A Domain-Driven, Physics-Backed, Proximity-Informed AI Model for PVT Predictions—Part II: Differential Liberation Expansion and Viscosity Tests

Sofianos Panagiotis Fotias , Eirini Maria Kanakaki * , Afzal Memon , Anna Samnioti , Jahir Khan , John Nighswander , Vassilis Gaganis

Posted Date: 10 March 2026

doi: 10.20944/preprints202603.0729.v1

Keywords: differential liberation expansion (DLE); viscosity test; pressure-volume-temperature (PVT) properties; reservoir fluids; local interpolation model; AI model; thermodynamic behavior; machine learning

 doi:10.20944/preprints202603.0729.v1

Article

A Domain-Driven, Physics-Backed, Proximity-Informed AI Model for PVT Predictions—Part II: Differential Liberation Expansion and Viscosity Tests

Sofianos Panagiotis Fotias [1,2], Eirini Maria Kanakaki [1,2,*], Afzal Memon [1], Anna Samnioti [1,2], Jahir Khan [1], John Nighswander [1] and Vassilis Gaganis [1,2]

[1] Fluidsdata, 1253 91 St SW unit 102, Edmonton, AB, Canada

[2] School of Mining and Metallurgical Engineering, National Technical University of Athens, 15780 Athens, Greece

* Correspondence: ekanakaki@metal.ntua.gr

Abstract

Differential Liberation Expansion (DLE) and viscosity tests are core elements of the Pressure–Volume–Temperature (PVT) laboratory suite used to characterize reservoir oils under depletion and to support compositional modeling and reservoir simulation. Nevertheless, both DLE and viscosity testing remain expensive and time-consuming due to specialized equipment, strict operating procedures, and the need for experienced laboratory personnel. Building on our prior work that introduced the proximity-informed Local Interpolation Model (LIM) framework for Constant Composition Expansion (CCE), this study demonstrates how the same end-to-end, neighborhood-based workflow is applied to DLE and viscosity test data. A target fluid is embedded in a compositional–thermodynamic descriptor space and paired with a small set of thermodynamically similar fluids drawn from a PVT data archive. Within this locality, LIM is used to infer DLE behavior by combining local interpolation for key scalar quantities (e.g., saturation-point and endpoint PVT values) with shape-preserving reconstruction of pressure-dependent curves. For viscosity, the same approach reconstructs the oil-viscosity curve $\mu_o(p)$ across the undersaturated and saturated regions. Evaluation on a proprietary database of DLE and viscosity tests shows strong agreement across diverse fluids for both DLE and oil viscosity trends. This supports reducing reliance on new DLE and viscosity measurements while maintaining engineering-grade fidelity in reservoir-engineering and simulation workflows. This approach has been fully automated through software so it can be set up and directly utilized by the field operators on their own databases to significantly reduce their fluid sampling and laboratory analysis costs. Moreover, the proposed AI model does not use others' data while respecting data privacy and data ownership.

Keywords: differential liberation expansion (DLE); viscosity test; pressure-volume-temperature (PVT) properties; reservoir fluids; local interpolation model; AI model; thermodynamic behavior; machine learning

1. Introduction

Reliable fluid characterization underpins essentially every quantitative decision in petroleum engineering[1–5]. Volumetric estimates, production forecasting, surface-network and separation design, and the predictive performance of reservoir simulators all depend on PVT descriptions that are internally consistent and representative of the field's depletion pathway. The pressure-dependent evolution of phase volumes and phase compositions controls gas liberation, oil shrinkage, and density and viscosity changes, and therefore governs the multiphase-flow response that ultimately drives well performance and oil recovery. For this reason, PVT properties are not merely inputs to

 doi:10.20944/preprints202603.0729.v1

simulation. They provide constitutive constraints that shape thermodynamic behavior in flow models and, by extension, directly affect the credibility of engineering predictions[6–11].

In practice, PVT behavior enters reservoir and production-system workflows through model-ready representations, whether in tabulated form for fast engineering calculations or via equation-of-state-based compositional descriptions when additional thermodynamic detail is applied. In both cases, laboratory PVT experiments provide the reference information: they establish pressure-dependent property behavior that is transferred into simulation inputs and they serve as the calibration targets used to tune thermodynamic models[12] so that predicted phase behavior and volumetric response remain consistent with measurements[6,7,9,10,13].

Within the standard PVT test suite, the DLE experiment (also termed Differential Vaporization) is particularly important for oil reservoirs depleted below saturation pressure at reservoir temperature[6,7,14]. In a DLE test, pressure is reduced stepwise, phases are allowed to equilibrate, and the evolved gas is removed at each stage. Consequently, the composition of the remaining oil evolves continuously along the depletion path. This depletion mechanism makes DLE indispensable for deriving the pressure-dependent relationships required in engineering workflows, including oil-phase properties and evolved-gas properties. These relationships are routinely used to build simulation inputs and, together with complementary measurements, are also used to constrain and validate thermodynamic model calibration.

Complementary to the DLE-derived volumetric and gas-property outputs, viscosity measurements provide the pressure-dependent oil viscosity curve $\mu_o(p)$ at reservoir temperature across the undersaturated and saturated regions. This curve is a central input to mobility calculations and flow modeling and, like DLE-derived property curves used in simulation tables, must remain smooth, physically plausible, and internally consistent across pressure because small inconsistencies can propagate into simulated performance and forecast behavior.

At the same time, DLE and viscosity tests are experimentally demanding because they require repeated stabilization, accurate metering, and strict gas handling across many pressure stages, followed by careful post-processing of measured volumes and compositions[13]. In addition, these tests are inherently destructive: the sample composition is altered after each pressure step due to progressive gas removal, so the original fluid cannot be reconstructed and the test cannot be repeated on the same specimen. As a result, repeat measurements (or re-runs after quality-control issues) require fresh aliquots of the original fluid, increasing the volume of expensive, limited-availability samples that must be consumed. Viscosity testing can also be operationally demanding, requiring specialized high-pressure viscometry equipment over multiple pressure stages, which further increases laboratory time and cost. This practical burden can limit how frequently DLE and viscosity measurements can be acquired or updated, even though fluid composition and depletion conditions may vary across wells and evolve over time.

These constraints motivate surrogate methodologies capable of reproducing DLE and viscosity test behavior with engineering fidelity when additional laboratory work is impractical. However, DLE prediction is inherently more structured than predicting isolated scalar properties. The outputs are coupled, pressure-dependent curves that must remain physically plausible and internally consistent. They should reflect a depletion path in which the liquid composition changes at each pressure step as gas is progressively removed. Pointwise prediction at a few pressures is therefore insufficient: nonphysical trends or small inconsistencies can propagate directly into the simulation inputs and subsequently into forecast behavior. A similar requirement applies to viscosity, where the objective is to infer a complete pressure-dependent curve $\mu_o(p)$ that is directly usable in engineering workflows.

