

Uncertain characterization of reservoir fluids due to brittleness of equation of state regression

Lívia Paiva Fulchignoni ^{*}, Daniel M. Tartakovsky

Department of Energy Science and Engineering, Stanford University, Stanford, CA 94305, USA

ARTICLE INFO

Keywords:

EoS
Regression
Parameter tuning
Optimization
Initialization

ABSTRACT

Equations of state (EoS) play a central role in modeling the phase equilibrium of fluid mixtures. Their parameterization involves fitting a model to experimental data, i.e., solving a nonlinear, non-convex, multivariate optimization problem. The latter requires one to select design variables, domains of definition for each variable, and weights assigned to individual measurements. We demonstrate that subjective choices of an optimization algorithm and an initial guess also impact the regression process. Consequently, EoS predictions are fundamentally uncertain even after the EoS tuning to a limited set of experimental data points. We demonstrate this observation for two hydrocarbon reservoir fluids, in which five properties of the heaviest carbon fraction are treated as design variables. While all the optimization algorithms and initial guesses match experimental data for the gas and liquid properties, the resulting EoS parameterizations lead to dramatically different predictions of the fluid's thermophysical behavior in the unsampled pressure and temperature regions. We propose the probabilistic treatment of design variables to quantify the predictive uncertainty of the resulting fluid models.

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 (Soria et al., 2011), planning carbon capture and storage operations (Xu et al., 2017), and quantitative forecasting of multiphase flow in the oil and gas industry (Santim et al., 2020). 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 ensure that available experimental data are sparse and 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 (Pedersen et al., 1984), while the heavier fractions are lumped together in the so-called C_{plus} fraction. Thus, petroleum compositions are most often reported to the C_{7+} , C_{10+} , or C_{20+} fractions and, in rare cases, to the C_{30+}

fraction (Pedersen et al., 1989). The data reported in this way pose an additional challenge to the EoS parameterization.

Cubic equations of state serve as the benchmark within the industry for modeling hydrocarbon reservoir fluids, yet their application is not without limitations. When employed to describe oil fractions with C_{20+} components, even state-of-the-art software packages and advanced correlations can yield unreliable phase equilibrium calculations (Allahyarzadeh-Bidgoli et al., 2021). Additionally, the analysis of CO_2 -rich systems suggests that highly asymmetric mixtures challenge the precision boundaries of cubic EoS when coupled with conventional mixing rules (da Silva et al., 2018). A noteworthy limitation is the demonstrated inadequacy of cubic EoS in accurately predicting asphaltene gradients in reservoirs (Mullins et al., 2013), an issue that might prompt consideration of alternative EoS models. One such alternative includes the Flory–Huggins–Zuo EoS in combination with the Yen–Mullins Model for a more reliable representation of asphaltenes (Mullins et al., 2017). Despite these recognized limitations, this study utilizes the modified Peng–Robinson EoS (Peng and Robinson, 1976; Robinson and Peng, 1978) is utilized in this study owing to its widespread acceptance and application in the industry.

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).

^{*} Corresponding author.

E-mail address: liviafp@alumni.stanford.edu (L.P. Fulchignoni).

Nomenclature**Abbreviations**

ADAM	Adaptive moment estimation method
C_{plus}	Hydrocarbon plus fraction
CMA-ES	Covariance matrix adaptation evolution strategy
DFP	Davidon–Fletcher–Powell method
DSM	Direct search method
EoS	Equation of state
PDF	Probability density function
PVT	Pressure–volume–temperature
SCN	Single carbon number

Superscripts

*	Normalized variable
---	---------------------

Subscripts

cr	Critical
comp	Components
g	Gas
m	Mixture
meas	Measurements
o	Oil
Pen	Peneloux correction

Variables

\mathbf{z}	Composition vector of fluid mixture
ω	Acentric factor
Ω_a, Ω_b	Cubic EoS parameters
a	Attraction parameter
b	Covolume parameter
N	Total number
γ	Specific gravity
\mathbf{x}	Design variables vector
\mathcal{E}	Percentage error
ρ	Density
B	Formation volume factor
c	Volume shift parameter
e	EoS prediction
F	Objective function
M	Molecular weight
N_{iter}	Total number of iterations
P	Pressure
P_{sat}	Saturation pressure
R	Universal gas constant
R_s	Solution gas–oil ratio
T	Temperature
V	Volume
v	Molar volume
w	Weight
y	Experimental value
Z	Gas compressibility factor
BIP	Binary interaction parameter

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 (Coats and Smart, 1986) 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 Brulé (2000) 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 (1999) 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. Aguilar Zurita and McCain (2002) 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 (2004) 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 (Deepstar, 2001).

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 (Coats and Smart, 1986; Whitson and Brulé, 2000; Christensen, 1999; Aguilar Zurita and McCain, 2002; Al-Meshari, 2004) on 30 fluid samples (19 oil and 11 gas) suggest that the approaches from Al-Meshari (2004) and Whitson and Brulé (2000) perform best for oil and gas fluids, respectively (Ali and El-Banbi, 2015). 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 hyper-parameters is known to affect the optimization trajectory and final destination. An initial guess can also impact the optimization path. In Section 2, we describe four representative optimization algorithms considered in this work: the adaptive moment estimation method (ADAM) (Kingma and Ba, 2014), the Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), the covariance matrix adaptation evolution strategy method (CMA-ES) (Hansen, 2006), and the direct search method (Hooke and Jeeves, 1961) (DSM). Section 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 5.

2. Methodology

To be specific, we consider the modified Peng–Robinson EoS (Peng and Robinson, 1976; Robinson and Peng, 1978),

$$P = \frac{RT}{v - b_m} - \frac{a_m}{v^2 + 2b_m v - b_m^2}, \quad (1a)$$

because of its widespread usage and simplicity. Here, P is the pressure, T is the temperature, v is the molar volume, R is the universal gas constant. The attraction, a_m , and covolume, b_m , parameters are computed for the fluid mixture of N_{comp} components with composition $\mathbf{z} = (z_1, \dots, z_{N_{\text{comp}}})^T$ according to the mixing rules

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

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

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

$$m_i = \begin{cases} 0.37464 + 1.54226\omega_i - 0.26992\omega_i^2, & \omega_i \leq 0.49 \\ 0.379642 + 1.48503\omega_i - 0.164423\omega_i^2 + 0.016666\omega_i^3, & \omega_i > 0.49, \end{cases} \quad (1d)$$

with $\Omega_a = 0.45724$ and $\Omega_b = 0.07780$. The properties of pure hydrocarbon components, generalized single carbon number fractions, and light gases are reported in [Appendix A](#).

For given P and T , the molar volume v predicted with Eq. (1) is shifted to v_{pen} according to [Péneloux et al. \(1982\)](#) and [Jhaveri and Youngren \(1988\)](#)

$$v_{\text{pen}} = v - \sum_{i=1}^{N_{\text{comp}}} z_i c_i, \quad (2)$$

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

2.1. Problem formulation

The design variables used to fit the EoS (1) 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 $\mathbf{x} = (P_{\text{cr},C_{\text{plus}}}^*, T_{\text{cr},C_{\text{plus}}}^*, \omega_{C_{\text{plus}}}^*, c_{C_{\text{plus}}}^*, M_{C_{\text{plus}}}^*)^T$ is a vector containing these design variables non-dimensionalized by their initial values, so that $\mathbf{x} = \mathbb{1}$ at the first iteration.

