2006.
- [56] Mohamad Heidarian, Masoud Karimnezhad, Mahin Schaffie, and Mohammad Ranjbar. A new empirical correlation for estimating bubble point pressure using the genetic algorithm. *Geology, Geophysics and Environment*, 43(1):33–33, 2017.
- [57] Robert Hooke and Terry A Jeeves. “Direct Search” Solution of Numerical and Statistical Problems. *Journal of the ACM (JACM)*, 8(2):212–229, 1961.
- [58] International Energy Agency. *Energy Technology Perspectives 2020 - Special Report on Carbon Capture Utilisation and Storage*. 2020.
- [59] Mamoru Ishii. One-dimensional drift-flux model and constitutive equations for relative motion between phases in various two-phase flow regimes. Technical report, Argonne National Lab., Ill.(USA), 1977.
- [60] HA Jacobson. The effect of nitrogen on reservoir fluid saturation pressure. *Journal of Canadian Petroleum Technology*, 6(03):101–105, 1967.
- [61] I e Jahanandish, B Salimifard, and H Jalalifar. Predicting bottomhole pressure in vertical multiphase flowing wells using artificial neural networks. *Journal of Petroleum Science and Engineering*, 75(3-4):336–342, 2011.
- [62] Bharat S Jhaveri and Gary K Youngren. Three-parameter modification of the Peng-Robinson equation of state to improve volumetric predictions. *SPE Reservoir Engineering*, 3(3):1033–1040, 1988.
- [63] Michael G Kesler and Byung Ik Lee. Improve prediction of enthalpy of fractions. *Hydrocarbon Processing*, 55:153–158, 1976.
- [64] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

- [65] Julia Klinkert. The characterization of uncertainty for steady state multiphase flow models in pipelines. 2018.
- [66] Georgios M Kontogeorgis, Xiaodong Liang, Alay Arya, and Ioannis Tsivintzelis. Equations of state in three centuries. are we closer to arriving to a single model for all applications? *Chemical Engineering Science: X*, page 100060, 2020.
- [67] Anthony R. Kovscek. *Thermodynamics of petroleum reservoir and storage formations: an equation of state approach*. Stanford University, 2021.
- [68] JA Lasater. Bubble point pressure correlation. *Journal of Petroleum Technology*, 10(05):65–67, 1958.
- [69] Anthony L Lee, Mario H Gonzalez, and Bertram E Eakin. The viscosity of natural gases. *Journal of petroleum technology*, 18(08):997–1000, 1966.
- [70] Huazhou Li and Daoyong Yang. Phase behaviour of C₃H₈/n-C₄H₁₀/heavy-oil systems at high pressures and elevated temperatures. *Journal of Canadian Petroleum Technology*, 52(01):30–40, 2013.
- [71] Alan G Marshall and Ryan P Rodgers. Petroleomics: the next grand challenge for chemical analysis. *Accounts of chemical research*, 37(1):53–59, 2004.
- [72] Emma Martin-Roberts, Vivian Scott, Stephanie Flude, Gareth Johnson, R Stuart Haszeldine, and Stuart Gilfillan. Carbon capture and storage at the end of a lost decade. *One Earth*, 4(11):1569–1584, 2021.
- [73] W.D. McCain, J.P. Spivey, and C.P. Lenn. *Petroleum Reservoir Fluid Property Correlations*. PennWell Corporation, 2011.
- [74] William D McCain. Analysis of black oil PVT reports revisited. In *SPE Annual Technical Conference and Exhibition*. OnePetro, 2002.
- [75] KK Meisingset. Uncertainties in reservoir fluid description for reservoir modeling. *SPE Reservoir Evaluation & Engineering*, 2(05):431–435, 1999.
- [76] M Mesbah and Alireza Bahadori. Tuning equations of state. 2017.