In Part I of this series[15], we introduced a domain-driven, physics-backed, proximity-informed Local Interpolation Model (LIM) for Constant Composition Expansion (CCE) tests, designed to leverage the most relevant evidence contained in a curated PVT archive, namely thermodynamically similar fluids. In this Part II, we focus on Differential Liberation Expansion and viscosity. In particular, we extend the same locality principle to DLE and viscosity data, while explicitly

 doi:10.20944/preprints202603.0729.v1

addressing the curve-structured and path-dependent nature of depletion under progressive gas removal. The proposed approach embeds a target fluid in a compositional-thermodynamic descriptor space, identifies a small neighborhood of similar archive fluids, and infers the target DLE tests response through local interpolation and reconstruction within this neighborhood. In parallel, the same neighborhood-based workflow is applied to reconstruct the pressure-dependent oil viscosity curve. To respect the way DLE tests data are generated and used, the methodology is formulated around anchor inference combined with shape-preserving curve reconstruction, so that predicted trends remain consistent with characteristic DLE tests behavior and directly usable in engineering workflows. For viscosity, the same anchor-and-reconstruction logic is used to obtain a smooth, physically plausible $\mu_o(p)$ curve. An additional practical feature is that model applicability is naturally tied to neighborhood strength: predictions are expected to be most reliable when close analogs exist and to degrade gradually as similarity support weakens, providing an interpretable indication of confidence that aligns with engineering practice.

The remainder of the paper is organized as follows. Section 2 summarizes the DLE experiment and the viscosity test and defines the corresponding target outputs used in this work, namely DLE test-derived properties and viscosity pressure curves. Section 3 presents the proposed proximity-informed methodology, including descriptor construction, neighbor selection, anchor inference, and curve reconstruction for both oil- and gas-phase properties as well as oil viscosity. Section 4 evaluates performance on the DLE and viscosity test database using endpoint-level and curve-level diagnostics, including tier-resolved behavior as neighborhood support decreases. Section 5 discusses practical implications and limitations for PVT practice and reservoir simulation workflows, and Section 6 provides conclusions arising from the work.

2. Laboratory Experiments

In this work, the prediction targets are pressure-dependent PVT outputs. These include the oil- and gas-phase property curves derived from the DLE experiment, as well as the oil-viscosity curve obtained from viscosity measurements at controlled temperature (typically reservoir temperature). This section summarizes the laboratory tests that conventionally generate these targets and clarifies the physical meaning and depletion-path context associated with each reported curve. The objective is to document what is measured in the laboratory and to define the target quantities used in the remainder of the study.

2.1. Differential Liberation Expansion Tests

The DLE experiment, also referred to as the Differential Vaporization (DV) test, is a standard volumetric PVT laboratory procedure. It is designed to represent depletion below the saturation (bubble-point) pressure at reservoir temperature under progressive gas removal. As pressure drops below saturation, gas previously dissolved in the oil is liberated; once the liberated gas becomes mobile, it separates from the oil phase. The DLE experiment reproduces this mechanism in a controlled setting by repeatedly allowing gas to evolve to equilibrium and then removing it from contact with the remaining liquid at successive pressure stages.

Figure 1 provides a schematic overview of the DLE experimental workflow. The experiment begins by charging a visual high-pressure PVT cell with the reservoir fluid sample and conditioning it to a single phase at reservoir temperature. Pressure is then reduced (typically by piston movement) until the saturation pressure is reached and the saturated-oil volume $V_{o,sat}$ is recorded. Following this, the pressure inside the cell is further reduced to a discrete pressure below the sample's saturation pressure at reservoir temperature and the two-phase system (evolved gas and residual reservoir fluid) is allowed to equilibrate. Prior to withdrawing gas, the equilibrium phase volumes in the cell at the current test pressure and reservoir temperature are measured. The liberated gas is then withdrawn from the cell through a needle valve (used to regulate the withdrawal rate) and directed to a metering device (e.g., a gasometer) and/or captured in a pycnometer for density determination. As the evolved gas is removed at each successively lower pressure stage, the overall composition of

 doi:10.20944/preprints202603.0729.v1

the oil sample is progressively altered, and the volume (ΔV_g), moles (ΔN_g) and specific gravity (S_g) of the expelled gas are measured at standard conditions. Note that the composition of the liberated gas is measured using a gas chromatograph (GC). Additionally, the volume of the remaining oil in the cell ($V_{o,i}$) is recorded. This sequence is repeated stepwise at reservoir temperature until the pressure reaches atmospheric conditions. At the end of the test, a residual-oil sample is withdrawn from the PVT cell at ambient conditions, and its density $\rho_{o,res}$ at 60 °F is measured using a separate density-measurement instrument with sub-ambient temperature control. The corresponding residual-oil volume $V_{o,res}$ is then determined. Finally, the residual oil composition is analyzed using a GC, after first vaporizing the sample prior to injection into the gas chromatograph.

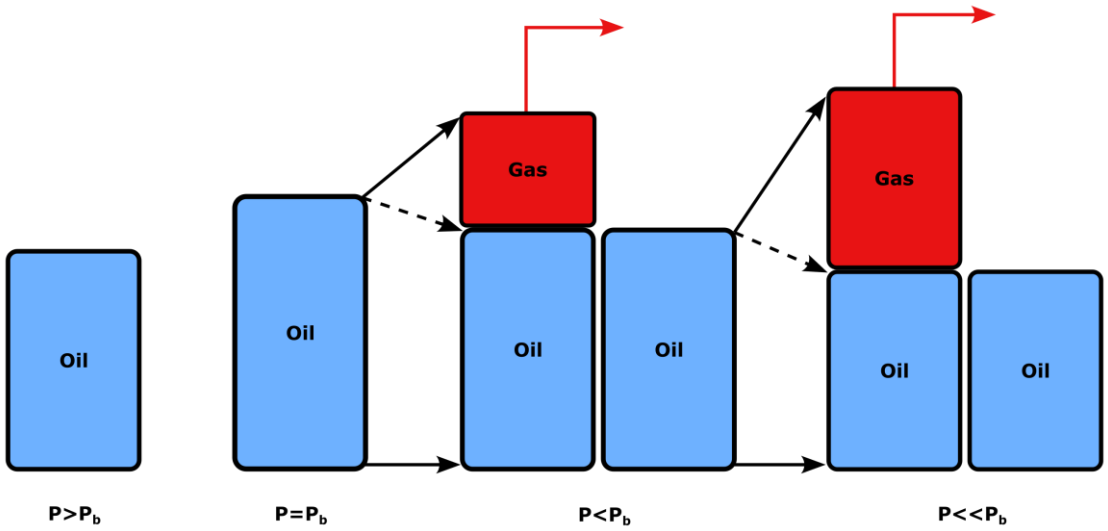


Figure 1. Differential liberation (DLE) experiment.