Eq. (1) 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} measurements,

$$\mathcal{E}_k(\mathbf{x}) = \frac{e_k(\mathbf{x}) - y_k}{y_k}, \quad k = 1, \dots, N_{\text{meas}}, \quad (3)$$

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 [Agarwal et al. \(1990\)](#), we consider the objective function

$$F(\mathbf{x}) = \sum_{k=1}^{N_{\text{meas}}} w_k^2 \mathcal{E}_k^2(\mathbf{x}), \quad (4)$$

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

$$\text{minimize } F(\mathbf{x}), \quad \text{subject to } 1 - \Delta x_i \leq x_i \leq 1 + \Delta x_i \quad \text{for all } i. \quad (5)$$

The initial values for $P_{\text{cr},C_{\text{plus}}}$ and $T_{\text{cr},C_{\text{plus}}}$ are obtained through Twu's correlation ([Twu, 1984](#)); for $\omega_{C_{\text{plus}}}$ through Lee–Kesler's correlation ([Kesler and Lee, 1976](#)); for the volume shift parameter $c_{C_{\text{plus}}}$ for all components through Jhaveri–Youngren's correlation ([Jhaveri and Youngren, 1988](#)). The initial guess for $M_{C_{\text{plus}}}$ is its measured value reported in the PVT analysis.

Table 1

Molar composition and plus fraction properties of reservoir fluids A and B.

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 ^a	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 ^b

^aPlus fraction. Density = 0.8515 g/cm³ at 60°F. Molecular weight = 218.

^bPlus fraction. Density = 0.9811 g/cm³ at 60°F. Molecular weight = 505.

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.

2.2. Alternative minimization strategies

We discuss the four alternative minimization algorithms used in our comparative study in [Appendix B](#). 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.

2.3. Numerical implementation

We implement the Peng–Robinson EoS ([Peng and Robinson, 1976](#); [Robinson and Peng, 1978](#)) 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.

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 [Whitson and Brulé \(2000\)](#), 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. Their compositions are described in [Table 1](#), along with the density and molecular weight of the plus fractions.

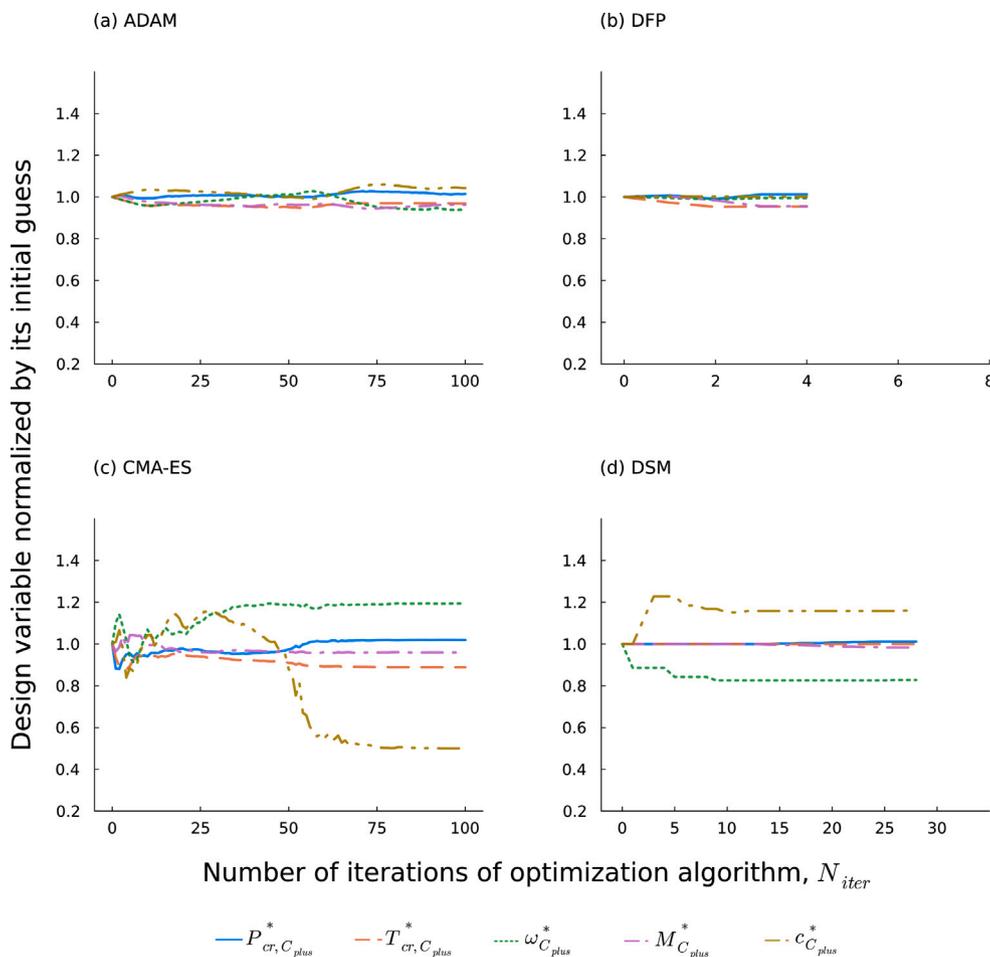


Fig. 1. Convergence history of fluid A's normalized design variables ($P^*_{cr, C_{plus}}$, $T^*_{cr, C_{plus}}$, $\omega^*_{C_{plus}}$, $M^*_{C_{plus}}$, and $c^*_{C_{plus}}$) for alternative optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

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 Whitson and Brulé (2000). Hydrocarbon–hydrocarbon BIPs are set to zero, except for fluid A between methane and C_{7+} , which is computed according to Arbabi and Firoozabadi (1995). Volume shift parameters were defined according to Whitson and Brulé (2000) for fluid A and computed according to Jhaveri and Youngren (1988) for fluid B.

We treat the C_{plus} fraction as a single pseudocomponent as done by Robinson and Peng (1978), Rowe (1978) and Stange et al. (1988). While this approach has its detractors (Aguilar Zurita and McCain, 2002; Christensen, 1999; Al-Meshari, 2004; Zuo and Zhang, 2000), 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 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 (Aguilar Zurita and McCain, 2002), we ignore this source of uncertainty. Third, the single-component approach is still seen in the oil and gas industry (e.g., Li and Yang, 2013).

3. Results

3.1. Optimization algorithm

Fig. 1 exhibits fluid A's convergence history of the normalized design variables ($P^*_{cr, C_{plus}}$, $T^*_{cr, C_{plus}}$, $\omega^*_{C_{plus}}$, $c^*_{C_{plus}}$, $M^*_{C_{plus}}$) for alternative optimization methods: ADAM (Kingma and Ba, 2014), DFP (Davidon, 1959; Fletcher and Powell, 1963), CMA-ES (Hansen, 2006), and DSM (Hooke and Jeeves, 1961). Fig. 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 2 for both fluids. This performance is to be expected given the high degree of nonlinearity of the optimization problem (Eqs. (1) and (2)).

