

ORIGINAL ARTICLE

DOI: <https://doi.org/10.18599/grs.2026.2.3>

Determining Gas Condensate Composition Using Well Test Data and Optimization Algorithms

B.N. Starovoytova^{1,2}, B.Kh. Imomnazarov^{1,2}, A.N. Baykin^{1,2}*¹*Novosibirsk State University, Novosibirsk, Russian Federation*²*Lavrentyev Institute of Hydrodynamics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russian Federation*

Abstract. This study proposes an optimization-based approach to determine the actual fluid composition of gas condensate reservoirs when obtaining representative samples are impossible. The method incorporates the well tests hydrodynamic modeling, laboratory analyses of non-representative lean samples, and field data, including the gas-condensate ratio (GCR). The reservoir composition is assumed to be a linear combination of lean gas and its equilibrium condensate. The proportionality (mixing) parameter is obtained by minimizing the discrepancy between observed and simulated GCR values obtained using tNavigator. Two variants are considered: 1) a scalar parameter, corresponding to mixing of equilibrium gas and condensate; 2) a vector-valued mixing parameter, permitting per-component adjustment for improved accuracy. For the vector mixing parameter, a check is performed for compliance with the gamma distribution of the obtained heavy component fractions relative to their molecular weight. The approach is verified for a synthetic case with a known reservoir composition. For a detailed 34-component “lean” sample model, the scalar parameter approach accurately reproduces key PVT properties such as the dew point pressure and condensate dropout curve from constant-volume depletion tests. Reduced-component fluid models require the vector-valued mixing parameter to achieve comparable accuracy. To evaluate robustness against field uncertainties, Gaussian noise is introduced into the actual GCR data. Numerical experiments confirm that the method remains reliable if the error in noisy data does not exceed 10% relative to actual GCR.

Keywords: Gas condensate reservoir, reservoir fluid composition, well test, lean sample, numerical optimization, NOMAD, PSO, DE

Recommended citation: Starovoytova B.N., Imomnazarov B.Kh., Baykin A.N. (2026). Determining gas condensate composition using well test data and optimization algorithms. *Georesursy = Georesources*, 28(2), pp. 186–198. <https://doi.org/10.18599/grs.2026.2.3>

Introduction

In the early stages of reservoir development, gas condensate well tests are used to estimate the well performance, production of gas and condensate, and their economic potential.

An important part of these well tests involves laboratory studies of the produced fluids samples (i.e., condensate and gas) to determine their composition and key PVT properties such as Z-factor, dew point pressure, gas viscosity etc. The work (Osfouri,

Azin, 2015) considers the main problems that arise during the sampling and subsequent recombination of the laboratory gas and condensate.

To obtain a representative fluid sample the pressure drawdown should not be excessive and, at the same time, the gas velocity should be high enough to push the liquid out of the wellbore bottomhole and avoid the well loading (AIP, 2003; R Gazprom 086–2010, 2011).

However, in low-permeability reservoirs that are rich in condensate, it is difficult and often impossible to satisfy these conditions simultaneously. In such cases, reservoir fluids are sampled under high pressure drawdown, resulting in lean samples due to the loss of heavy components either in the formation or at the wellbore bottomhole. It is important to note that the thermophysical (PVT) properties of reservoir fluids

*Corresponding author: Botagoz N. Starovoytova
e-mail: b.starovoytova@nsu.ru

© 2026 The Authors. Published by Georesursy LLC

This is an open access article under the Creative Commons Attribution 4.0 License (<https://creativecommons.org/licenses/by/4.0/>)

depend directly on the molar fractions and molecular weights of heavy components (Elsharkawy, 2002).

Thus, the lean sample is not representative and does not exhibit the same PVT properties as the actual reservoir fluid. A way out of this situation can be an approach in which the lean sample is saturated with heavy components until the actual reservoir composition is obtained. In (Brusilovskiy, Yushchenko, 2016; Promzelev et al., 2018) numerical methods for restoring the actual reservoir fluid compositions are presented. In these studies, the actual (true) dew point pressure (P_{dew}) is assumed to be equal to the reservoir pressure. First, a PVT model of the initial gas is created by calibrating an equation of state (EoS) against the laboratory data of the lean fluid sample. Next, using the PVT model, the equilibrium gas and condensate compositions at the initial P_{dew} can be found, and mixed until the desired P_{dew} is reached. However, none of these methods take into account the filtration processes in the reservoir or actual well test data. It should also be noted that these methods cannot be applied to undersaturated gas condensate reservoirs, where P_{dew} of the reservoir fluid is not equal to the reservoir pressure.

In (Gimazov et al., 2024) authors use an approach in which the reservoir fluid composition is selected by comparing the actual well test data (gas and condensate flow rates, and gas-condensate ratio – GCR) with the results of the hydrodynamic simulation of the well test. Note that numerical hydrodynamic simulations are now considered reliable for describing the development of various types of reservoirs, provided that all necessary data are of sufficient accuracy. Given reservoir fluid composition, the mathematical model can be used to predict well flow rates at any time. Therefore, if the results of the well test simulation match the actual data for a certain fluid composition, we can accept this composition as an actual one.

In (Gimazov et al., 2024) a set of composition candidates is pre-generated by mixing gas and equilibrium condensate in different proportions. Then the one is selected based on the smallest discrepancy between the actual and calculated production data. Additionally, the sensitivity analysis is carried out to evaluate the impact of uncertainties in key hydrodynamic model parameters on the selection of the acceptable fluid composition. The results show that the choice of the fluid composition based on GCR discrepancies is the least sensitive to variations in the hydrodynamic model parameters. On the contrary, the selection of fluid composition by the discrepancies in gas or condensate flow rates depends strongly on the hydraulic fracture half-length, net-to-gross ratio, reservoir temperature and two relative permeability parameters: minimum water saturation and relative permeability for gas at residual condensate saturation.

The main disadvantage of this approach is that the limited number of composition candidates restricts the possible P_{dew} of the reservoir fluid.

The aim of the present work is to generalize the approach from (Gimazov et al., 2024). It is proposed to seek a suitable fluid composition as a solution for the nonlinear optimization problem without generating the possible fluid compositions variants. As an objective function to minimize we regard the discrepancy between the simulated and some reference or measured GCR.

The optimization problem is to be solved using popular methods such as NOMAD (Nonlinear Optimization by Mesh Adaptive Direct Search), PSO (Particle Swarm Optimization), DE (Differential Evolution), and basic Local Search (LS) algorithms. For hydrodynamic simulation of the well test, we employ the industry acceptable simulator tNavigator (tNavigator, 2023) under an academic license.

While laboratory fluid analyses typically identify 30-50 distinct components, compositional reservoir simulations often use reduced-component (lumped) representations to optimize computational efficiency. These lumping schemes group multiple components into pseudo-components while preserving the fluid's thermodynamic behavior – a key requirement validated through comparative PVT tests. The paper applies the proposed approach using both detailed compositions and their lumped counterparts with varying numbers of pseudo-components.

Materials and Methods

In this study a well test (WT) conducted in a low permeability gas condensate reservoir is modeled. The test design incorporates frequent measurements of gas and condensate production rates, enabling precise calculation of the GCR profile.

During the WT operation, the reservoir fluid samples are collected under high pressure drawdown conditions. It is assumed that the subsequent laboratory analysis of these samples, combined with Equation of State (EoS) calibration, yields the PVT model. However, when sampling conditions deviate from the recommended protocols, the obtained samples tend to be leaner than the original reservoir fluid. As a result, the derived PVT model may not entirely represent the actual reservoir fluid properties.

According to (Gimazov et al., 2024) to define the actual reservoir fluid composition, one should find such a composition that minimizes the discrepancy between the actual and simulated GCR.

Determining the reservoir fluid composition only from the well flowing data without any physical constraints on the feasible solution space is generally challenging, as various compositions may yield the similar gas and condensate flow rates.

Following (Promzelev et al., 2018; Brusilovsky, Yushchenko, 2016) we assume that any admissible reservoir fluid compositions can be obtained by mixing the equilibrium gas and condensate. Mathematically, the mole fraction of k -th component in the mixture, Z_k^m , is defined as follows:

$$Z_k^m = \frac{Y_k^g + \alpha X_k^f}{\sum_{i=1}^N Y_i^g + \alpha X_i^f},$$

$$\sum_{i=1}^N Y_i^g = \sum_{i=1}^N X_i^f = \sum_{i=1}^N Z_i^m = 1, 0 \leq \alpha \leq 1, \quad (1)$$

where Y_k^g , and X_k^f are the mole fractions of k -th component in the equilibrium gas and condensate phases, respectively; N is the total number of components, α is a mixing parameter representing the amount (in moles) of condensate added to the one mole of gas. In (1), it is assumed that the actual gas composition and the lean sample consist of the same components with the same properties.