Beyond the raw measurements, several pressure-dependent properties that are directly used in reservoir engineering workflows are calculated from raw volumetric and composition data and included in DLE reports. These include the following:

- Differential oil formation volume factor (B_o)

B_o is computed by referencing the recorded oil volume ($V_{o,i}$) of the oil at the i^{th} pressure step to the residual oil volume ($V_{o,res}$):

$$B_{o,i} = \frac{V_{o,i}}{V_{o,res}} \tag{2.1}$$

- Differential solution gas–oil ratio (R_s)

R_s is defined as the cumulative standard gas volume that will be liberated from the current pressure step down to the final atmospheric step, per unit residual oil volume ($V_{o,res}$):

$$R_{s,i} = \frac{\sum_{j=i}^{N} \Delta V_j}{V_{o,res}} \tag{2.2}$$

where $k = N$ denotes that the final stage of the DLE test atmospheric pressure and reservoir temperature. The sum in the nominator represents the amount of gas that remains dissolved in the remaining fluid, at current pressure step.

- Oil density (ρ_o)

Oil density is defined as:

$$\rho_{o,i} = \frac{m_{o,i}}{V_{o,i}} = \frac{m_{o,res} + \sum_{j=i}^{N} \Delta N_{g,j} MW_{g,j}}{V_{o,i}} \tag{2.3}$$

 doi:10.20944/preprints202603.0729.v1

where $m_{o,res}$ is determined from the measured density and volume of the residual oil. The moles of the gas, ΔN_g, removed at each step, along with the gas molecular weight, MW_g, typically in the form of its specific gravity, $\gamma_{g,i}$, and the oil volume at each stage, are all measured and known.

- Compressibility factor (Z) and gas formation volume factor (B_g)

The compressibility factor of the produced gas at each step accounts for its non-ideal behavior and is determined from the real-gas relation at the step conditions:

$$Z_i = \frac{\Delta V_{g_i}}{V_{sc}} \frac{P_i}{P_{sc}} \frac{T_{sc}}{T_{sc}} \tag{2.4}$$

where P and T are the pressure and temperature of the hydrocarbon system, ΔV_g is the volume of the liberated gas at P and T and P_{sc}, V_{sc} and T_{sc} correspond to standard conditions.

From the compressibility factor, the gas formation volume factor (B_g) can be determined by definition as follows:

$$B_{g_i} = Z_i \frac{P_{sc}}{P_i} \frac{T}{T_{sc}} \tag{2.5}$$

- Gas specific gravity (S_g)

The gas specific gravity of the produced gas at each step is used as an indicator of gas "heaviness" and for mass accounting, and is used to compute the amount of mass released at each pressure step:

$$S_{g,i} = \frac{\mathrm{MW}_{g,i}}{MW_{air}} \tag{2.6}$$

where MW_{air} is the air molar mass, i.e., 28.97 g/mole.

The liberated gas phase density can be directly computed as a function of the gas molar mass (originating from the S_g curve) and its deviation factor (Z) using the real gas law. Note that some laboratories measure gas density directly at each pressure step. Here, we follow the calculation-based procedure consistent with our dataset.

2.2. Viscosity Tests

Dynamic (absolute) viscosity, denoted by μ_o, accounts for the resistance to flow exerted by a fluid and is typically reported in centipoise (cP). Viscosity is strongly influenced by temperature, as well as by the amount of solution gas retained in the oil.

In PVT laboratory practice, viscosity is measured at a controlled test temperature (typically the reservoir one) and reported as a function of pressure over a prescribed pressure grid that spans the undersaturated region above the bubble point and the saturated region below it. At pressures above p_{sat}, the oil remains single-phase and μ_o often decreases modestly as pressure declines toward p_{sat}. Once pressure falls below p_{sat}, gas evolves from the liquid phase and the amount of gas remaining in solution decreases with continued depletion. Because dissolved light and intermediate components act as a diluent/solvent, their progressive liberation reduces the dilution effect and enriches the remaining liquid in heavier fractions. As a result, the viscosity pressure curve below the initial pressure commonly exhibits a minimum at the bubble point followed by an increasing trend at lower pressures.

Operationally, viscosity pressure data are generated by conditioning the sample at the test temperature and performing measurements at successive pressure stages. At each stage, the system is allowed to equilibrate, and the reported viscosity corresponds to the laboratory definition of oil viscosity at those pressure–temperature conditions. Below the bubble point, additional care is required to ensure that the measured value corresponds to a well-defined oil phase at equilibrium conditions (e.g., adequate equilibration time and avoidance of entrained gas in the measurement). The outcome of the laboratory procedure is therefore a curve-structured quantity, $\mu_o(p)$, that is directly tabulated and transferred into engineering calculations and simulation workflows.

 doi:10.20944/preprints202603.0729.v1

A variety of viscometry principles can be used to obtain viscosity pressure data, depending on viscosity level and the operating pressure–temperature range[16]. Common approaches include capillary-flow viscometers, where viscosity is inferred from the pressure drop (or gravity head) required to drive laminar flow through a capillary, falling- or rolling-body viscometers, where viscosity is deduced from the motion of a body through the fluid under gravity, oscillating-piston (electromagnetic) viscometers, where viscosity is inferred from the piston's dynamic response as viscous forces resist its motion in a narrow annulus, and rotational viscometers or rheometers, where viscosity is obtained from measured torque under controlled shear. High-pressure, temperature-controlled configurations exist for these methods, enabling measurements on live oils over elevated-pressure grids. Regardless of instrument type, calibration or validation against viscosity standards and routine quality-control checks such as temperature stability, equilibration at each condition, repeatability of readings, and avoidance of sample contamination or alteration during handling are required to ensure consistent viscosity reporting across the pressure grid.

3. Methodology

3.1. Methodology Features

The PVT properties appearing in a typical DLE test, as well as the oil viscosity curves, are functions of pressure, typically modelled with appropriate function forms, based on the individual values computed at each pressure step. Once a new (target) fluid appears requiring the prediction of its full DLE-derived PVT properties and viscosity behavior, a prediction system must provide full curves of all oil and gas properties, with the former defined from the initial pressure down to the atmospheric one (in combination to the CCE data) and the latter being defined between the first and the last pressure steps below p_{sat}, together with the corresponding oil viscosity curve over the reported pressure range.

Importantly, the proposed framework is designed for inference-time deployment: the PVT database gets curated and standardized once, and every new prediction is produced on demand using only a small subset of relevant analogs, without retraining a global model for each newly encountered fluid. Apart from few values like the density of the residual oil, all properties such as B_o, R_s and μ_o vary as a function of pressure. As a result, in an AI modeling context, the accurate estimation of pressure-dependent PVT properties in the DLE test and viscosity measurements needs to be formulated as a multi-level prediction problem.