The non-convexity of the objective function is depicted in Fig. 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

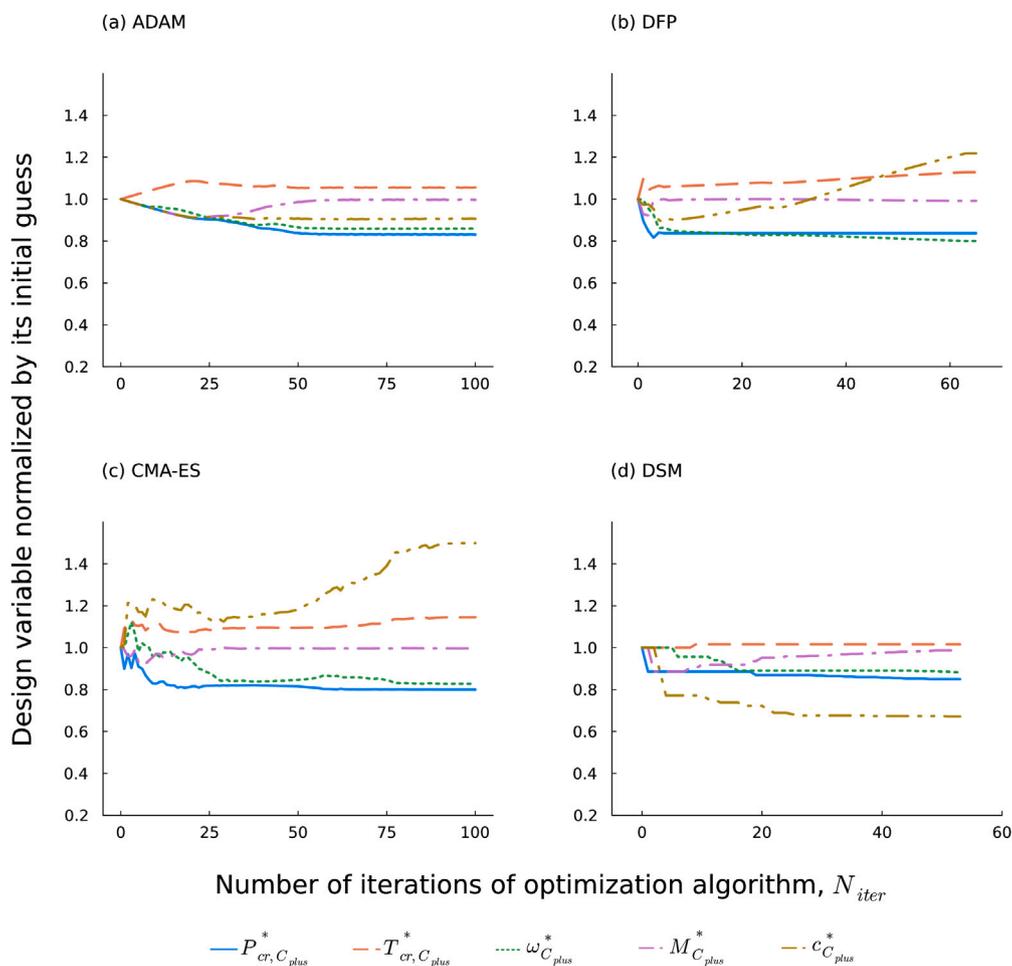


Fig. 2. Convergence history of fluid B's normalized design variables ($P_{cr, C_{plus}}^*$, $T_{cr, C_{plus}}^*$, $\omega_{C_{plus}}^*$, $M_{C_{plus}}^*$, and $c_{C_{plus}}^*$) for alternative optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

Table 2
EoS regression results obtained with four alternative optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

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

two solutions. All in all, the results presented in Figs. 1–3 and Table 2 demonstrate that results of the EoS regression depend on the choice of an optimization algorithm.

Fig. 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 (Eqs. (1)–(2)). 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 2). 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, Figs. 5–6 and Table 3 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

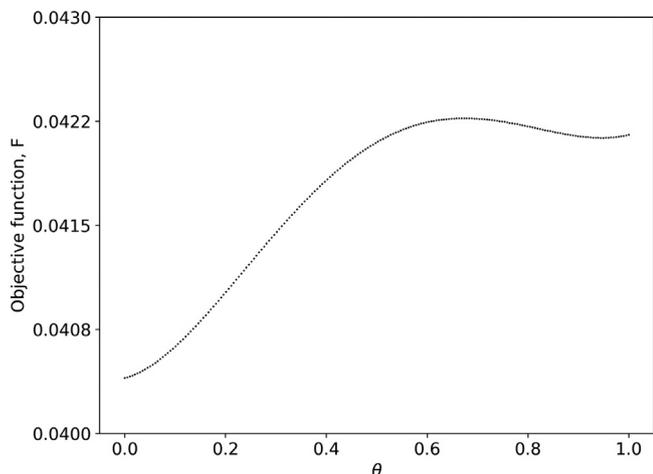


Fig. 3. Objective function (squared relative error) evaluated for fluid B at convex combinations between the solutions obtained by the covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006) ($\theta = 0$) and the direct search method (DSM) (Hooke and Jeeves, 1961) ($\theta = 1$). The curve is generated using 200 objective function evaluations.

Table 3

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) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

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

these regimes. We investigate the ability of the EoS models, tuned with the four alternative optimization techniques on “training” data, to extrapolate from such regimes, i.e., to represent the unseen data.

Fig. 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.

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 (Agarwal et al., 1990).

The phase envelopes for thus refined EoS are plotted in Fig. 8 for fluids A and B, demonstrating the sensitivity of DFP minimization to an

Table 4

EoS-regression results obtained via the two-step regression in which the optimal values of the decision variables obtained with ADAM (Kingma and Ba, 2014), CMA-ES (Hansen, 2006), or DSM (Hooke and Jeeves, 1961) serve as the initial guess for the subsequent DFP minimization (Davidon, 1959; Fletcher and Powell, 1963).

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

initial guess. The phase envelopes optimized using the initial guesses from ADAM, CMA-ES, and DSM differ significantly. Table 4 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.

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 ($P_{cr,C_{plus}}^*$, $T_{cr,C_{plus}}^*$, $\omega_{C_{plus}}^*$, $M_{C_{plus}}^*$, and $c_{C_{plus}}^*$) of Fluid B, for which the optimized values are alternatively obtained by the four optimization algorithms (Section 2.2) and by the two-stage EoS-tuning procedure (Section 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 Fig. 9.