We also consider a more general approach in which we add pure substances corresponding to each component to the equilibrium gas individually rather than adding their mixture (the condensate). In this case, the mole fraction of each component in the mixture is adjusted independently with its own value. Thus, the mixing parameter becomes an N -dimensional vector $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_N)$, and relation (1) takes the following form:

$$Z_k^m = \frac{Y_k^g + \alpha_k X_k^f}{\sum_{i=1}^N Y_i^g + \alpha_i X_i^f},$$

$$\sum_{i=1}^N Y_i^g = \sum_{i=1}^N X_i^f = \sum_{i=1}^N Z_i^m = 1, 0 \leq \alpha_k \leq 1. \quad (2)$$

To ensure the physical consistency of the composition from (2), the molecular weight distribution of the C_7+ fraction (components with a Single-Carbon-Number more than 7) must follow a gamma distribution (Whitson, 1983), where the exponential distribution is a special case. This constraint is enforced by adding a penalty term to the objective function. The magnitude of this penalty is proportional to the deviation from the theoretical gamma distribution, which is quantified using the R^2 metric. An R^2 value of at least 0.98 is required to accept the fit.

Also, since we consider methane-rich gas condensates, the fluid compositions with low methane content are excluded from consideration via a penalty method. A penalization value of 10^6 is imposed if the methane mole fraction Z_{C1}^m for some fluid composition is below or equal the threshold value of 0.6.

The discrepancy between the actual and calculated GCR is evaluated using the relative mean squared error as

$$\text{rMSE} = \frac{1}{n} \sum_{i=1}^n g \left(\frac{\text{GCR}_i - \text{GCR}_i^*}{\text{GCR}_i^*} \right)^2,$$

where n is the number of time steps, GCR_i and GCR_i^* are the calculated and actual values of GCR at the i -th time step, respectively; $g = 100$ is the weighting factor.

Thus, the problem of determining the actual composition of the reservoir fluid is formulated as follows

Find α (or α), $0 \leq \alpha \leq 0.5$ ($\forall k: 0 \leq \alpha_k \leq 0.5$):

Objective function $\rightarrow \min$, (3)

where

$$\text{Objective function} = \begin{cases} \text{rMSE} + B \cdot |R^2 - 0.98|, & \text{if } Z_{C1}^m > 0.6, \\ \text{rMSE} + B \cdot |R^2 - 0.98| + 10^6, & \text{otherwise.} \end{cases} \quad (4)$$

Here R^2 represents the deviation from the theoretical gamma distribution, $B \geq 0$ is the weighting factor. When the mixing parameter α is a scalar, B equals 0 because the linear combination of two gamma distributions remains a gamma distribution, making regularization unnecessary. For a vector-valued α , B is set to 10^4 to impose a meaningful constraint.

The upper constraint on the optimization parameter α in (3) prevents the cases where the resulting fluid composition is typical for oil rather than gas condensate one.

In this paper, we consider only the synthetic case, in which the compositions of both the actual reservoir fluid and the lean sample are known. The actual data are also generated by the simulating of WT for the given actual reservoir fluid composition. The objective function (4) for each feasible fluid composition is calculated using tNavigator.

Once the mixing parameter is found, the corresponding fluid composition is obtained by (1) or (2). To evaluate the quality of the received fluid, its predicted P_{dew} , liquid dropout curve, Z -factor and gas viscosity from the constant-volume depletion (CVD) simulations are compared with the actual data using the Mean Absolute Percentage Error defined as

$$\text{MAPE} = 100\% \cdot \frac{1}{N'} \sum_{i=1, y_{\text{actual},i} \neq 0}^{N'} \left| \frac{y_{\text{actual},i} - y_{\text{model},i}}{y_{\text{actual},i}} \right|$$

where $y_{\text{actual},i}$ and $y_{\text{model},i}$ are the reference and calculated values at point i , respectively, and N' is the number of points with $y_{\text{actual},i} \neq 0$.

Fluid Composition

In our work, we take the 34-component Lean Gas Condensate (LGC) model with $P_{\text{dew}} = 22.575$ MPa from (Alavian et al., 2014; Hoffmann, 2019) as the lean fluid sample. The LGC contains 9 “light” components including two non-hydrocarbons (N_2 and CO_2), the traditional two-isomers of butanes (iC_4 , nC_4) and pentanes (iC_5 , nC_5). Heavier fractions are represented by the Single-Carbon-Number (SCN) groups ranging from C_6 to C_{30} . The Peng-Robinson EoS is used to describe the gas condensate along with the Lorentz-Bray-Clark viscosity correlation. Following (Alavian et al., 2014;

Hoffmann, 2019), the fluid temperature is set at 93.33 °C. We denote this detailed LGC model as EoS 34 Lean and also assume that it represents the laboratory data with good accuracy.

By adding 0.23 moles of equilibrium condensate per mole of LGC, a Rich Gas Condensate (RGC) model with $P_{\text{dew}} = 29.253$ MPa is created. This RGC model called EoS 34 Rich is regarded as an actual reservoir composition.

In addition to the detailed EoS 34 Lean, we consider 4 reduced-component EoS models with 15, 12, 9, and 6 components. The lumping rules and P_{dew} are listed in Table 1, where EoS 9A and EoS 9B are obtained by different lumping schemes.

We perform the lumping procedure and numerical simulation of CVD experiments using PVT-designer module of tNavigator. It is worth noting that EoS 15, EoS 9A, and EoS 6 models employ the optimal lumping

rules from (Hoffmann, 2019). All lumped models (Table 1) accurately predict CVD test data compared with the detailed EoS 34 Lean model (Figure 1, a–c) and adhere to a gamma distribution for heavy fractions (Figure 1d), where filled points represent the molar fraction and dashed lines – theoretical distributions. For the lumped models, the heavy fractions start from the first component containing C_7 : components C_7 – C_{10} , C_7 – C_8 and C_7 – C_9 for EoS 15, EoS 12, and EoS 9B, respectively; and iC_4 – C_7 for EoS 9A and EoS 6.

Reservoir Simulation

The reservoir simulation model is represented as one horizontal, homogeneous rock layer of $4000 \times 4000 \times 25$ m along the Ox , Oy , and Oz axes, respectively. It contains a centrally located horizontal production well with six transverse, equally spaced planar hydraulic fractures. The reservoir and fracture parameters are listed in Table 2.

EoS model	Lumping Schema	P_{dew} , MPa
EoS 15	$N_2, CO_2, C_1, C_2, C_3, iC_4, nC_4, iC_5, nC_5, C_6, C_7$ – C_{10}, C_{11} – C_{15}, C_{16} – C_{21}, C_{22} – C_{29}, C_{30}	22.416
EoS 12	C_1 – N_2, C_2 – CO_2 – C_3, iC_4 – nC_4, iC_5 – nC_5 – C_6, C_7 – C_8, C_9 – C_{10}, C_{11} – C_{12}, C_{13} – C_{15}, C_{16} – C_{19}, C_{20} – C_{24}, C_{25} – C_{29}, C_{30}	21.482
EoS 9A	N_2, CO_2, C_1, C_2 – C_3, iC_4 – C_7, C_8 – C_{10}, C_{11} – C_{15}, C_{16} – C_{25}, C_{26} – C_{30}	21.658
EoS 9B	C_1 – N_2, C_2 – CO_2 – C_3, iC_4 – nC_4, iC_5 – nC_5 – C_6, C_7 – C_9, C_{10} – C_{15}, C_{16} – C_{23}, C_{24} – C_{29}, C_{30}	22.035
EoS 6	C_1 – N_2, C_2 – CO_2 – C_3, iC_4 – C_7, C_8 – C_{13}, C_{14} – C_{23}, C_{24} – C_{30}	21.522

Table 1. Reduced-component EoS models with their lumping schemes and dew point pressures