In this work, the task is decomposed into predicting physically meaningful anchor quantities that define scale and boundary behavior and inferring the intermediate curve trajectory in a normalized shape preserving space (i.e. in the unitary pressure-property square), so that the reconstructed curve remains consistent with the expected physical trends of each target property. Firstly, each property is split into its diphasic and monophasic parts (where applicable), so that each part to be predicted constitutes a monotonic curve of pressure for all oil phase properties (i.e. B_o, R_s and ρ_o). Oil viscosity is also treated as a pressure-dependent curve, $\mu_o(p)$, but the full curve is not monotonic over the entire pressure range because viscosity typically reaches a minimum close to the bubble point. To preserve monotonicity, viscosity is treated as two monotonic branches that meet at p_{sat}: a saturated (two-phase) branch for $p_{atm} \leq p \leq p_{sat}$, where μ_o increases as pressure declines below p_{sat}, and an undersaturated (single-phase) branch for $p \geq p_{sat}$, where μ_o declines as pressure declines. Each branch is constrained by physically meaningful anchors, and the intermediate trajectory is reconstructed using a branch-appropriate model.

Subsequently, a small set of physically meaningful anchor values (endpoints) at reference pressures are defined for each target curve. These anchors reflect the way DLE and viscosity data is interpreted and used in practice: they correspond to reference pressures that carry clear physical meaning and strongly constrain the curve, reducing the risk that a purely pointwise model produces nonphysical trends. For the example of R_s, the endpoint values considered are $R_s@p_{atm}$ and $R_s@p_b$. Similarly, for the diphasic region oil density, the endpoints required are $\rho_o@p_{atm}$ and $\rho_o@p_b$. For

 doi:10.20944/preprints202603.0729.v1

oil viscosity, the corresponding anchors are $\mu_o@p_{atm}$ and $\mu_o@p_b$. Unless the point values are fixed or known (such as $R_s@p_{atm} = 0$), appropriate Local Interpolation Models (LIMs) are developed to provide those point values for any new test fluid.

Finally, the normalized intermediate diphasic curve shape, constrained by those anchors, is predicted by an appropriate AI model. Here, physics-backed refers to enforcing the correct boundary behavior through anchors and learning only within the pressure window where the property is defined, while using monotonic, shape-preserving representations so that the reconstructed curve respects the expected depletion behavior. Once the endpoints and the curve shapes have been predicted by the AI models, the full curve can now be accurately reconstructed by combining those separate pieces of inferred information.

The endpoint and curve shape modules which lie at the heart of the prediction system are based on a physics backed logic which combines AI techniques, often considered as data-centric methods, to the workflows which incorporate a PVT Expert's approach that introduces principles and laws from the fluid engineering domain. Unlike "blind" and all-purpose AI methods, it utilizes the exact amount of information living in a PVT database which is needed for the specific target fluid under study only, thus justifying its "local" nature. This locality is intentional: it favors interpolation within thermodynamically coherent families of fluids when close analogs exist, and it avoids forcing a single global mapping to extrapolate across regimes that are poorly represented in the PVT database.

The developed prediction system also provides estimates of the uncertainty accompanying all its predictions. In particular, confidence is tied to neighborhood support: predictions are considered most reliable when the target is surrounded by several close analogs, and progressively less reliable as the target moves away from the archive manifold. For that task, when a target fluid arrives, the number of neighbors and their distances from the new fluid are evaluated in contrast to the inter-distances between all dBase fluids. A big number of close neighboring fluids ensures safe predictions within a minimum uncertainty and highest accuracy. On the other hand, a target fluid which is entirely different from any other existing in the dBase, will still receive a prediction by the LIM model but it will be accompanied by high uncertainty and lack of confidence on the accuracy of the predicted PVT values. The terms Tier 1 to Tier 6 are used to order a test fluid according to its similarity to the dBase ones, as will be discussed later in detail. Operationally, tiering is derived from the target-to-neighbor distances evaluated against the characteristic inter-fluid distance scale of the database, so tiers encode both similarity and how well the database supports the target region.

3.2. Methodology Details

At a more detailed level, the required PVT repository is expected to contain full compositional, DLE, and viscosity analyses for each of the contained fluids. The compositional data (composition plus some fluid quality describing values like API, GOR, etc, obtained through the course of a compositional analysis) constitute the descriptor space (i.e. the LIM input). Once the input of the test fluid appears, the PVT properties anchor values are predicted using LIM, which identifies similar fluids (i.e. neighbors) in the dBase descriptor space and performs smart local interpolation of their DLE target endpoints, and the corresponding viscosity anchors.

To demonstrate neighbor support, we compute the similarity between every fluid in our proprietary archive and a target (unknown) fluid in the full descriptor space. Rather than using a simple Euclidean distance metric, this similarity is evaluated using the same multi-metric, weighted distance formulation introduced in Part I [15], so that different descriptor blocks contribute in a controlled way. Each archive fluid is then assigned a percentile rank based on its distance to the target one (relative to all target-to-archive distances), which is mapped to qualitative similarity clusters. For example, fluids assigned to Cluster A exhibit higher similarity to the test fluid than those in Clusters B to F).

Once Clusters have been assigned to each archive fluid, the algorithm proceeds by providing the prediction of the PVT property under study to the test fluid, followed by the assignment of the appropriate Tier number. For the prediction, LIM can exploit many fluids living in a broad