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

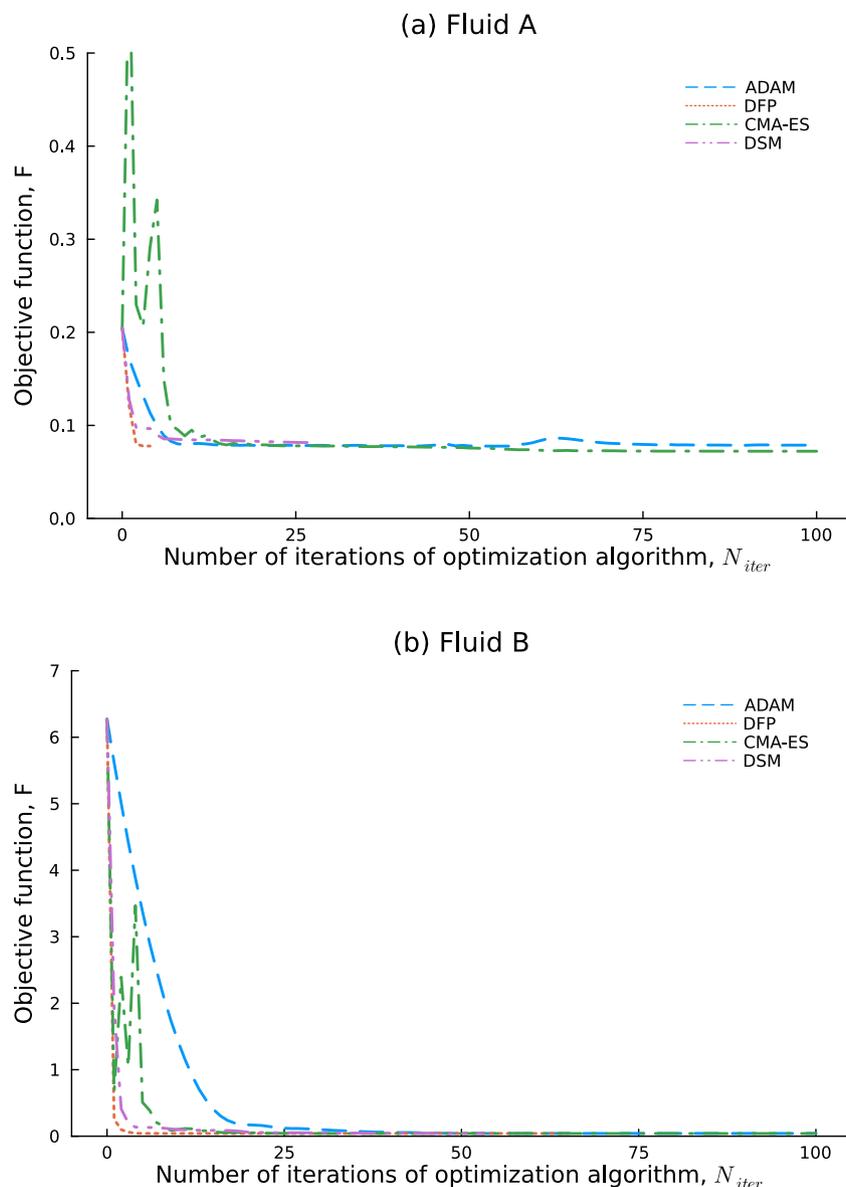


Fig. 4. Convergence history of the objective function F for (a) fluid A and (b) fluid B, and four optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

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 (Alawadhi et al., 2018; Fulchignoni et al., 2023) and porous media flow (Yang et al., 2020, 2022) models. Quantifying the EoS prediction uncertainty and its impact on the overall flow predictions is also an area of future work.

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) (Kingma and Ba, 2014), the Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), the covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and the direct search method (DSM) (Hooke and Jeeves, 1961). These were deployed to fit the Peng–Robinson EoS (Robinson and Peng, 1978) to experimental data for two hydrocarbon reservoir fluids. Our study leads to the following major conclusions.

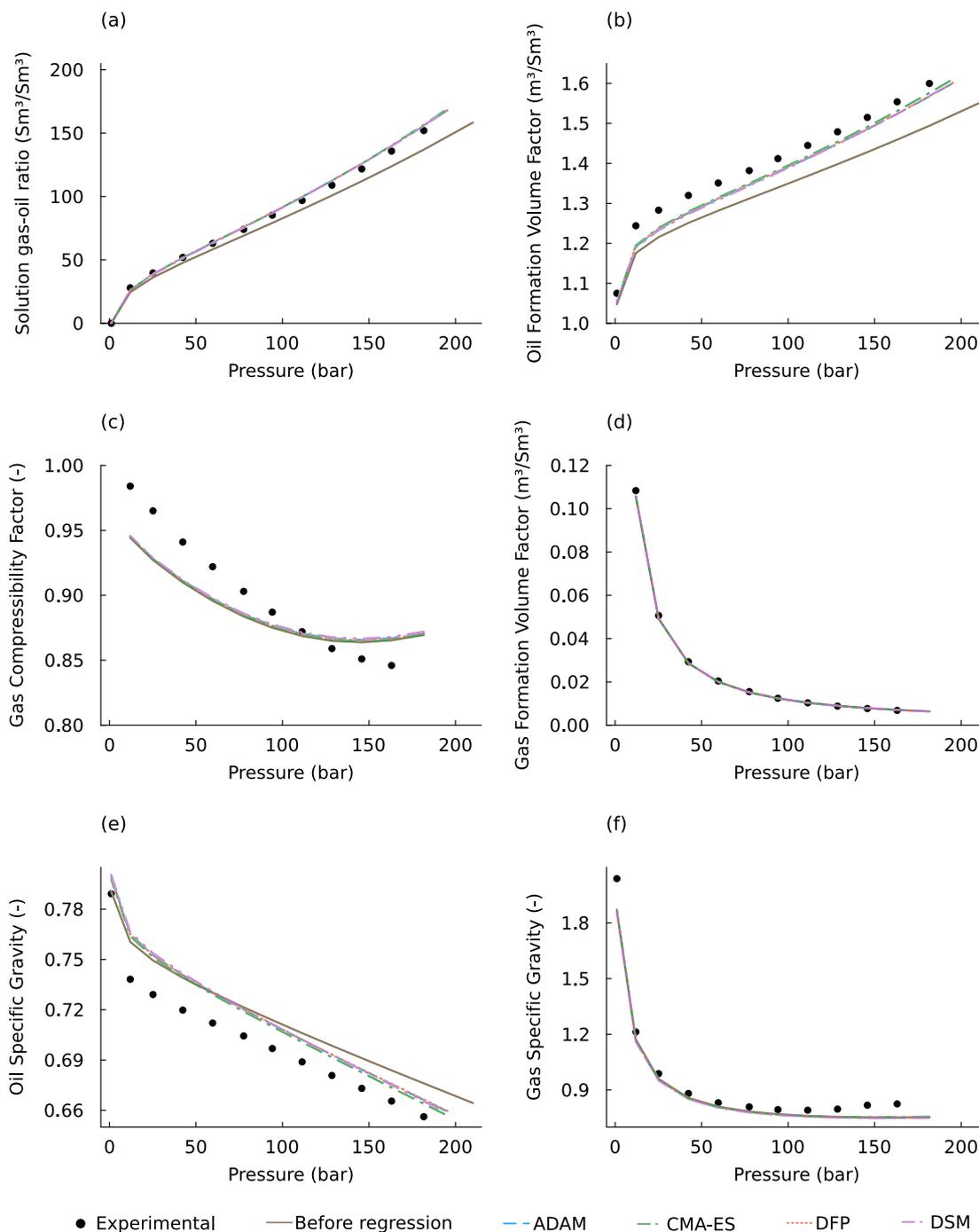


Fig. 5. Fluid A EoS predictions and experimental data for (a) solution gas–oil ratio, (b) oil formation volume factor, (c) gas compressibility factor, (d) gas formation volume factor, (e) oil specific gravity, and (f) gas specific gravity. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