- [77] Michael L Michelsen. Saturation point calculations. *Fluid Phase Equilibria*, 23(2-3):181–192, 1985.
- [78] D O Monteiro, M C M A Duque, Gabriela S Chaves, Juliana S Baioco, and Virgílio JM Ferreira Filho. Model Adjusting and Parameters Estimation under Uncertainties for Multiphase Flow. In *Rio Oil & Gas Expo and Conference*, 2018.
- [79] Danielle D Monteiro, Maria Machado Duque, Gabriela S Chaves, Virgílio M Ferreira Filho, and Juliana S Baioco. Using data analytics to quantify the impact of production test uncertainty on oil flow rate forecast. *Oil & Gas Science and Technology–Revue d'IFP Energies nouvelles*, 75:7, 2020.
- [80] Dan Vladimir Nichita. Phase envelope construction for mixtures with many components. *Energy & fuels*, 22(1):488–495, 2008.
- [81] M C K Oliveira, Lívia Fulchignoni Paiva, F A Meireles, R Mendes, P M D Silva, and L G Machado. Subsea Demulsifier Injection to Reduce Emulsion Viscosity and Enhance Crude Oil Production. In *Offshore Technology Conference*, 2017.
- [82] Dean S Oliver and Yan Chen. Recent progress on reservoir history matching: a review. *Computational Geosciences*, 15:185–221, 2011.
- [83] MI Omar and AC Todd. Development of new modified black oil correlations for Malaysian crudes. In *SPE Asia Pacific oil and gas conference*. OnePetro, 1993.
- [84] R D Ostermann, C A Ehlig-Economides, and O O Owolabi. Correlations for the reservoir fluid properties of Alaskan crudes. In *SPE California regional meeting*. OnePetro, 1983.
- [85] Art B. Owen. *Monte Carlo theory, methods and examples*. 2013.
- [86] Charles Martin Palmer. Evaluation of inclined pipe two-phase liquid holdup correlations using experimental data. Master's thesis, University of Tulsa, 1975.

- [87] Karen Schou Pedersen, Ann Lisbeth Blilie, and Knut Kristian Meisingset. PVT calculations on petroleum reservoir fluids using measured and estimated compositional data for the plus fraction. *Industrial & Engineering Chemistry Research*, 31(5):1378–1384, 1992.
- [88] Karen Schou Pedersen, Peter Lindskou Christensen, and Jawad Azeem Shaikh. *Phase behavior of petroleum reservoir fluids*. CRC press, 2014.
- [89] Karen Schou Pedersen, Peter Lindskou Christensen, Jawad Azeem Shaikh, and Peter L Christensen. *Phase behavior of petroleum reservoir fluids*. CRC press, 2006.
- [90] Karen Schou Pedersen, Aage Fredenslund, and Per Thomassen. *Properties of oils and natural gases*, volume 5. Gulf Publishing Company, 1989.
- [91] Karen Schou Pedersen, Per Thomassen, and Aage Fredenslund. Thermodynamics of petroleum mixtures containing heavy hydrocarbons. 1. Phase envelope calculations by use of the Soave-Redlich-Kwong equation of state. *Industrial & Engineering Chemistry Process Design and Development*, 23(1):163–170, 1984.
- [92] Karen Schou Pedersen, Per Thomassen, and Aage Fredenslund. Thermodynamics of petroleum mixtures containing heavy hydrocarbons. 3. efficient flash calculation procedures using the srk equation of state. *Industrial & Engineering Chemistry Process Design and Development*, 24(4):948–954, 1985.
- [93] KS Pedersen, J Milter, and H Sørensen. Cubic equations of state applied to ht/hp and highly aromatic fluids. *SPE Journal*, 9(02):186–192, 2004.
- [94] José Luís Peña Díez, Daniel Merino García, Verónica Benito Iglesias, and Paula Sanz Sanz. Computer implemented method for characterizing a target fluid of a hydrocarbon reservoir under uncertainty, December 6 2022. US Patent 11,519,897.
- [95] André Péneloux, Evelyne Rauzy, and Richard Fréze. A consistent correction for Redlich-Kwong-Soave volumes. *Fluid Phase Equilibria*, 8(1):7–23, 1982.

- [96] Ding-Yu Peng and Donald B Robinson. A new two-constant equation of state. *Industrial & Engineering Chemistry Fundamentals*, 15(1):59–64, 1976.
- [97] Petrobras. COMUNICAÇÃO TÉCNICA TR 025/2004. Internal report, Petrobras, Brazil, 2004.
- [98] GE Petrosky and FF Farshad. Pressure-Volume-Temperature correlations for Gulf of Mexico crude oils. volume 1, page 416–420. OnePetro, 1998.
- [99] Davide Picchi and Pietro Poesio. Uncertainty quantification and global sensitivity analysis of mechanistic one-dimensional models and flow pattern transition boundaries predictions for two-phase pipe flows. *International Journal of Multiphase Flow*, 90:64–78, 2017.
- [100] J K Pucknell, J N E Mason, and E G Vervest. An Evaluation of Recent “Mechanistic” Models of Multiphase Flow for Predicting Pressure Drops in Oil and Gas Wells. 09 1993. SPE-26682-MS.
- [101] Arnald Puy, Samuele Lo Piano, Andrea Saltelli, and Simon A. Levin. sensobol: An R package to compute variance-based sensitivity indices. *Journal of Statistical Software*, 102(5):1–37, 2022.
- [102] Sina Rashidi, Mohammad Mehrad, Hamzeh Ghorbani, David A Wood, Nima Mohamadian, Jamshid Moghadasi, and Shadfar Davoodi. Determination of bubble point pressure & oil formation volume factor of crude oils applying multiple hidden layers extreme learning machine algorithms. *Journal of Petroleum Science and Engineering*, 202:108425, 2021.
- [103] MR Riazi. *Characterization and properties of petroleum fractions*, volume 50. ASTM international, 2005.
- [104] MR Riazi, HA Al-Adwani, and A Bishara. The impact of characterization methods on properties of reservoir fluids and crude oils: options and restrictions. *Journal of Petroleum Science and Engineering*, 42(2-4):195–207, 2004.