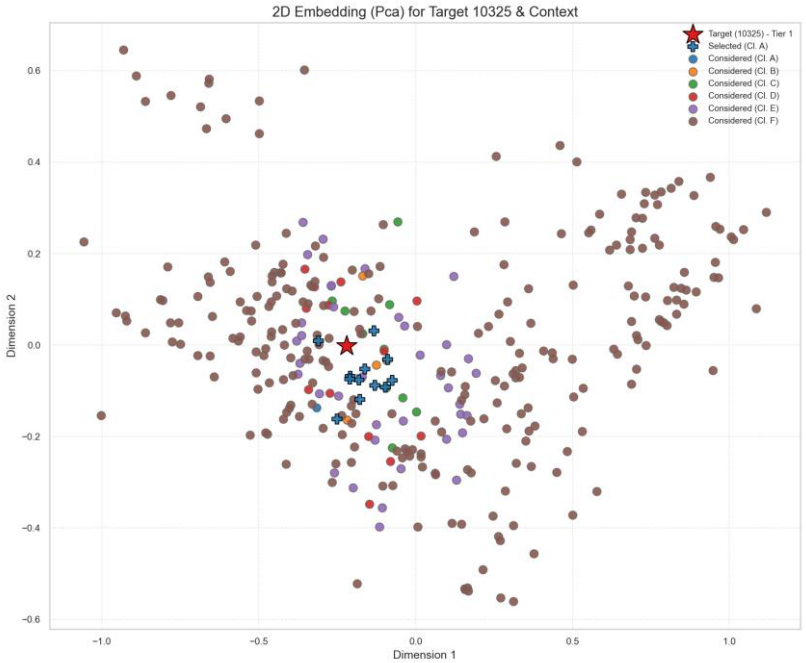
 doi:10.20944/preprints202603.0729.v1

neighborhood. The neighbors PVT values are projected using Taylor expansion (see Eq. 3.1 in Part I [15]) and the projections are then weighted according to the neighbors' distance to the test fluid. In contrast, the Tier number of the test fluid is defined only from the few nearest neighbors, which quantify the strength of the immediate local support due to their significant contribution as op-posed to that of distant archive fluids: If enough neighbors in Cluster A are available, the test fluid is attributed to Tier 1, whereas Tier 2 appears when not enough Cluster A neighbors are available, but some Cluster B fluids are required to complete the picture. Conversely, when the nearest neighbors live up to Cluster F, the test fluid is assigned to Tier 6, since progressively weaker clusters need to considered.

To demonstrate the Clusters and Tiers definitions, we visualize neighborhood structure for three fluids of increasing Tier (1, 2 and 3) in Figures 2–7. Because the descriptor space is high dimensional, via 2D embeddings of the input descriptors using Principal Component Analysis (PCA) and Uniform Manifold Approximation and Projection (UMAP). These plots are illustrative only: all neighbor ranking, clustering, and tiering are performed in the original descriptor space. The plots highlight the target fluid, its top tiering neighbors, and the surrounding archive fluids colored according to their similarity cluster. For each target, both the PCA and UMAP embeddings are shown as separate figures for clarity.

Figure 2 demonstrates the location of the test fluid and all archive fluids in the PCA while keeping only the two dominant directions (i.e., components). The test fluid is shown by the star marker whereas the dBase fluids are indicated by circles, the color of each is defined according to the Cluster it belongs to for the specific test fluid. The closest points (in the full descriptor space) are shown in "plus" markers. It can be readily seen that enough Cluster A neighbors (shown in blue plus marker) are available to the test fluid, thus rendering it as a Tier 1 one. Interestingly, shrinking the high dimensional descriptor space using only 2 linear principal components still captures the structure of the fluids inter-distances map. A similar situation holds when the UMAP dimensionality reduction technique replaces PCA (Figure 3).

When it comes to Tier 2, Figures 4 and 5 demonstrate a fluid which has only few Cluster A neighbors (shown in blue plus markers), thus introducing the need to ac-count for few Cluster B ones to arrive to the minimum number of neighbors needed to assign a Tier value. In this case, the test fluid is assigned Tier 2. Similarly, Figures 6 and 7 demonstrate the location of a Tier 3 fluid in the reduced dimensionality space, where "plus" marks of three consecutive clusters (A, B and C) are needed to get the Tier number.



 doi:10.20944/preprints202603.0729.v1

Figure 2. Two-dimensional PCA embedding of the full input descriptor space for Target 10325. The star marks the target fluid. The crosses mark the top five nearest neighbors in the original descriptor space used for tier assignment and as core local support for LIM. All five crosses belong to similarity Cluster A, which means the target has a dense set of highest-quality analogs and is therefore classified as Tier 1. The remaining archive fluids are shown as circles and colored by similarity cluster from A as most similar to F as least similar. The PCA map is used only for visualization and all distances and tiers are computed in the original descriptor space.

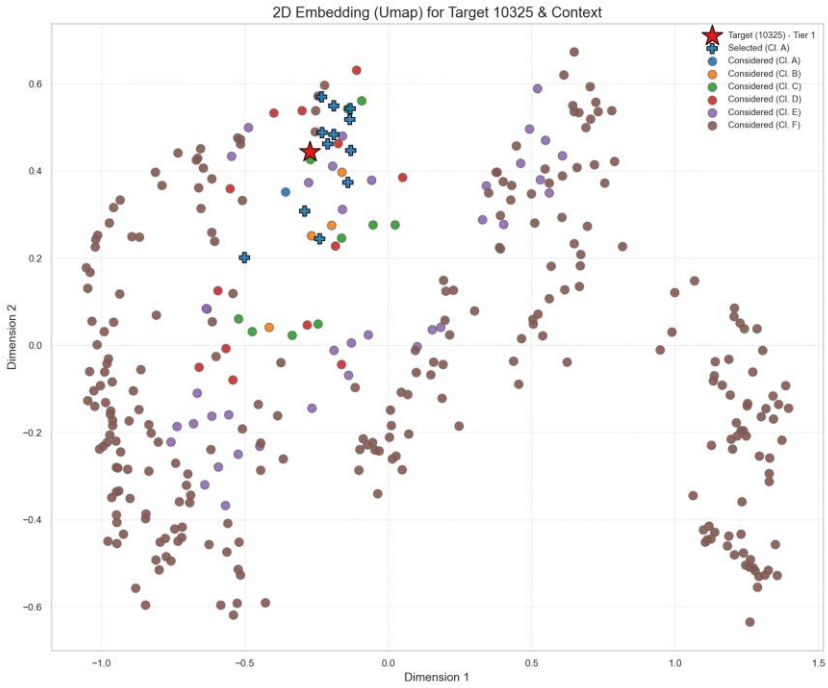
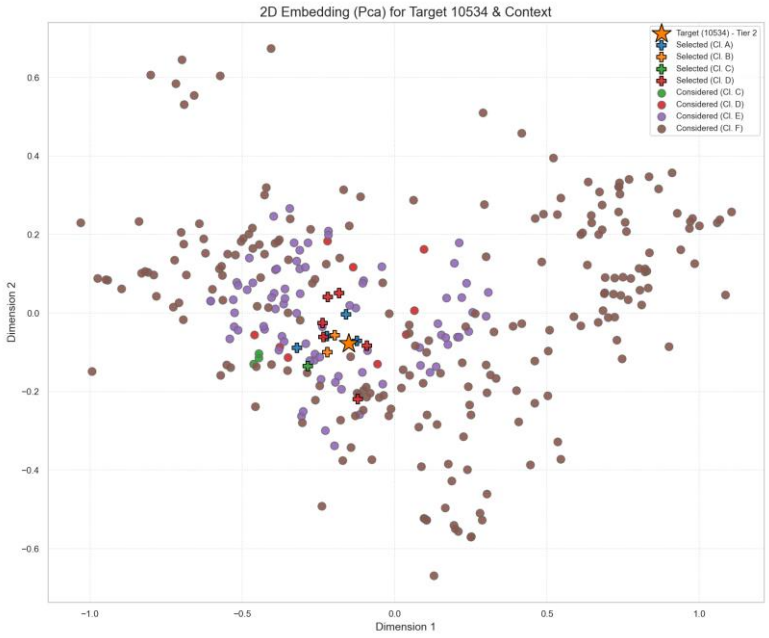