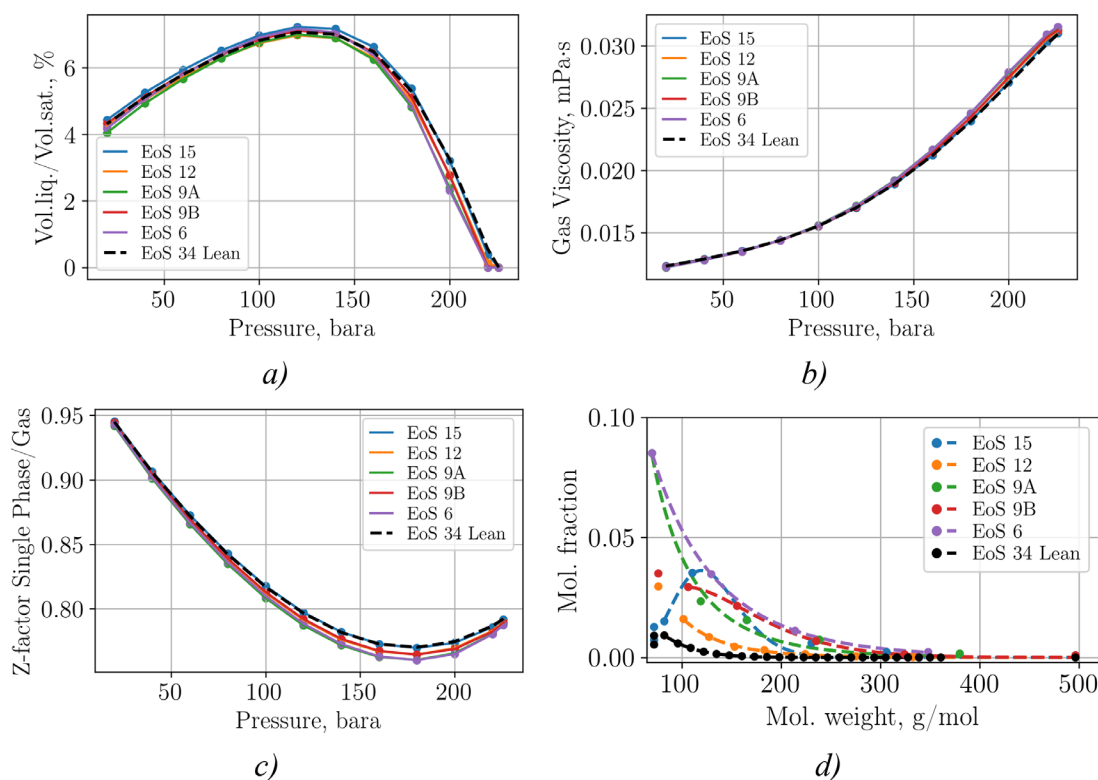


Figure 1. CVD experiment simulations for the detailed and lumped models: a) liquid dropout curve; b) gas viscosity; c) Z-factor; d) gamma distribution for detailed and lumped models. The filled points represent the mole fraction data, the dashed lines – the theoretical distribution

We consider the two-phase flow (gas and condensate) both in the reservoir and fractures to be isothermal and governed by Darcy's law. The water is presented only as an irreducible phase. The relative permeabilities (RPs) for both gas-condensate and water-condensate systems are modeled using Corey correlations. For the sake of simplicity, capillary pressures are neglected. Additionally, the minimal condensate saturation S_{oL} , the minimal (S_{gL}) and critical (S_{gcr}) gas saturations are assumed to be zero. The irreducible water saturation S_{wr} is equal to 0.2, so the maximal gas saturation is found as $S_{gU} = 1 - S_{wr} = 0.8$. To provide the water immobility throughout the well test we set the critical water saturation S_{wcr} to 0.25 and the endpoint of the water RP (k_{rwr}) at minimal condensate and gas saturations (S_{oL} and S_{gL}) to 0.001. The critical condensate saturation for water-condensate displacement is of 0.3, and condensate RP at connate water and minimal gas saturations is equal to 1. The RPs for the gas-condensate system are defined as follows:

$$k_{rg} = 0.5 \left(\frac{S_g}{1 - S_{wcr}} \right)^2, \quad k_{rog} = \left(\frac{1 - S_g}{1 - S_{wcr} - S_{ogcr}} \right)^2,$$

where k_{rg} , k_{rog} denote the RPs for gas and condensate, respectively; S_g is a gas saturation, and $S_{ogcr} = 0.2$ is the critical condensate saturation.

The producer is controlled by the bottomhole pressure (P_{bh}) assumed to be uniform along the horizontal part of the wellbore, neglecting friction. The WT is conducted in three stages: 1) first stage (4 days) with $P_{bh} = 16.5$ MPa; 2) second stage (3 days) with $P_{bh} = 19.5$ MPa; 3) final stage (5 days) with $P_{bh} = 15$ MPa.

The reservoir sector is assumed to be isolated with no-flow boundary conditions. The model is initialized under the equilibrium conditions with the gas-condensate contact (GOC) at the reservoir top of 3000 m and the water-condensate contact (WOC) positioned at the foot bed. Also, the one-stage separator is implemented, where the reservoir fluid is immediately brought to the standard surface conditions of 20 °C and 0.101325 MPa.

The grid size near the wellbore and fractures is of 10 m, increasing to 200 m at the domain boundaries. Figure 2 demonstrates the domain discretization and the simulated pressure field distribution in the Oxy plane in

the vicinity of the fractures with the drawdown regions at some moment of the WT.

Optimization Algorithms

Mathematically, any optimization problem can be written in the following form.

Find $\mathbf{x} = (x_1, x_2, \dots, x_D)$ to optimize $f(\mathbf{x})$, subject to $\mathbf{x} \in X$,

where “optimize” means to minimize or maximize, $f(\mathbf{x})$ is an objective function (or fitness function), X is a search subspace of $\mathbb{R}^D$ taking into account the constraints on the vector of free optimization variables $\mathbf{x}$.

DE method proposed in (Storn, Price, 1997; Das, Suganthan, 2011) belongs to the class of evolutionary algorithms, where a set of potential solutions is called a population, and iterations are called generations. DE starts by initializing a random population and evaluating its members by the fitness function values. In standard DE, the transition between generations is governed by three fundamental operations: mutation, crossover, and selection. In this paper, we apply one of the popular strategies of DE, named “best/1/bin”. “Best/1” means that the mutation operator uses the best vector of the population and 1 pair of other vectors to produce the mutation vector: $\mathbf{x}_m^i = \mathbf{x}_{best}^i + F(\mathbf{x}_1^i - \mathbf{x}_2^i)$. Here i denotes the generation number, $\mathbf{x}_{best}^i$ is the best vector in the population at i -th generation, $\mathbf{x}_1^i$ and $\mathbf{x}_2^i$ are randomly selected vectors from the population, and $F \in [0, 1]$ is a constant mutation factor. “Bin” means that DE applies the binomial crossover to construct a trial vector, $\mathbf{x}_{trial}^i$,

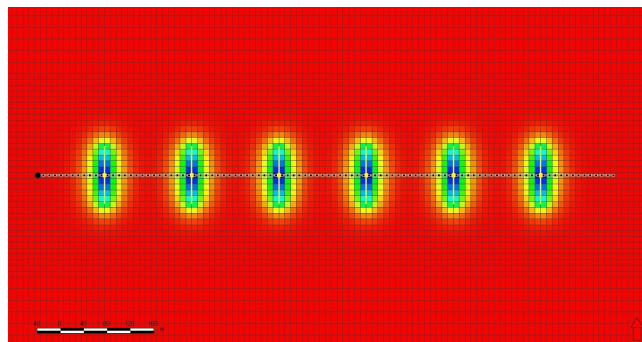


Figure 2. Reservoir layout and well configuration in the Oxy plane with the pressure distribution colormap at some time moment

Parameter	Value	Parameter	Value
Reservoir depth H , m	3000	Reservoir temperature T_{res} , °C	93.33
Wellbore length L , m	1000	Reservoir pressure P_{res} , MPa	30
Reservoir thickness h , m	25	Fracture permeability k_f , mD	300
Porosity ϕ	0.2	Fracture half-length x_f , m	50
Permeability $k_x = k_y = k_z$, mD	0.1	Fracture width w_f , m	0.03
Net-to-gross ratio NTG	0.5		

Table 2. The reservoir and fracture parameters in the hydrodynamic model

using $\mathbf{x}_m^i$ and a target vector, $\mathbf{x}_p^i$, which is randomly selected from the population and differs from the vectors used in the mutation operator. Given the crossover rate, $CR \in [0, 1]$, we randomly select a coordinate index j_{rand} from $[1, 2, \dots, D]$, and a number r from $[0; 1)$. Then the j -th element $\mathbf{x}_{trail}^i$ is determined as follows. If $j = j_{rand}$ or $r \leq CR$, then $\mathbf{x}_{trail,j}^i = \mathbf{x}_{m,j}^i$, and $\mathbf{x}_{trail,j}^i = \mathbf{x}_{p,j}^i$, otherwise. Finally, the selection operator compares the trial vector with the target one and keeps the “better”, e.g. with the better fitness function value. This procedure is repeated for each individual in the population.