- 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

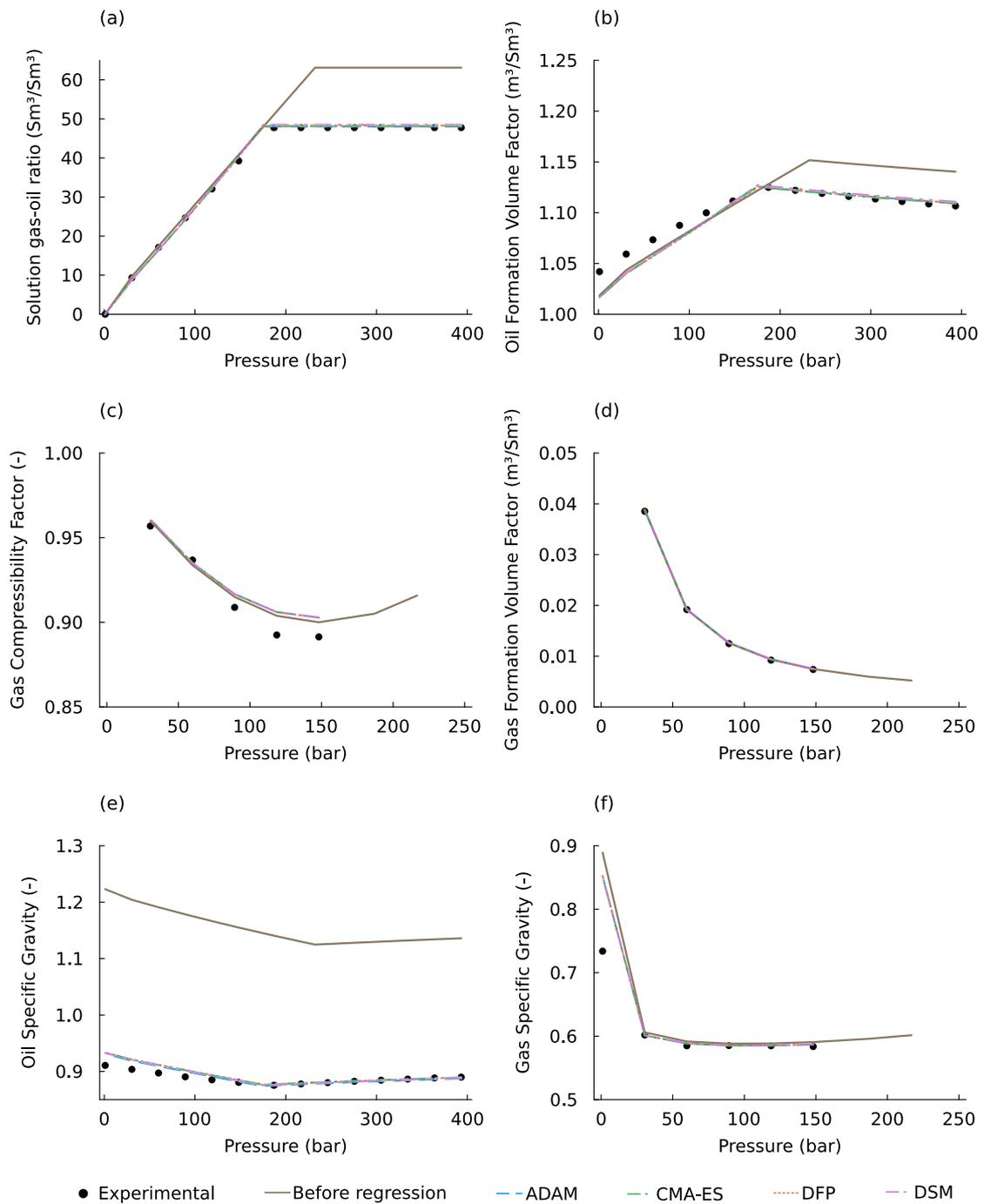


Fig. 6. Fluid B EoS predictions and experimental data for (a) solution gas-oil ratio, (b) oil formation volume factor, (c) gas compressibility factor, (d) gas formation volume factor, (e) oil specific gravity, and (f) gas specific gravity. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon-Fletcher-Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961).

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.

CRedit authorship contribution statement

Lívia Paiva Fulchignoni: Conceptualization, Methodology, Software, Investigation, Writing – original draft. **Daniel M. Tartakovsky:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors thank Dr. Khalid Aziz for the fruitful discussions and Petrobras for the experimental data and permission to publish this paper. This research was supported in part by the Office of Energy