Figure 3. Two-dimensional UMAP embedding of the descriptor space for Target 10325. The star marks the target and the crosses mark the top five nearest neighbors used to define the tier. The selected crosses are entirely from Cluster A, confirming strong immediate local support and therefore a Tier 1 classification. Circles show the broader archive context colored by similarity cluster A to F. The UMAP projection is illustrative only and all neighbor ranking and tiering are performed in the original descriptor space.



 doi:10.20944/preprints202603.0729.v1

Figure 4. Two-dimensional PCA embedding of the descriptor space for Target 10534. The star marks the target fluid. The crosses mark the top five nearest neighbors in the original descriptor space that define the tier. Here the top five cannot be filled using Cluster A alone and Cluster B neighbors are required to complete the five, which is the definition of Tier 2 in our framework. Circles show the remaining archive fluids colored by similarity cluster A to F. PCA is shown only to visualize neighborhood structure and does not affect the distance calculations.

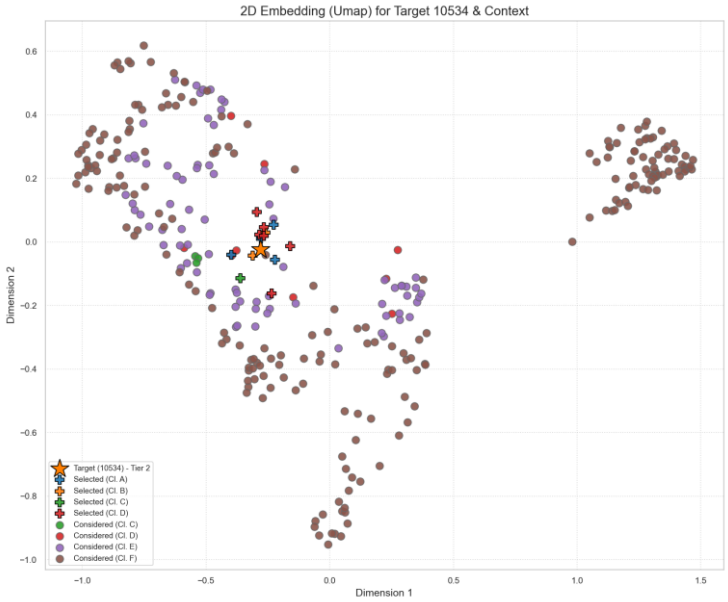


Figure 5. Two-dimensional UMAP embedding for Target 10534. The target is marked by the star and the crosses indicate the top five nearest neighbors used for tiering. The presence of Cluster B among these five neighbors indicates that the immediate high-quality support is not sufficient in Cluster A alone, which justifies the Tier 2 label. The background archive fluids are shown as circles and colored by similarity cluster A to F. The embedding is for visualization only and all tiers are computed in the original descriptor space.

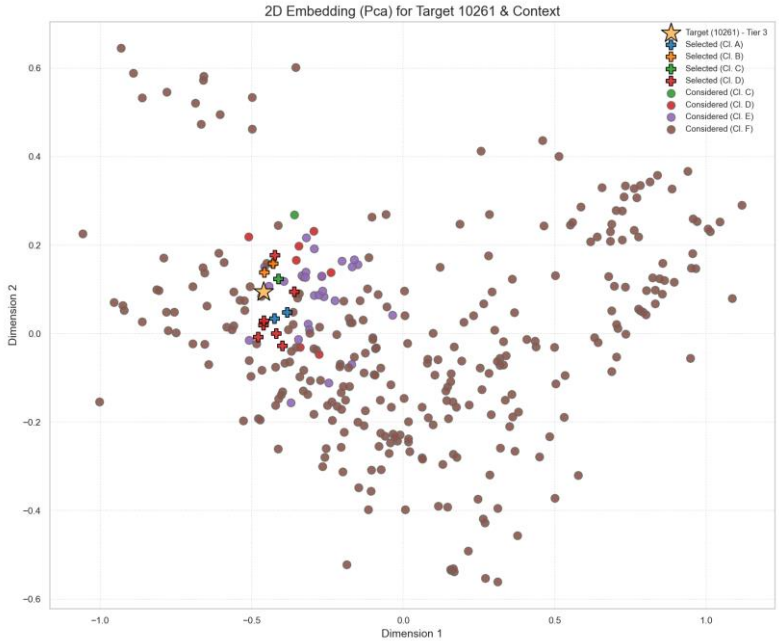


Figure 6. Two-dimensional PCA embedding of the descriptor space for Target 10261. The star denotes the target fluid. The crosses denote the top five nearest neighbors used to define the tier. In this case, completing the top

 doi:10.20944/preprints202603.0729.v1

five requires drawing beyond clusters A and B and into weaker similarity clusters, which indicates reduced immediate local support and therefore a Tier 3 classification. Circles show the remaining archive fluids colored by similarity cluster A to F. The PCA map is shown for visualization only and does not change the underlying distance ranking.

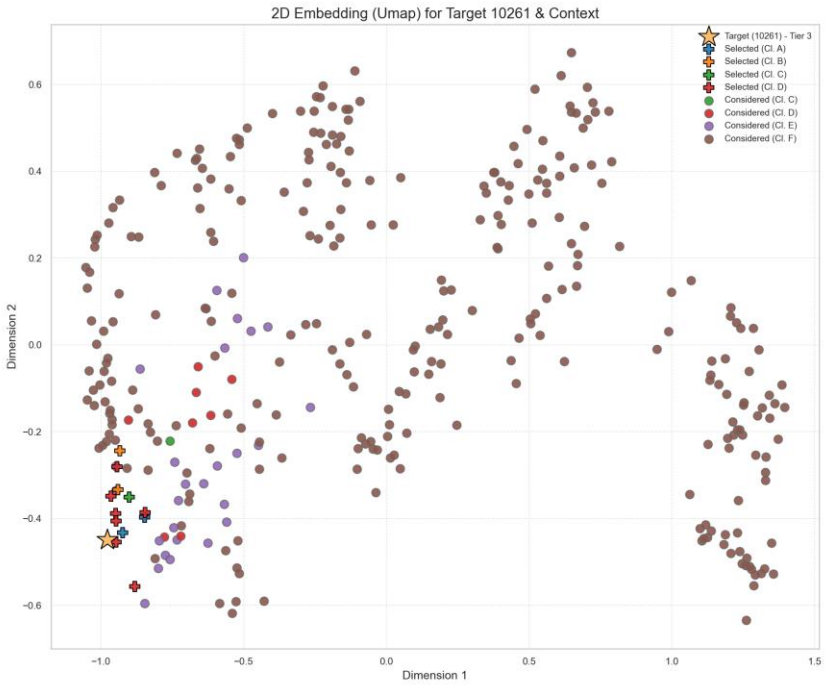


Figure 7. Two-dimensional UMAP embedding for Target 10261. The target is shown as a star and the crosses mark the top five nearest neighbors used for tier assignment. The need to include neighbors from clusters weaker than A and B in the top five is consistent with a sparser or less supported neighborhood and justifies the Tier 3 label. The archive context is shown as circles colored by similarity cluster A to F. The 2D embedding is illustrative only and all tiering is computed in the original descriptor space.