The PSO method (Kennedy, Eberhart, 1995) is based on the simulation of the bird’s behavior within a flock. The set of potential solutions is called a swarm of particles. The i -th particle ($i = 1, \dots, N$) is defined by its position $\mathbf{x}_i$ in the search space, its velocity $\mathbf{v}_i$, and its fitness function $f(\mathbf{x}_i)$. PSO, as well as DE, starts with generating an initial swarm. The position of particle $\mathbf{x}_i$ changes according to its own experience and the experience of the entire swarm as follows:

$$\mathbf{x}_i(t+1) = \mathbf{x}_i(t) + \mathbf{v}_i(t+1) \quad (5)$$

$$\mathbf{v}_i(t+1) = C_{in}\mathbf{v}_i(t) + C_{cog}r_1[\mathbf{x}_{pbest,i}(t) - \mathbf{x}_i(t)] + C_{soc}r_2[\mathbf{x}_{gbest,i}(t) - \mathbf{x}_i(t)]$$

Here t indicates a pseudo-time representing the iteration step; $\mathbf{x}_i(t)$ is the particle’s position at the time t ; $\mathbf{v}_i(t)$ is the velocity of the i -th particle at the time t ; $\mathbf{x}_{pbest,i}(t)$ is the personal best position for the i -th particle at the time t ; $\mathbf{x}_{gbest,i}(t)$ is the best position for the whole swarm at the time t ; C_{in} is the inertia factor, which shows the influence of the previous direction; C_{cog} is a cognitive coefficient reflecting the aspiration of the particle to move towards its best position; C_{soc} is a social coefficient that corresponds to the movement towards the most successful particle in the swarm; r_1 and r_2 are random numbers uniformly distributed in $(0; 1)$.

NOMAD from (Digabel, 2011) is a deterministic algorithm, that explores the search space by adaptively adjusting a mesh according to the previous iteration outcomes. At the beginning, the algorithm is initialized with a starting point as the current best solution, along with a mesh of a base mesh size. At each iteration, NOMAD generates several trial points by shifting the current solution along chosen directions such as standard coordinate axes, random orientations, or orthogonal vectors. If one of the trial points improves the current solution, it is taken as the new incumbent, and the mesh may be locally coarsened or left unchanged to facilitate the broader exploration. Otherwise, the mesh size is reduced, prompting a more detailed local exploration around the incumbent. This adaptive mechanism effectively balances global search and local refinement, reducing the risk of convergence to local minima. Each iteration consists of two complementary phases: a broad

search phase for preliminary exploration of promising solutions, followed, if necessary, by a detailed poll phase that intensively examines the neighborhood around the current best solution. The search terminates automatically when the mesh size becomes sufficiently small, suggesting the convergence is reached. In our work the Python NOMAD implementation PyNomad is used.

Local search algorithm is an iterative optimization procedure that begins with an initial feasible solution and progressively improves it through successive local modifications. The algorithm initializes by defining a starting point in the search space as the current solution. At each iteration, it systematically generates neighboring solutions according to a certain rule, then evaluates each of them by the objective function value and selects the best one. If the best solution in the neighborhood yields an improvement over the current solution, it replaces the current solution for the next iteration.

Results and Discussion

In this section we try to restore the actual fluid composition from the lean sample, EoS 34 Lean, by solving the optimization problem (3).

The actual GCR is found by using EoS 34 Rich for the well test simulation. We use not only EoS 34 Lean as a start composition but also its reduced-component versions (Table 1).

It should be noted that the aim is not to reproduce exactly the actual composition, but to find a composition that has the same PVT properties as EoS 34 Rich.

To solve the optimization problem, we implement a program in Python that calls the console version of the tNavigator to determine the gas and condensate flow rates and compute the objective function (4). All the parameters used in the optimization methods are determined after the set of test calculations, which are omitted here. The direction type for the poll phase in the optimization algorithm NOMAD is set to ORTHO 2N, which corresponds to the OrthoMADS strategy. It means that 2n mutually orthogonal polling directions forming a maximal positive basis in $\mathbb{R}^n$ is constructed. The swarm size for PSO is set to 50 particles. Coefficients C_{in} , C_{cog} and C_{soc} in formula (5) are considered to be constant values chosen as follows: $C_{in} = 0.5$, $C_{cog} = C_{soc} = 2.0$. The maximal possible velocity is set to 0.1. For DE the population size is of 100 vectors, the mutation factor F and CR rate are of 0.2 and 0.8, respectively. For LS the number of neighbors is equal to 20 and the number of iterations is set to 100.

The restored fluid composition calculations are performed and the resulting fluid compositions are denoted by adding “R” to the name of the model, e.g. EoS 15-R.

The Case of Scalar Mixing Parameter

In this section we consider the case of scalar α in order to find the actual fluid composition according to (1). The NOMAD optimization algorithm is applied with an initial guess of $\alpha^0 = 0$, corresponding to the lean sample composition as the starting point. Table 3 shows the obtained values of α , the final objective function values, MAPE for P_{dew} and CVD tests including the liquid dropout curve, single-phase Z-factor and gas viscosity.

As expected, the restored EoS 34 Lean-R model accurately reproduces both the P_{dew} and CVD test results of the reference EoS 34 Rich model. In this case, the obtained α is close to the value that we used to generate the RGC (0.23 moles of equilibrium condensate per a mole of the gas). Consequently, the compositions of the actual EoS 34 Rich and the restored EoS 34-R are nearly identical, as shown in Figure 3 plotted in a logarithmic scale on the y-axis.

Among all of the lumped schemes, EoS 9B-R exhibits the best predictions for P_{dew} and CVD test simulation. The largest discrepancies are observed for EoS 6-R. Figures 4a, 4b, and 4c compare the gas and condensate flow rates in standard cubic meters per day (Sm^3/d), as well as GCR (Sm^3/Sm^3) obtained for actual (EoS 34 Rich) and restored fluid compositions, respectively.

The numerical analysis demonstrates that the method of restoring the fluid composition by adding equilibrium condensate to the lean sample is effective for the detailed compositional model. However, the results for reduced-component models depend heavily on the particular lumping rule applied. For example, EoS 9B-R model exhibits an acceptable level of accuracy whereas other models show unacceptable errors in predicting liquid dropout curve.

The Case of Vector-Valued Mixing Parameter

For a more accurate reproduction of composition properties, we employ a vector-valued mixing parameter α expanding our optimization toolkit beyond NOMAD. In this case the weighting factor B in formula (4) is set

to 10^4 , ensuring the physical correctness of the received composition.

We illustrate this approach for reduced-component models only. NOMAD is considered in two variants, NOMAD₁ and NOMAD₂, which differ in their choice of an initial guess for the iterations. NOMAD₁ starts from the lean sample composition, i.e., the initial α^0 is a zero vector. NOMAD₂ initializes α^0 with all elements equal to the scalar solution α (Table 3). For the LS algorithm the initial point is selected as in NOMAD₂.

The final objective function values for each of the algorithms are listed in Table 4. All methods can find such a fluid composition that, when used in the well test simulation, produces a GCR profile close to the actual one.

The CVD test results for the EoS models that performed poorly with a scalar α (EoS 15-R, 9A-R, 6-R, and 12-R) are represented in Figures 5–7. These figures compare the restored compositions predicted by all optimization algorithms with the actual composition, highlighting the scalar α case.