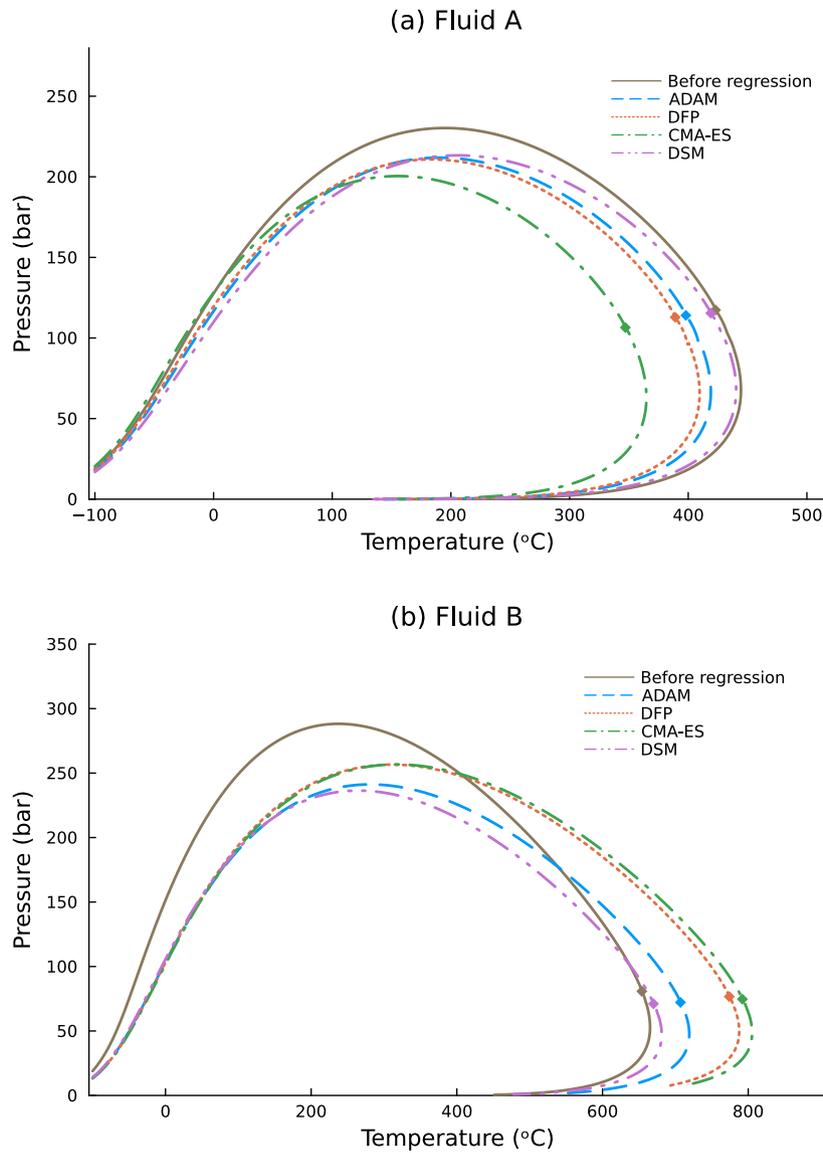


Fig. 7. Phase envelopes (vapor liquid equilibrium) for the EoS models before and after regression of (a) fluid A and (b) fluid B. The EoS tuning is carried out with four alternative optimization algorithms: adaptive moment estimation (ADAM) (Kingma and Ba, 2014), Davidon–Fletcher–Powell method (DFP) (Davidon, 1959; Fletcher and Powell, 1963), covariance matrix adaptation evolution strategy (CMA-ES) (Hansen, 2006), and direct search method (DSM) (Hooke and Jeeves, 1961). The markers indicate the critical point.

Efficiency and Renewable Energy (EERE), USA of the U.S. Department of Energy (DOE) under project number FY19 AOP 3.1.8.7.

Appendix A

For the sake of completeness, we collate in Table A.5 the relevant properties of pure hydrocarbon components, generalized single carbon number fractions, and light gases.

Appendix B

This Appendix describes the four optimization algorithms implemented in our study.

The adaptive moment estimation method (Kingma and Ba, 2014). 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

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

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

are computed in order to update the design point,

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

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 (Davidon, 1959; Fletcher and Powell, 1963). DFP is a second-order gradient decent method. It rests on the

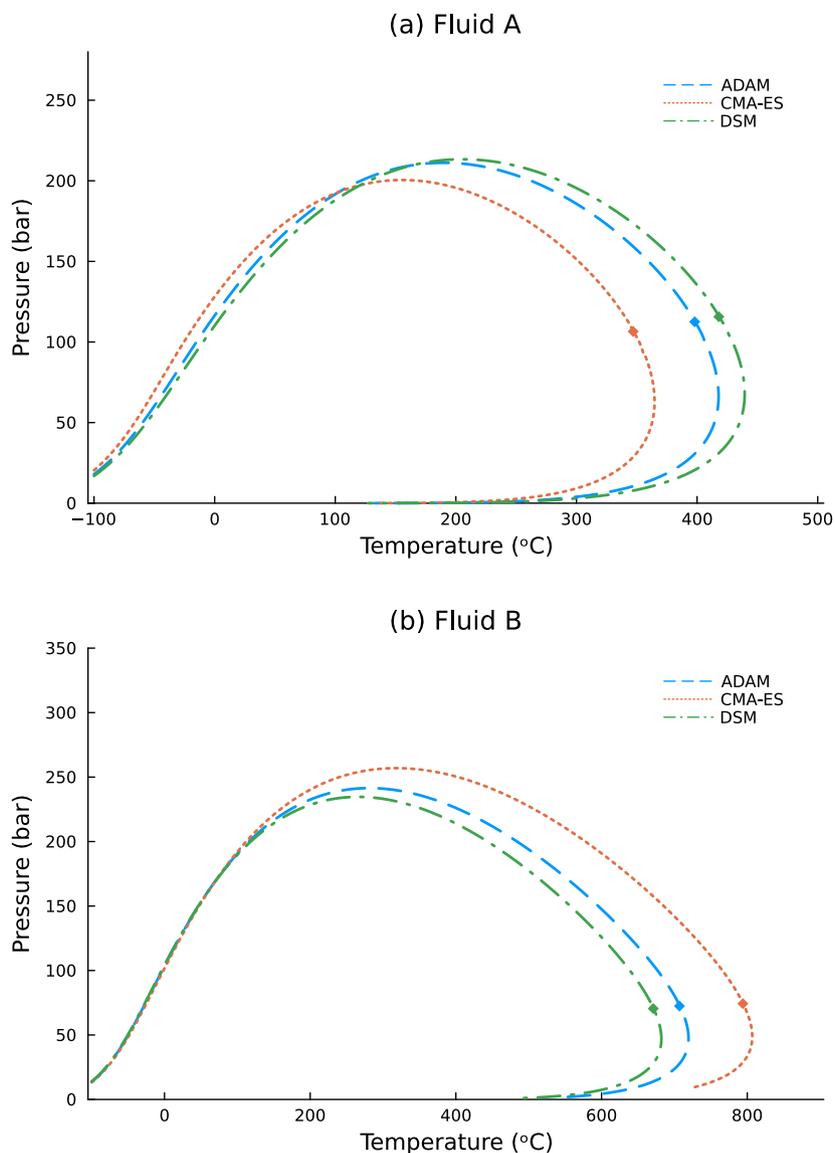


Fig. 8. Phase envelopes (vapor liquid equilibrium) obtained for (a) fluid A and (b) fluid B via the two-step EoS regression in which the optimal values of the decision variables obtained with ADAM (Kingma and Ba, 2014), CMA-ES (Hansen, 2006), or DSM (Hooke and Jeeves, 1961) serve as the initial guess for the subsequent GD20 minimization (Agarwal et al., 1990). The markers indicate the critical point.

Newton’s method,

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - (\mathbf{H}_t)^{-1} \nabla F(\mathbf{x}_t), \tag{B.2}$$

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

$$\mathbf{x}_{t+1} \leftarrow \mathbf{x}_t - \alpha_t \mathbf{Q}_t \nabla F(\mathbf{x}_t), \tag{B.3}$$

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

The covariance matrix adaptation evolution strategy method (Hansen, 2006). 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 (Eiben et al., 2003).

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 Hansen (2006).

Direct search method. The Hooke–Jeeves algorithm (Hooke and Jeeves, 1961), 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 $\mathbf{x} \in \mathbb{R}^5$, $F[\mathbf{x} + (\alpha, 0, 0, 0, 0)^T]$, $F[\mathbf{x} + (-\alpha, 0, 0, 0, 0)^T]$, ..., $F[\mathbf{x} + (0, 0, 0, 0, \alpha)^T]$ and $F[\mathbf{x} + (0, 0, 0, 0, -\alpha)^T]$ 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 step size decreases for a finer search.