To make the LIM anchor-prediction stage explicit, we briefly summarize here the key steps introduced in our previous work[15] and illustrated in Figure 8. When a test fluid is provided, its input vector (composition, temperature, and auxiliary fluid-quality descriptors) is embedded in the same descriptor space as the dBase fluids, after which a trust region is defined using the neighborhood criteria and distance metric described in[15] . The resulting neighbor set is used to transfer endpoint information to the test fluid through a local, physics-backed correction: neighbor anchor values are first adjusted toward the test-fluid location via the proposed Taylor-step extrapolation and then aggregated using distance-based weights to yield the final LIM estimate. This weighting accounts for both proximity (more similar fluids contribute more) and experimental noise dispersion across neighbors (through averaging), producing robust anchor predictions that are consistent with domain-expert practice when working with a compositional PVT library. These LIM-predicted anchors are subsequently used as the fixed endpoints and denormalization constraints for reconstructing all target curves in their physical pressure domain, including both DLE-derived properties and oil viscosity. For viscosity in particular, reconstruction is controlled by two reference-pressure anchors, namely the bubble-point oil viscosity $\mu_{o,b} \equiv \mu_o(p_b)$ and the atmospheric-pressure reference $\mu_{o,atm} \equiv \mu_o(p_{atm})$.

 doi:10.20944/preprints202603.0729.v1

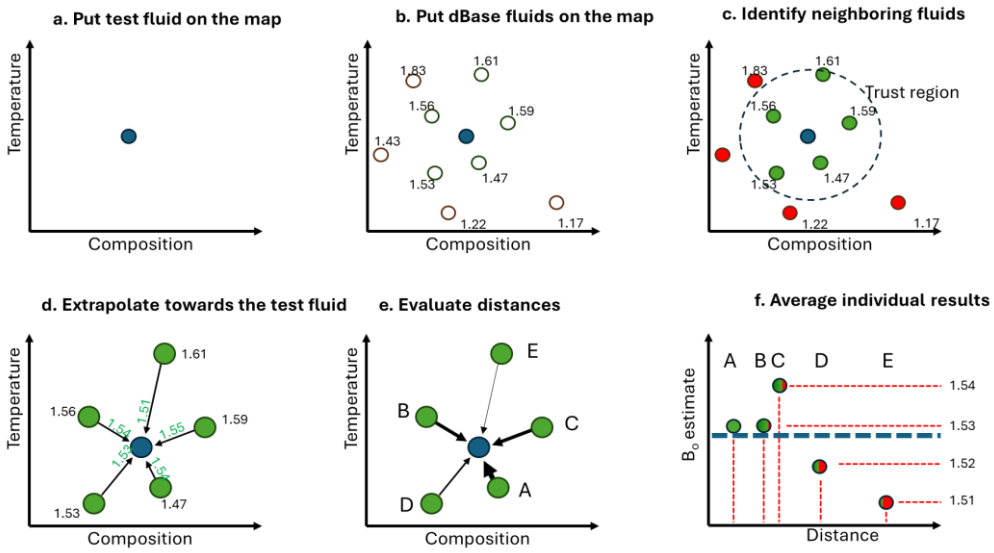


Figure 8. PVT predictions workflow.

Unlike the predicted anchors, the intermediate property profile (e.g. R_s pressure curve values between p_{atm} and p_{sat}) is treated as a shape-learning task: each dBase fluid property curve is sampled on a fixed pressure grid within its pressure window and mapped into a normalized unit-square representation (Figure 9). A physics backed regressor predicts the normalized trajectory of the target fluid by obtaining shape information from the most similar curves found in the PVT library. Finally, the predicted normalized curve is denormalized (i.e. transformed back to the original pressure space) using the LIM-predicted anchors, yielding the full curve in physical units.

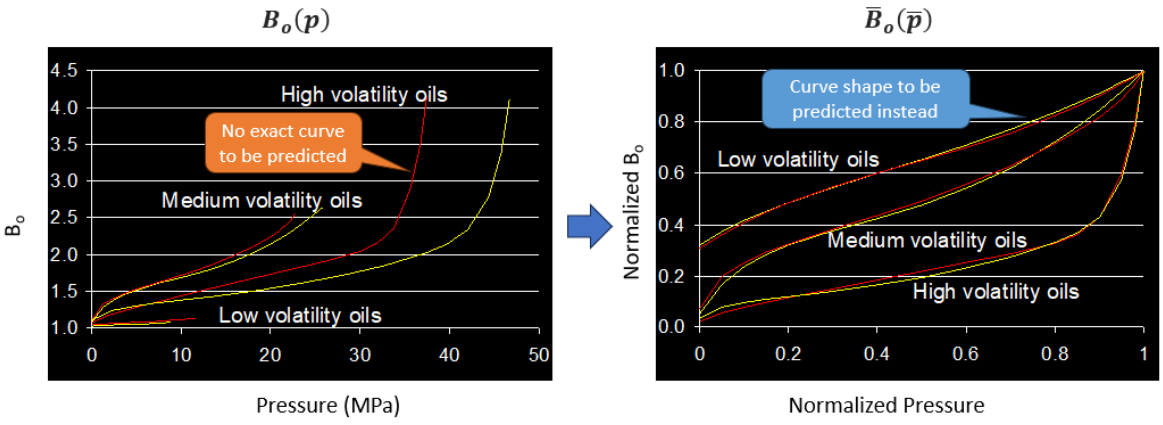


Figure 9. Unit square normalization of a Differential Liberation B_o curve for shape learning.

Note that the gas phase properties curves over pressure need some extra manipulation since they are not defined at p_{sat}. For properties such as the Z factor or the S_g, the curves are trimmed a specific pressure relative values and the original properties are transformed appropriately to ensure monotonicity (e.g. Z/p is predicted, i.e. the molar volume, rather than Z itself). Moreover, due to their abrupt increase when arriving at p_{atm}, as is the case with S_g or Z/p , the gas properties curves are further trimmed below some specific near-atmospheric relative pressure value. The values at p_{atm}, can then be inferred by domain's expertise (e.g., $Z/p \,@\, p_{atm} \approx 1/p_{atm}$) or using a separate LIM model (e.g., predict $S_g \,@\, p_{atm}$).