Note that EoS 12-R, obtained using the DE algorithm, accurately predicts the Z-factor and gas viscosity. However, it produces significant errors (exceeding 10%) when simulating the liquid dropout curve in the CVD test (Figure 5d). The PSO-derived EoS 6-R model approximates the liquid drop out curve well (Figure 5c),

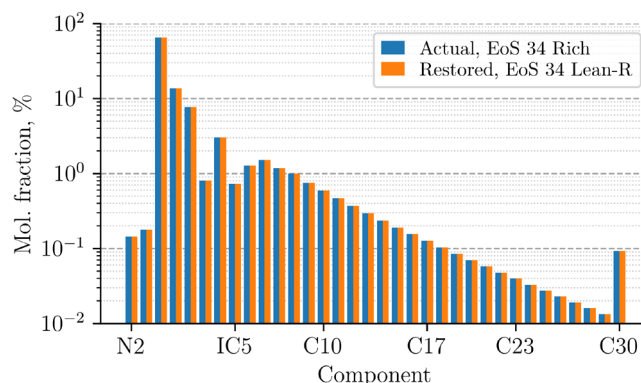


Figure 3. Molar composition of the actual, EoS 34 Rich, and restored, EoS 34 Lean-R models

Model	α	GCR rMSE	P_{dew} MAPE	Liq.dropout MAPE	Z-factor MAPE	Gas viscosity MAPE
EoS 34 Lean-R	0.23075	$2.379 \cdot 10^{-10}$	$7.9 \cdot 10^{-7}$	$1.58 \cdot 10^{-3}$	$4.115 \cdot 10^{-6}$	$8.978 \cdot 10^{-5}$
EoS 15-R	0.17167	0.082	2.27	19.75	0.09	1.02
EoS 12-R	0.18784	0.053	2.63	18.48	0.43	1.32
EoS 9A-R	0.17086	0.444	9.432	26.82	0.84	2.52
EoS 9B-R	0.32504	$3.295 \cdot 10^{-4}$	0.424	4.491	1.13	3.21
EoS 6-R	0.13911	1.034	11.493	35.97	0.84	2.98

Table 3. The solution α , objective function values, MAPE of P_{dew} , CVD liquid dropout curve, single-phase Z-factor and the gas viscosity

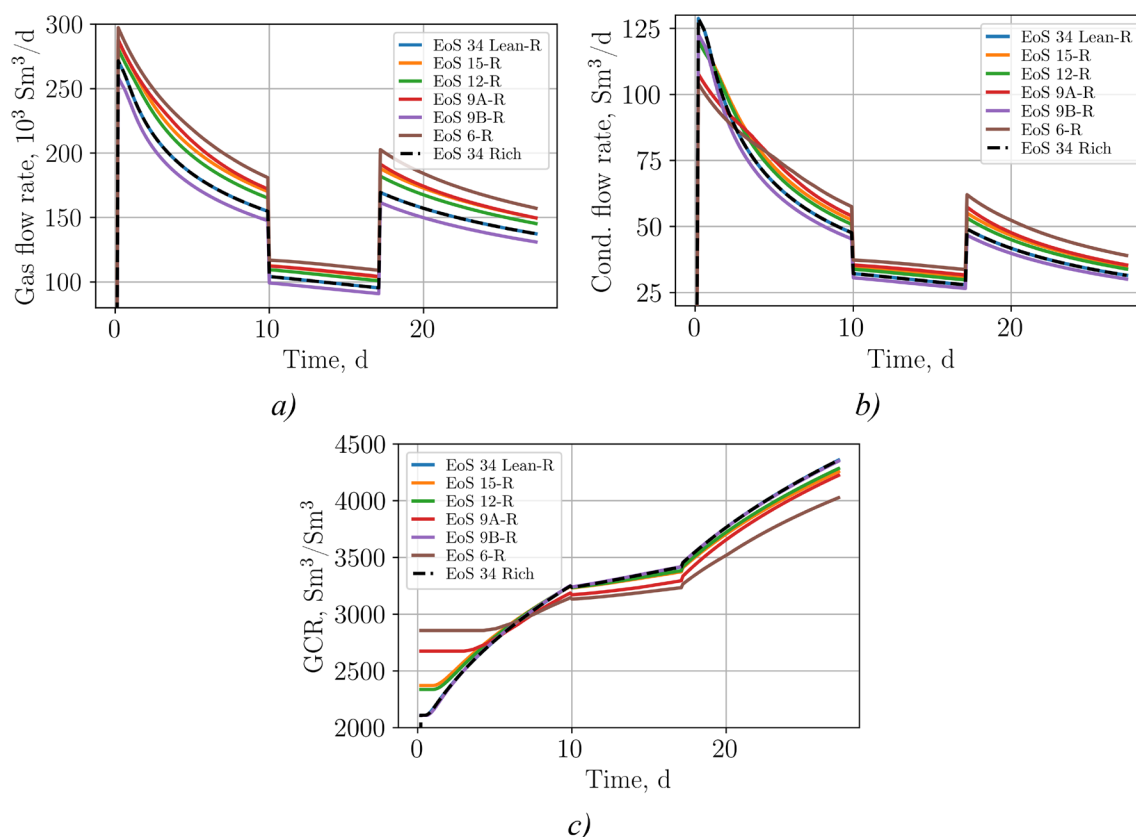


Figure 4. The WT simulation results for actual (dashed black line) and restored fluid compositions for scalar α : a) gas flow rates; b) condensate flow rates; c) GCR

	GCR errors				
	NOMAD ₁	NOMAD ₂	PSO	DE	LS
EoS 15-R	$1.297 \cdot 10^{-3}$	$1.537 \cdot 10^{-5}$	$2.29 \cdot 10^{-5}$	$1.924 \cdot 10^{-3}$	$3.91 \cdot 10^{-4}$
EoS 12-R	$4.218 \cdot 10^{-6}$	$1.075 \cdot 10^{-6}$	$1.01 \cdot 10^{-5}$	$1.953 \cdot 10^{-4}$	$2.918 \cdot 10^{-3}$
EoS 9A-R	0.014	$4.14 \cdot 10^{-3}$	$1.266 \cdot 10^{-3}$	$3.907 \cdot 10^{-3}$	$6.17 \cdot 10^{-3}$
EoS 9B-R	$7.513 \cdot 10^{-5}$	$8.959 \cdot 10^{-6}$	$5.875 \cdot 10^{-5}$	$1.531 \cdot 10^{-5}$	$3.295 \cdot 10^{-3}$
EoS 6-R	$4.547 \cdot 10^{-3}$	$3.26 \cdot 10^{-3}$	$1.523 \cdot 10^{-3}$	$3.731 \cdot 10^{-3}$	0.092

Table 4. Objective function values for optimization algorithms for vector-valued α

but shows a larger discrepancy of 7% in gas viscosity simulation (Figure 6c), although its Z factor prediction remains within 5%.

Another important characteristic of a fluid sample is the potential condensate content, PC_5+ , representing the amount of C_5+ hydrocarbons that can be obtained from a unit volume of reservoir gas under standard surface conditions (+20 °C and 0.101325 MPa). It is measured in grams per cubic meter (g/m³) and calculated using the formula

$$PC_5+ = \frac{10}{24.04} \sum_{n=5}^N M_{Cn} z_{Cn}, \quad (6)$$

where n denotes SCN of a hydrocarbon component C_n ; z_{Cn} and M_{Cn} are its molar fraction (in percent) and molecular weight (g/mole), respectively; 24.04 is the

molar volume (dm³/mole) at standard surface conditions. Note that formula (6) calculates the value of PC_5+ for reservoir gas.

Figure 8 shows the absolute percentage errors in calculated P_{dew} and PC_5+ content (based on relationship (6)) for the restored compositions, compared to their actual values of 29.252 MPa and 489.41 g/m³, respectively. A logarithmic scale is used to enhance the visualization of the errors. Note that the lumping rules for EoS 9A and EoS 6 do not support PC_5+ calculations, therefore comparisons for PC_5+ are only shown for EoS 15-R, EoS 12-R, and EoS 9B-R.

The results reveal that the predicted P_{dew} and calculated PC_5+ content for reservoir gas for all compositions achieves acceptable accuracy, with deviations below 10%.