- [105] Donald B Robinson and Ding-Yu Peng. *The characterization of the heptanes and heavier fractions for the GPA Peng-Robinson programs.* Gas Processors Association, 1978.
- [106] Murray Rosenblatt. Remarks on a multivariate transformation. *The annals of mathematical statistics*, 23(3):470–472, 1952.
- [107] J G Ross. Guidelines for the evaluation of petroleum reserves and resources, 2001.
- [108] Marco Rotondi, Giovanna Nicotra, Antonella Godi, F Michela Contento, Martin Julian Blunt, Mike Christie, et al. Hydrocarbon production forecast and uncertainty quantification: A field application. In *SPE Annual Technical Conference and Exhibition*. Society of Petroleum Engineers, 2006.
- [109] Allen McGhee Rowe. Internally consistent correlations for predicting phase compositions for use in reservoir composition simulators. In *SPE Annual Fall Technical Conference and Exhibition*, pages SPE-7475-MS. OnePetro, 1978.
- [110] Leonardo de Pádua Agripa Sales, Anselmo Ramalho Pitombeira-Neto, and Bruno de Athayde Prata. A genetic algorithm integrated with monte carlo simulation for the field layout design problem. *Oil & Gas Sciences and Technology-Revue d'IFP Energies nouvelles*, 73:24, 2018.
- [111] C G S Santim, L P Fulchignoni, E S Rosa, and E F Gaspari. Transient multiphase flow modeling and validation in a real production system with high CO₂ content using the drift-flux model. *Journal of Petroleum Science and Engineering*, 188, 2020.
- [112] Schlumberger. PIPESIM® Version 2017.2 User Guide. 2017.
- [113] Laio Oriel Seman, Luis Kin Miyatake, Eduardo Camponogara, Caio Merlini Giuliani, and Bruno Ferreira Vieira. Derivative-free parameter tuning for a well multiphase flow simulator. *Journal of Petroleum Science and Engineering*, 192:107288, 2020.

- [114] Mack Shippen and William J Bailey. Steady-State Multiphase Flow - Past, Present, and Future, with a Perspective on Flow Assurance. *Energy & Fuels*, 26(7):4145–4157, 2012.
- [115] Giorgio Soave. Equilibrium constants from a modified redlich-kwong equation of state. *Chemical engineering science*, 27(6):1197–1203, 1972.
- [116] T M Soria, A E Andreatta, S Pereda, and S B Bottini. Thermodynamic modeling of phase equilibria in biorefineries. *Fluid Phase Equilibria*, 302(1-2):1–9, 2011.
- [117] Marshall Burton Standing. *Volumetric and phase behavior of oil field hydrocarbon systems*. Society of Petroleum Engineers of AIME, 1977.
- [118] MB Standing. A pressure-volume-temperature correlation for mixtures of California oils and gases. In *Drilling and Production Practice*. OnePetro, 1947.
- [119] E Strange, A Majeed, and S Overa. Theory and application of C7+ characterization in a North Sea transportation and terminal system. Technical report, New York, NY; American Institute of Chemical Engineers, 1988.
- [120] E Stoian. Fundamentals and applications of the Monte Carlo method. *Journal of Canadian Petroleum Technology*, 4(03):120–129, 1965.
- [121] Andreas Strand, Ivar Eskerud Smith, Tor Erling Unander, Ingelin Steinsland, and Leif Rune Hellevik. Uncertainty Propagation through a Point Model for Steady-State Two-Phase Pipe Flow. *Algorithms*, 13(3):53, 2020.
- [122] Daniel M Tartakovsky. Assessment and management of risk in subsurface hydrology: A review and perspective. *Advances in Water Resources*, 51:247–260, 2013.
- [123] S. Taverniers, S. B. M. Bosma, and D. M. Tartakovsky. Accelerated multilevel Monte Carlo with kernel-based smoothing and Latinized stratification. *Water Resour. Res.*, 56(9):e2019WR026984, 2020.