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

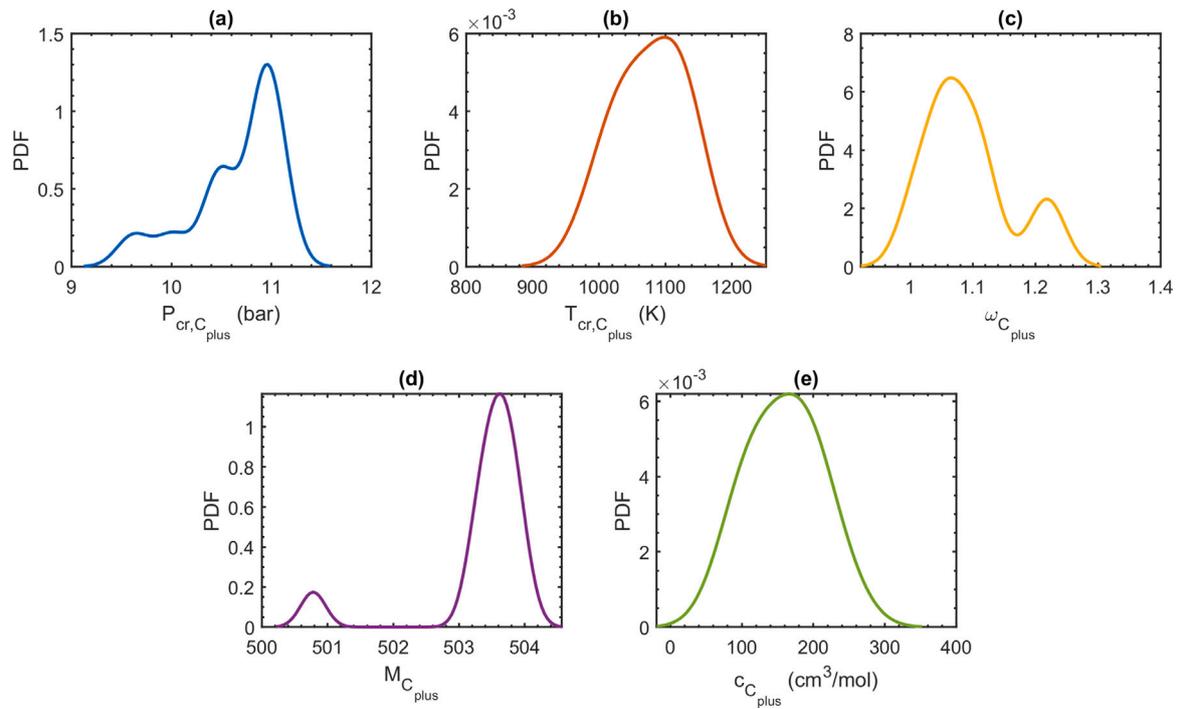


Fig. 9. Fitted probability density function (PDF) for the C_{plus} fraction properties: (a) critical pressure, (b) critical temperature, (c) acentric factor, (d) molecular weight, and (e) volume shift parameter.

Table A.5

Properties of the reservoir fluid components (CO_2 , N_2 , and hydrocarbons up to C_{19} fraction).

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

Algorithm 1 Adaptive Moment Estimation Method (ADAM) optimization algorithm (Kingma and Ba, 2014)

- 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:** $F(x)$: Objective function with parameters x
- 5: **Require:** x_0 : Initial guess
- 6: $m_0 \leftarrow 0$ (Initialize 1st moment vector)
- 7: $v_0 \leftarrow 0$ (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: $m_t \leftarrow \beta_1 m_{t-1} + (1 - \beta_1) \nabla F(x_{t-1})$ (Update biased first moment estimate)
- 12: $v_t \leftarrow \beta_2 v_{t-1} + (1 - \beta_2) \nabla F(x_{t-1}) \odot \nabla F(x_{t-1})$ (Update biased second raw moment estimate)
- 13: $\hat{m}_t \leftarrow m_t / (1 - (\beta_1)^t)$ (Compute bias-corrected first moment estimate)
- 14: $\hat{v}_t \leftarrow v_t / (1 - (\beta_2)^t)$ (Compute bias-corrected second raw moment estimate)
- 15: $x_t \leftarrow x_{t-1} - \alpha \hat{m}_t \oslash (\sqrt{\hat{v}_t} + \epsilon)$ (Update parameters)
- 16: **end while**
- 17: **Return** x_t (Resulting parameters)

Algorithm 2 Davidon–Fletcher–Powell (DFP) optimization algorithm (Davidon, 1959; Fletcher and Powell, 1963)

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: $\delta_t \leftarrow \mathbf{x}_{t+1} - \mathbf{x}_t$ (Compute change in parameter vector)
 12: $\boldsymbol{\gamma}_t \leftarrow \mathbf{g}_{t+1} - \mathbf{g}_t$ (Compute change in gradient)
 13: $\mathbf{Q}_{t+1} \leftarrow \mathbf{Q}_t + \frac{\delta_t \delta_t^\top}{\delta_t^\top \boldsymbol{\gamma}_t} - \frac{\mathbf{Q}_t \boldsymbol{\gamma}_t \boldsymbol{\gamma}_t^\top \mathbf{Q}_t}{\boldsymbol{\gamma}_t^\top \mathbf{Q}_t \boldsymbol{\gamma}_t}$ (Update inverse Hessian approximation)
 14: $t \leftarrow t + 1$ (Increment iteration counter)
 15: **end while**
 16: **Return** \mathbf{x}_t

Algorithm 3 Covariance Matrix Adaptation Evolution Strategy (CMA-ES) optimization algorithm (Hansen, 2006) (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: $w'_i \leftarrow \ln \frac{\lambda + 1}{2} - \ln i$ 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: $w_i = \begin{cases} \frac{1}{\sum_{j=1}^{\lambda} |w'_j|} w'_i & \text{if } w'_i \geq 0 \\ \frac{\min(\alpha_\mu^-, \alpha_{\mu_{\text{eff}}}^-, \alpha_{\text{pos def}}^-)}{\sum_{j=1}^{\lambda} |w'_j|} w'_i & \text{if } w'_i < 0 \end{cases}, i = 1, \dots, \lambda$
 20: $E \leftarrow \sqrt{n} \left(1 - \frac{1}{4n} + \frac{1}{21n^2} \right)$
 21: $\mathbf{p}_\sigma, \mathbf{p}_\Sigma, \boldsymbol{\Sigma} \leftarrow \mathbf{0}, \mathbf{0}, \mathbf{I}$
 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 (Hansen, 2006) (Part 2)

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

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: $\mathbf{y}_B \leftarrow F(\mathbf{x}_t)$
 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: $\mathbf{y}_T \leftarrow F(\mathbf{x}_T)$
 15: **if** $\mathbf{y}_T < \mathbf{y}_B$ **then**
 16: $\mathbf{x}_B \leftarrow \mathbf{x}_T$ (Update best solution)
 17: $\mathbf{y}_B \leftarrow \mathbf{y}_T$
 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