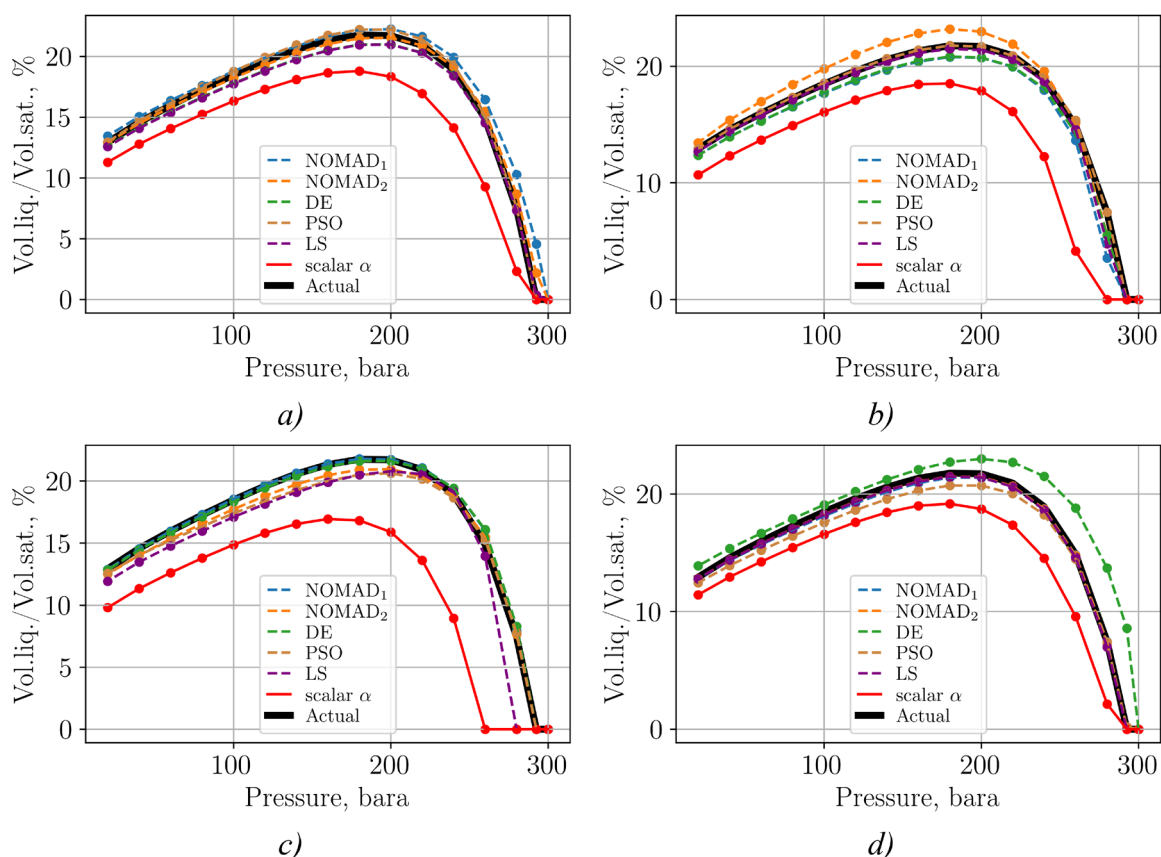


Figure 5. Condensate dropout curves for restored compositions obtained by NOMAD₁, NOMAD₂, DE, PSO, LS, and for the scalar α case: a) EoS 15-R; b) EoS 9A-R; c) EoS 6-R; d) EoS 12-R

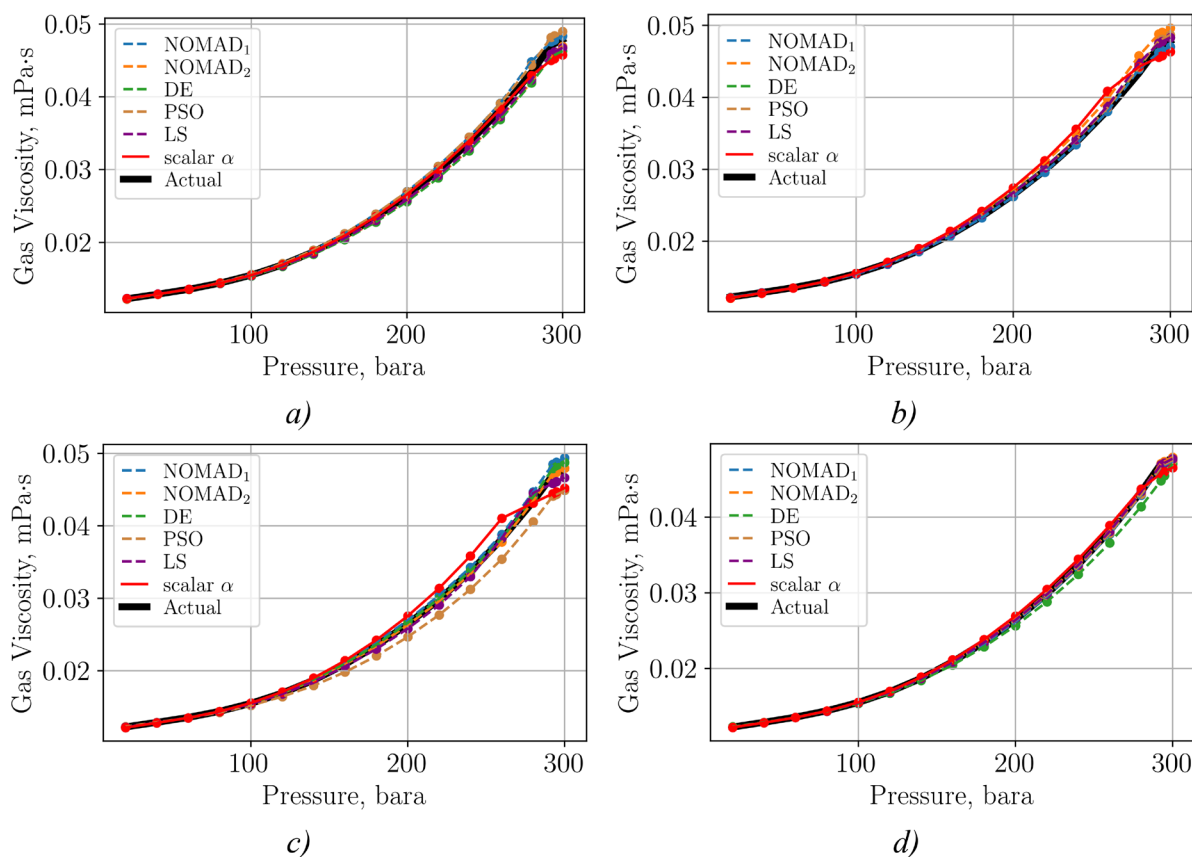


Figure 6. Gas viscosity for restored compositions obtained by NOMAD₁, NOMAD₂, DE, PSO, LS, and for the scalar α case: a) EoS 15-R; b) EoS 9A-R; c) EoS 6-R; d) EoS 12-R

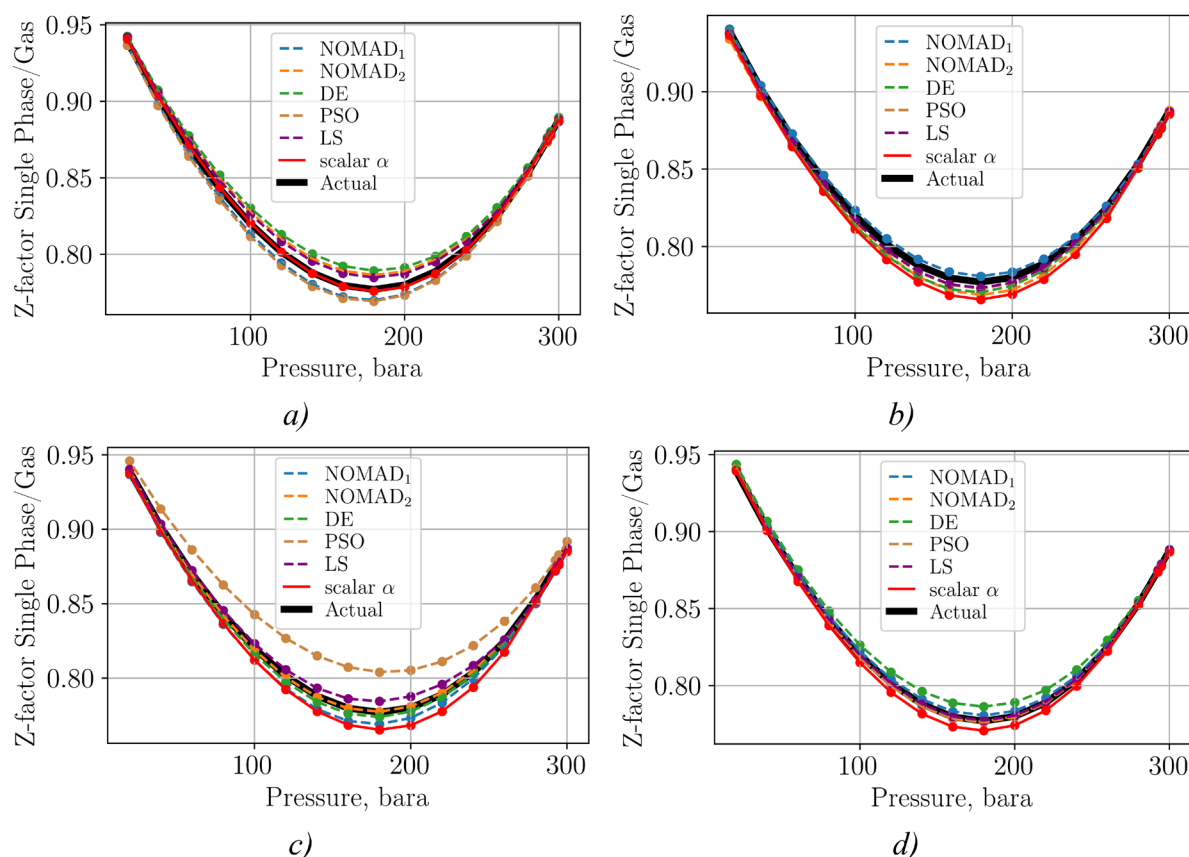


Figure 7. Single-phase Z-factor for restored compositions obtained by NOMAD₁, NOMAD₂, DE, PSO, LS, and for the scalar α case: a) EoS 15-R; b) EoS 9A-R; c) EoS 6-R; d) EoS 12-R