- [124] S. Taverniers and D. M. Tartakovsky. Estimation of distributions via multilevel Monte Carlo with stratified sampling. *J. Comput. Phys.*, 419:109572, 2020.
- [125] Chorng H Twu. An internally consistent correlation for predicting the critical properties and molecular weights of petroleum and coal-tar liquids. *Fluid Phase Equilibria*, 16(2):137–150, 1984.
- [126] Johannes Diderik Van der Waals. *Over de Continuiteit van den Gas-en Vloeistofstoestand*, volume 1. Sijthoff, 1873.
- [127] N Van Wingen. Viscosity of Oil, Water, Natural Gas, and Crude Oil at Varying Pressures and Temperatures. *The Secondary Recovery of Oil in the United States*, 1950.
- [128] Milton Vazquez and H Dale Beggs. Correlations for fluid physical property prediction. In *SPE Annual Fall Technical Conference and Exhibition*. OnePetro, 1977.
- [129] J Velarde, TA Blasingame, and WD McCain. Correlation of Black Oil Properties at Pressures Below Bubble Point Pressure-A New Approach. In *Annual Technical Meeting*. OnePetro, 1997.
- [130] G Venkatarathnam. Density marching method for calculating phase envelopes. *Industrial & engineering chemistry research*, 53(9):3723–3730, 2014.
- [131] GB Wallis. One-Dimensional Two-Phase Flow, 1969.
- [132] Curtis H Whitson. Characterizing hydrocarbon plus fractions. *Society of Petroleum Engineers Journal*, 23(04):683–694, 1983.
- [133] Curtis H Whitson and Michael R Brûlé. *Phase behavior*, volume 20. Henry L. Doherty Memorial Fund of AIME, Society of Petroleum Engineers, 2000.
- [134] Grant M Wilson. A modified redlich-kwong equation of state, application to general physical data calculations. In *65th National AIChE Meeting, Cleveland, OH*, volume 15, 1969.

- [135] C. L. Winter, A. Guadagnini, D. Nychka, and D. M. Tartakovsky. Multivariate sensitivity analysis of saturated flow through simulated highly heterogeneous groundwater aquifers. *J. Comput. Phys.*, 217(1):166–175, 2006.
- [136] William Woelflin. The viscosity of crude-oil emulsions. In *Drilling and Production Practice*. OnePetro, 1942.
- [137] Xiaochun Xu, Silvia Lasala, Romain Privat, and Jean-Noël Jaubert. E-PPR78: A proper cubic EoS for modelling fluids involved in the design and operation of carbon dioxide capture and storage (CCS) processes. *International Journal of Greenhouse Gas Control*, 56:126–154, 2017.
- [138] H.-J. Yang, F. Boso, H. A. Tchelepi, and D. M. Tartakovsky. Method of distributions for quantification of geologic uncertainty in flow simulations. *Water Resources Research*, 56(7):e2020WR027643, 2020.
- [139] H.-J. Yang, H. A. Tchelepi, and D. M. Tartakovsky. Method of distributions for two-phase flow in heterogeneous porous media. *Water Resour. Res.*, 58(12):e2022WR032607, 2022.
- [140] Larry C Young. Compositional reservoir simulation: A review. *SPE Journal*, pages 1–47, 2022.
- [141] Zhixue Zheng, Yuan Di, and Enyi Yu. A dl-kf modeling for acceleration of flash calculations in phase equilibrium using deep learning methods. Available at SSRN 3991785, 2023.
- [142] RG Ziervogel and BE Poling. A simple method for constructing phase envelopes for multicomponent mixtures. *Fluid Phase Equilibria*, 11(2):127–135, 1983.
- [143] Novak Zuber and J ASME Findlay. Average volumetric concentration in two-phase flow systems. *J. Heat Transf.-Trans. ASME*, 87:453–468, 1965.
- [144] Julian Y Zuo and Dan Zhang. Plus fraction characterization and PVT data regression for reservoir fluids near critical conditions. In *SPE Asia Pacific Oil and Gas Conference and Exhibition*, pages SPE-64520-MS. OnePetro, 2000.