References

- Agarwal, R.K., Li, Y.K., Nghiem, L., 1990. A regression technique with dynamic parameter selection for phase-behavior matching. *SPE Reserv. Eng.* 1, 115–120.
- Aguiar Zurita, R.A., McCain, W.D., 2002. An efficient tuning strategy to calibrate cubic EOS for compositional simulation. In: *SPE Annual Technical Conference and Exhibition*. OnePetro, SPE-77382-MS.
- Al-Meshari, A.A., 2004. New Strategic Method to Tune Equation-Of-State to Match Experimental Data for Compositional Simulation. Texas A & M University.
- Alawadhi, A., Boso, F., Tartakovsky, D.M., 2018. Method of distributions for water-hammer equations with uncertain parameters. *Water Resour. Res.* 54 (11), 9398–9411. <http://dx.doi.org/10.1029/2018WR023383>.
- Ali, M.T., El-Banbi, A.H., 2015. EOS tuning - comparison between several valid approaches and new recommendations. In: *SPE North Africa Technical Conference and Exhibition*. OnePetro, SPE-175877-MS.
- Allahyarzadeh-Bidgoli, A., Heidaryan, E., Yanagihara, J.I., Pessôa Filho, P.d.A., 2021. Assessment of correlations and simulation software to calculate phase diagrams of pre-salt fluids. *Petrol. Sci. Technol.* 39 (11–12), 410–420.
- Arbabi, S., Firoozabadi, A., 1995. Near-critical phase behavior of reservoir fluids using equations of state. *SPE Adv. Technol. Ser.* 3 (01), 139–145.
- Christensen, P., 1999. Regression to experimental PVT data. *J. Can. Pet. Technol.* 38 (13).
- Coats, K.H., Smart, G.T., 1986. Application of a regression-based EOS PVT program to laboratory data. *SPE Reserv. Eng.* 1 (3), 277–299.
- da Silva, V.M., do Carmo, R.P., Fleming, F.P., Daridon, J.-L., Pauly, J., Tavares, F., 2018. High pressure phase equilibria of carbon dioxide+ n-alkanes mixtures: Experimental data and modeling. *Fluid Phase Equilib.* 463, 114–120.
- Davidon, W.C., 1959. Variable Metric Method for Minimization. Argonne National Laboratory, <http://dx.doi.org/10.2172/4252678>.
- Deepstar, 2001. Flow Assurance Design Guideline, Deepstar IV Project, DSIV CTR 4203b-1.
- Eiben, A.E., Smith, J.E., et al., 2003. *Introduction to Evolutionary Computing*, Vol. 53. Springer.
- Fletcher, R., Powell, M.J., 1963. A rapidly convergent descent method for minimization. *Comput. J.* 6 (2), 163–168.
- Fulchignoni, L.P., Santim, C.G.d.S., Tartakovsky, D.M., 2023. Probabilistic forecasting of cumulative production of reservoir fluid with uncertain properties. *Geoenergy Sci. Eng. J.*
- Hansen, N., 2006. The CMA evolution strategy: a comparing review. In: *Towards a New Evolutionary Computation*. pp. 75–102.
- Hooke, R., Jeeves, T.A., 1961. Direct search solution of numerical and statistical problems. *J. ACM* 8 (2), 212–229.
- Jhaveri, B.S., Youngren, G.K., 1988. Three-parameter modification of the Peng–Robinson equation of state to improve volumetric predictions. *SPE Reserv. Eng.* 3 (3), 1033–1040.
- Kesler, M.G., Lee, B.I., 1976. Improve prediction of enthalpy of fractions. *Hydrocarbon Process.* 55, 153–158.
- Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Li, H., Yang, D., 2013. Phase behaviour of C3H8/n-C4H10/heavy-oil systems at high pressures and elevated temperatures. *J. Can. Pet. Technol.* 52 (01), 30–40.
- Mullins, O.C., Seifert, D.J., Zuo, J.Y., Zeybek, M., 2013. Clusters of asphaltene nanoaggregates observed in oilfield reservoirs. *Energy Fuels* 27 (4), 1752–1761.
- Mullins, O.C., Zuo, J.Y., Pomerantz, A.E., Forsythe, J.C., Peters, K., 2017. Reservoir fluid geodynamics: The chemistry and physics of oilfield reservoir fluids after trap filling. *Energy Fuels* 31 (12), 13088–13119.
- Pedersen, K.S., Fredenslund, A., Thomassen, P., 1989. *Properties of Oils and Natural Gases*, Vol. 5. Gulf Publishing Company.
- Pedersen, K.S., Thomassen, P., Fredenslund, A., 1984. Thermodynamics of petroleum mixtures containing heavy hydrocarbons. 1. Phase envelope calculations by use of the soave-redlich-kwong equation of state. *Ind. Eng. Chem. Process Des. Dev.* 23 (1), 163–170.
- Pénéloux, A., Rauzy, E., Fréze, R., 1982. A consistent correction for Redlich–Kwong–Soave volumes. *Fluid Phase Equilib.* 8 (1), 7–23.
- Peng, D.-Y., Robinson, D.B., 1976. A new two-constant equation of state. *Ind. Eng. Chem. Fundam.* 15 (1), 59–64.
- Robinson, D.B., Peng, D.-Y., 1978. The Characterization of the Heptanes and Heavier Fractions for the GPA Peng–Robinson Programs. Gas Processors Association.
- Rowe, A.M., 1978. Internally consistent correlations for predicting phase compositions for use in reservoir composition simulators. In: *SPE Annual Fall Technical Conference and Exhibition*. OnePetro, SPE-7475-MS.
- Santim, C., Fulchignoni, L., Rosa, E., Gaspari, E., 2020. Transient multiphase flow modeling and validation in a real production system with high CO₂ content using the drift-flux model. *J. Pet. Sci. Eng.* 188, 106903.
- Soria, T.M., Andreatta, A.E., Pereda, S., Bottini, S.B., 2011. Thermodynamic modeling of phase equilibria in biorefineries. *Fluid Phase Equilib.* 302 (1–2), 1–9.
- Stange, E., Majeed, A., Overa, S., 1988. Theory and Application of C7+ Characterization in a North Sea Transportation and Terminal System. Tech. Rep., American Institute of Chemical Engineers, New York, NY.
- Twu, C.H., 1984. An internally consistent correlation for predicting the critical properties and molecular weights of petroleum and coal-tar liquids. *Fluid Phase Equilib.* 16 (2), 137–150.
- Whitson, C.H., Brulé, M.R., 2000. *Phase Behavior*, Vol. 20. Society of Petroleum Engineers Inc., Henry L. Doherty Memorial Fund of AIME.
- Xu, X., Lasala, S., Privat, R., Jaubert, J.-N., 2017. E-PPR78: A proper cubic EoS for modelling fluids involved in the design and operation of carbon dioxide capture and storage (CCS) processes. *Int. J. Greenh. Gas Control* 56, 126–154.
- Yang, H.-J., Boso, F., Tchelepi, H.A., Tartakovsky, D.M., 2020. Method of distributions for quantification of geologic uncertainty in flow simulations. *Water Resour. Res.* 56 (7), <http://dx.doi.org/10.1029/2020WR027643>.
- Yang, H.-J., Tchelepi, H.A., Tartakovsky, D.M., 2022. Method of distributions for two-phase flow in heterogeneous porous media. *Water Resour. Res.* 58 (12), <http://dx.doi.org/10.1029/2022WR032607>.
- Zuo, J.Y., Zhang, D., 2000. Plus fraction characterization and PVT data regression for reservoir fluids near critical conditions. In: *SPE Asia Pacific Oil and Gas Conference and Exhibition*. OnePetro, SPE-64520-MS.