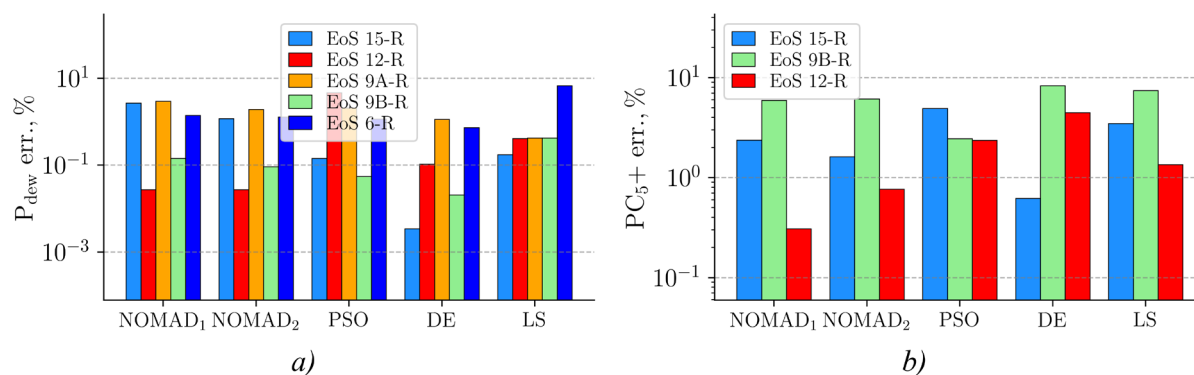


Figure 8. Absolute percentage errors in P_{dew} (a) and PC_5+ (b) for the restored compositions relative to the actual data

Also, the accuracy in restoring both P_{dew} and PC_5+ values depends not only on the optimization method used but also on the EoS model including the lumping rules and the number of components.

Correctness Analysis

In this section we evaluate the correctness of the proposed method of restoring the actual fluid composition by modeling situations where actual well test data contain uncertainties and errors. The correctness test is conducted for EoS 12 model and using the

NOMAD₂ optimization method. As a reference case, we use the solution of the optimization problem (3) with a vector-valued parameter α , EoS 12-R. The reference GCR is obtained from well test simulations using the reference EoS.

We perturb the actual GCR data, $GCR_{\text{actual}}(t)$, with synthetic noise and analyze the solution of the optimization problem. In inverse problem theory, this approach is known as a model resolution test (Aster et al., 2018), which helps characterize the bias and stability of the inverse solution.

We generate three noisy datasets by adding zero-mean Gaussian noise ε_i , $i = 1, 2, 3$, with standard deviations of $\sigma_1 = 0.1$, $\sigma_2 = 0.2$, and $\sigma_3 = 0.4$, representing low, medium and high noise level, respectively

$$\text{GCR}_{\text{noisy}}(t) = \text{GCR}_{\text{actual}}(t) + \varepsilon_i(t), \quad i = 1, 2, 3.$$

The optimization problem (3) is then solved for each case, and the properties of the obtained compositions are compared to actual ones. Figure 9 depicts the reference GCR, the noisy data, and the simulated GCR profiles calculated using the optimization solution compositions, $\text{GCR}(t)$, and denoted according to the noise level (e.g., “Opt. 0.1”).

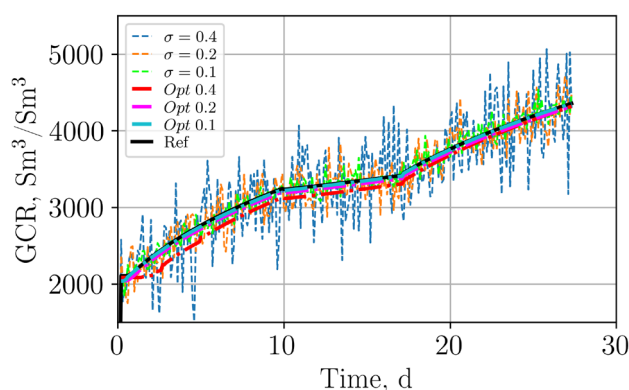


Figure 9. Comparison of the reference GCR, the noisy GCR, and GCR obtained using the optimization solution composition for three noise levels: low ($\sigma = 0.1$), medium ($\sigma = 0.2$), and high ($\sigma = 0.4$)

Table 5 presents error metrics (rMSE, MAPE, and maximum absolute error) comparing both $\text{GCR}_{\text{noisy}}(t)$ and the simulated $\text{GCR}(t)$ against the true values $\text{GCR}_{\text{actual}}(t)$. The maximum absolute error (AE) is the largest amplitude deviation defined as $\max \text{AE} = \max |\text{GCR}_{\text{actual}}(t) - \text{GCR}(t)|$.

The data in Table 5 show that the MAPE of $\text{GCR}_{\text{noisy}}$ remains below 10% for any noise level, even when the maximum deviation amplitude for $\sigma = 0.4$ exceeds 36% of the average $\text{GCR}_{\text{actual}}$ of $3374.724 \text{ Sm}^3/\text{Sm}^3$. Note, that an error level below 10% is generally considered acceptable for field data. For the simulated $\text{GCR}(t)$, rMSE metric represents the final objective function value which increases as the standard deviation rises. The value of max AE and the MAPE for all noise levels confirm that obtained $\text{GCR}(t)$ achieves acceptable accuracy relative to $\text{GCR}_{\text{actual}}(t)$.

Figure 10 illustrates the effect of noise level on the properties of the restored fluid compositions, labeled according to the noise level (e.g., “EoS $\sigma = 0.1$ ”). The condensate dropout curves from CVD simulations along with histograms for P_{dew} and PC_{5+} content for reservoir gas demonstrate that discrepancies from the actual EoS (EoS 34 Rich) remain within acceptable accuracy for considered noise levels.

Figure 11 represents the molar fraction (%) of the compositions obtained for different noise level. It is worth noting that the differences in component mole fractions become significant only for heavier components, such as C_{20} - C_{24} and C_{25} - C_{29} . However, their

Noise st. deviation	$\text{GCR}_{\text{noisy}}$			simulated		$\text{GCR}(t)$
	rMSE	MAPE	max AE, Sm^3/Sm^3	rMSE	MAPE	max AE, Sm^3/Sm^3
$\sigma=0$	0	0	0	$4.218 \cdot 10^{-6}$	0.013	2.456
$\sigma=0.1$	0.094	2.326	329.269	$2.699 \cdot 10^{-3}$	0.355	71.993
$\sigma=0.2$	0.382	4.879	603.934	0.0208	1.38	66.776
$\sigma=0.4$	1.479	9.429	1207.151	0.1861	3.67	266.148

Table 5. Error metrics of $\text{GCR}_{\text{noisy}}$ and simulated $\text{GCR}(t)$ with respect to $\text{GCR}_{\text{actual}}(t)$

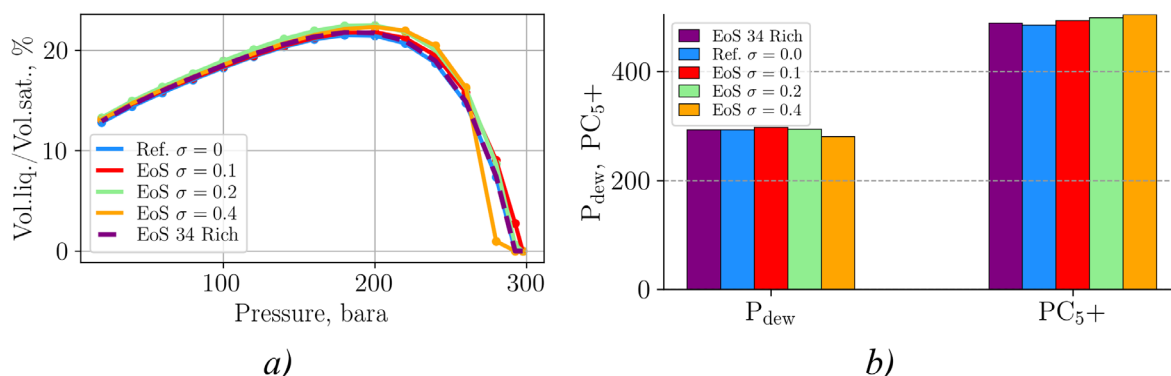


Figure 10. Comparison of the actual composition (EoS 34 Rich) and restored compositions for each noise level: a) condensate dropout curves from CVD simulation; b) P_{dew} and PC_{5+} content

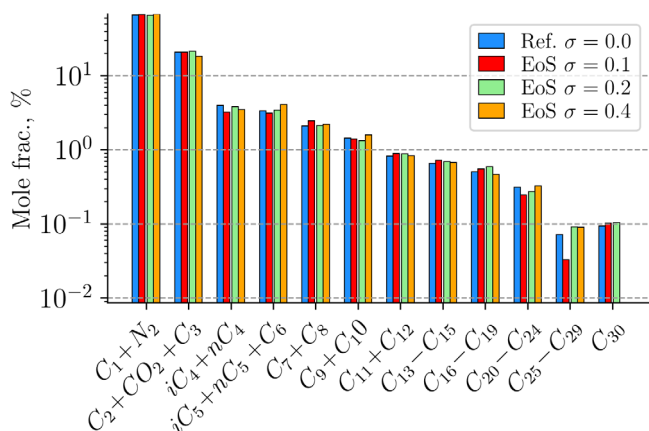


Figure 11. Molar fraction of the compositions obtained for the different noise levels

total contribution to PC_5+ content is relatively small and does not significantly affect its final value.

Conclusion

The paper proposes a method using optimization algorithms for determining gas-condensate reservoir composition for cases where only a lean sample of reservoir fluid and well test data are available. Given the calibrated EoS against the lean sample, the actual reservoir fluid is obtained by minimizing the discrepancy between measured and simulated GCR performed using tNavigator.

The required reservoir fluid is sought as a linear combination of lean gas and its equilibrium condensate compositions. Both scalar and vector-valued proportionality coefficients are considered. The scalar mixing parameter corresponds to the standard approach in which the lean gas is mixed with its equilibrium condensate. The vector-valued proportionality coefficient makes it possible to calibrate the amount of each component individually which may not follow a linear trend relative to the condensate composition. In this case, we introduce a check to confirm that the mole fraction as a function of molecular weight follows a gamma distribution.

The method is verified in a synthetic case for which the reservoir fluid is known. While reduced-component (lumped) EoS models are common in practice, the study regards both a detailed 34-component model and its lumped versions (15, 12, 9, and 6 components).

Numerical results demonstrate that for the 34-component model, the scalar mixing parameter approach reproduces the reference dew point pressure, condensate dropout curve, Z-factor and gas viscosity with good accuracy. For lumped models the accuracy of obtained fluids depends on the lumping scheme and the number of components.

For reduced-component EoS, applying a vector-valued mixing parameter method improves the accuracy

in reproducing the reference PVT data. In this case various optimization methods are tested, and it is shown that all of them achieve acceptable fidelity in recovering the actual fluid composition.

To analyze the correctness a model resolution test is conducted by perturbing the actual GCR with zero-mean Gaussian noise at low, medium and high noise level. The results show that if the absolute percentage error between noisy and actual CGR remains below 10%, the method can reliably determine the reservoir fluid composition with sufficient accuracy. It should also be emphasized that this approach requires further research and validation with real well test data.

Data Availability Statement

Supplementary materials, including EoS 34 Rich, EoS 34 Lean, its reduced-components versions and obtained compositions can be accessed online at <https://doi.org/10.5281/zenodo.17457522>

Acknowledgements

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Project No. FSUS-2025-0016), and by the NSU Advanced Engineering School.

The authors thank Rock Flow Dynamics for providing an academic license for the tNavigator simulator.

References

- Alavian S. A., Whitson C. H., Martinsen S. O. (2014). Global component lumping for eos calculations. *SPE annual technical conference and exhibition*, Amsterdam, the Netherlands. P. 170912-MS. <https://doi.org/10.2118/170912-MS>
- API recommended practice for sampling petroleum reservoir fluids (2003). Second ed. N.Y.: API Publishing Services.
- Aster R., Borchers B., Thurber C. (2018). Parameter estimation and inverse problems (3rd ed.). Amsterdam: Elsevier. <https://doi.org/10.1016/B978-0-12-804651-7.00015-8>
- Brusilovskiy A., Yushchenko T. (2016). Two-phase deposits: Methodology approach to the identification of composition and pvt properties of reservoir hydrocarbon fluids using limited initial information. *PROneft. Professionally about Oil*, (1), pp. 68–74. (In Russ.)
- Das S., Suganthan P.N. (2011). Differential evolution: A survey of the state-of-the-art. *IEEE Transactions on Evolutionary Computation*, 15(1), pp. 4–31. <https://doi.org/10.1109/TEVC.2010.2059031>
- Digabel S.L. (2011). Algorithm 909: Nomad: Nonlinear optimization with the mads algorithm. *ACM Transactions on Mathematical Software*, 37(4), 44. <https://doi.org/10.1145/1916461.1916468>
- Elsharkawy A.M. (2002). Predicting the dew point pressure for gas condensate reservoirs: Empirical models and equations of state. *Fluid Phase Equilibria*, 193(1–2), pp. 147–165. [https://doi.org/10.1016/S0378-3812\(01\)00724-5](https://doi.org/10.1016/S0378-3812(01)00724-5)
- Gimazov A., Imomnazarov B., Starovoytova B., Baykin A., Babin V., Khamidullin D., Kuporosov D. (2024). Solution of the inverse problem of determining the initial hydrocarbons composition in a gas-condensate reservoir using field data. *Georesursy = Georesources*, 26(3), pp. 73–86. (In Russ.) <https://doi.org/10.18599/grs.2024.3.9>
- Hoffmann A. (2019). Eos lumping optimization using a genetic algorithm and a tabu search. *Journal of Petroleum Science and Engineering*, 174, pp. 495–513. <https://doi.org/10.1016/j.petrol.2018.11.021>
- Kennedy J., Eberhart R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 – International conference on neural networks*, 4, pp. 1942–1948 <https://doi.org/10.1109/ICNN.1995.488968>
- Osfouri S., Azin R. (2015). An overview of challenges and errors in sampling and recombination of gas condensate fluids. *Journal of Oil, Gas*

and *Petrochemical Technology*, 3(1), pp. 1–14. <https://doi.org/10.22034/JOGPT.2016.43155>

Promzelev I., Brusilovsky A., Kuporosov D., Yushchenko T. (2018). Peculiarities of identification of reservoir fluids properties of two-phase with oil rim and gas cap deposits. *SPE Russian petroleum technology conference*, Moscow, Russia. SPE-191566-18RPTC-MS. <https://doi.org/10.2118/191566-18RPTC-MS>

R Gazprom 086–2010. (2011). Instruction for comprehensive gas and gas condensate well studies. In 2 Parts. Moscow: Gazprom. (In Russ.)

Storn R., Price K. (1997). Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11(4), pp. 341–359. <https://doi.org/10.1023/A:1008202821328>

tNavigator 23.1 (2023). Simulator. Technical Manual, RFD: Rock Flow Dynamics, 3855 p.

Whitson C.H. (1983). Characterizing hydrocarbon-plus fractions. *Soc. Petrol. Eng. J.*, 23, pp. 683–694. <https://doi.org/10.2118/12233-PA>

About Authors

Botagoz N. Starovoytova – PhD (Physics and Mathematics), senior researcher, Novosibirsk State University

1, Pirogova st., Novosibirsk, 630090 Russian Federation

e-mail: b.starovoitova@nsu.ru

Buned Kh. Imomnazarov – Junior Researcher, Novosibirsk State University

1, Pirogova st., Novosibirsk, 630090 Russian Federation

e-mail: b.imomnazarov@g.nsu.

Alexey N. Baykin – PhD (Physics and Mathematics), Head of the Laboratory for optimizing hydrocarbon production software systems, Novosibirsk State University

1, Pirogova st., Novosibirsk, 630090 Russian Federation

e-mail a.baikin@g.nsu.ru

Manuscript received 17 June 2025;

Accepted 27 November 2025;

Published 30 June 2026

© 2026 The Authors. This article is published in open access under the terms and conditions of the Creative Commons Attribution (CC BY) License (<https://creativecommons.org/licenses/by/4.0/>)