The same unit-square mapping and shape reconstruction is applied to oil viscosity in the saturated region, using $\mu_{o,atm}$ and $\mu_{o,sat}$ as the normalizing anchor points. Since viscosity exhibits a

 doi:10.20944/preprints202603.0729.v1

minimum at the bubble point p_{sat}, the full $\mu_o(p)$ curve is treated as two monotonic curves that meet at the bubble point, so that the unit-square formulation remains well posed.

Viscosity workflows commonly rely on fixed reference viscosities to tune correlation-based models before applying them across pressure, for example when implementing and tuning LBC-type viscosity correlations [17] against laboratory measurements. In our setting, reference anchors at p_{sat} and p_{atm} are considered well defined, and the corresponding viscosity constraints are introduced in a workflow-consistent manner: $\mu_{atm} \equiv \mu_o(p_{atm})$ and $\mu_{sat} \equiv \mu_o(p_{sat})$ are treated as known. Given that it might be easier to provide a PVT measured value at some pressure p_s above p_{sat} rather than at p_{sat} itself, a variation of the main method is developed by considering a reported single phase fluid viscosity value $\mu_s = \mu_o(p_s)$. Then, μ_{sat} is obtained by back-calculating the bubble-point anchor from the monophasic viscosity measurement by inversely solving the Beal equation [18].

For the undersaturated region viscosity is reconstructed using the Beal correlation [18] anchored at the LIM-predicted μ_{sat} :

$$\mu_o(p) = \mu_{sat} + 10^{-5}(p - p_{sat})(2.4\mu_{sat}^{1.6} + 3.8\mu_{sat}^{0.56}) \quad (3.1)$$

For the saturated region ($p \le p_{sat}$), the unit-square reconstruction is applied over [p_{sat}, p_{atm}]using μ_{sat} and μ_{atm} as normalizing anchors. The normalized trajectory is inferred from the proximity-selected neighbor curves and then mapped back to physical units. In practice, a low-pressure viscosity value is often available from routine PVT reporting as the dead (stock-tank) oil viscosity, i.e., a gas-free viscosity near atmospheric pressure. When needed, it can be adjusted to reservoir temperature using standard dead-oil viscosity correlations. Beggs–Robinson [19] is commonly used in this context: it provides a dead-oil viscosity estimate and, together with R_s, a relation to obtain the bubble-point (saturated) viscosity μ_{sat} from the dead-oil reference. Because a direct flash does not follow the same thermodynamic path as depletion below p_{sat}, the resulting liquid composition (and thus viscosity) can differ. Therefore, when μ_{atm} is taken from flashed/stock-tank oil it should be regarded as a practical low-pressure proxy anchor rather than an exact endpoint of the sub-p_{sat} viscosity path.

The endpoints-curves decomposition of the DLE and viscosity prediction problem renders the validation of the approach explicit: endpoint parity plots quantify anchors prediction accuracy, while curve-level errors quantify the fidelity of the whole reconstructed curve between the anchor points. The anchor definitions and pressure windows used for each property are summarized in Table 1, and the endpoint-constrained reconstruction concept is illustrated schematically in Figure 10.

Table 1. Summary of property-specific endpoints.

Property	Pressure range	Endpoints	Curve shape reconstruction
B_o (rb/stb)	$p_{atm} \le p \le p_{sat}$	$B_o@p_{sat}$ (LIM) $B_o@p_{atm}$ (LIM)	Monotone increasing; normalized to unit square; kNN shape between endpoints
R_s (scf/stb)	$p_{atm} \le p \le p_{sat}$	$R_s@p_{sat}$ (LIM) $R_s@p_{atm}$ = 0 (domain expertise)	Value fixed at p_{atm}; kNN reconstructs normalized rise to p_{sat}
ρ_o (g/cc)	$p_{atm} \le p \le p_{sat}$	$\rho_o@p_{sat}$ (CCE) $\rho_o@p_{atm}$ - (LIM)	Monotone decreasing; normalized to unit square
Z factor	$0.1p_{sat} \le p \le 0.95p_{sat}$	$Z@0.1p_{sat}$ (LIM) $Z@0.95p_{sat}$ (LIM) $Z@p_{atm}$= 1 (domain expertise)	Model Z in gas-present range; extend smoothly to atmospheric
S_g	$0.1p_{sat} \le p \le 0.95p_{sat}$	$S_g@0.1p_{sat}$ (LIM) $S_g@0.95p_{sat}$ (LIM) $S_g@p_{atm}$ - (LIM)	Model S_g in gas-present range; extend smoothly to atmospheric

 doi:10.20944/preprints202603.0729.v1

B_g (rcf/scf)	$0.1p_{sat} \leq p \leq 0.95p_{sat}$	Derived (no direct LIM)	Computed from $B_g = ZRT/p$ on same pressure grid
μ_o (cP)	$p_{atm} \leq p \leq p_{sat}$ $p \geq p_{sat}$	$\mu_o@p_{sat}$ (inverse-Beal) $\mu_o@p_{atm}$ (well-defined)	Neighbors interpolation on unit-square for $p_{atm} \leq p \leq p_{sat}$ and Beal for $p \geq p_{sat}$

To standardize learning across the fluids in the PVT dBase available, which exhibit different saturation pressures and property magnitudes, each dBase property curve $y(p)$ is sampled on a consistent grid. Let p_{sat} denote saturation pressure at the test temperature and p_{atm} the atmospheric-pressure final step. For a selected pressure interval $[p_{min}, p_{max}] = [p_{atm}, p_b]$, N uniformly spaced pressures p_i are sampled and normalized as

$$\bar{p_i} = \frac{p_i - p_{min}}{p_{max} - p_{min}} \tag{1.1}$$

so that, $\bar{p_1} = 0$ and $\bar{p_N} = 1$.

The property axis is normalized to [0,1] using bounds (y_{min}, y_{max}) determined from the anchor points, e.g. $y_{min} = B_o@p_{atm}$ and $y_{max} = B_o@p_{sat}$. For properties that increase with pressure over the modeled interval,

$$\bar{y}_i = \frac{y_i - y_{min}}{y_{max} - y_{min}} \tag{1.2}$$

When the limiting values are set to the min/max values (as in the equations above), the normalized points are guaranteed to lie within a unit square, i.e. $\bar{x}, \bar{y} \in [0,1]$, provided the property is monotonic in pressure. After $\hat{\bar{y}}$ at the sampled $\bar{x}$ points is predicted, the curve is reconstructed in physical units using the inverse mapping based on the same bounds. The same normalization–reconstruction logic is used for all modeled targets in this work, including DLE-derived properties and the monotonic branches of the viscosity curve.

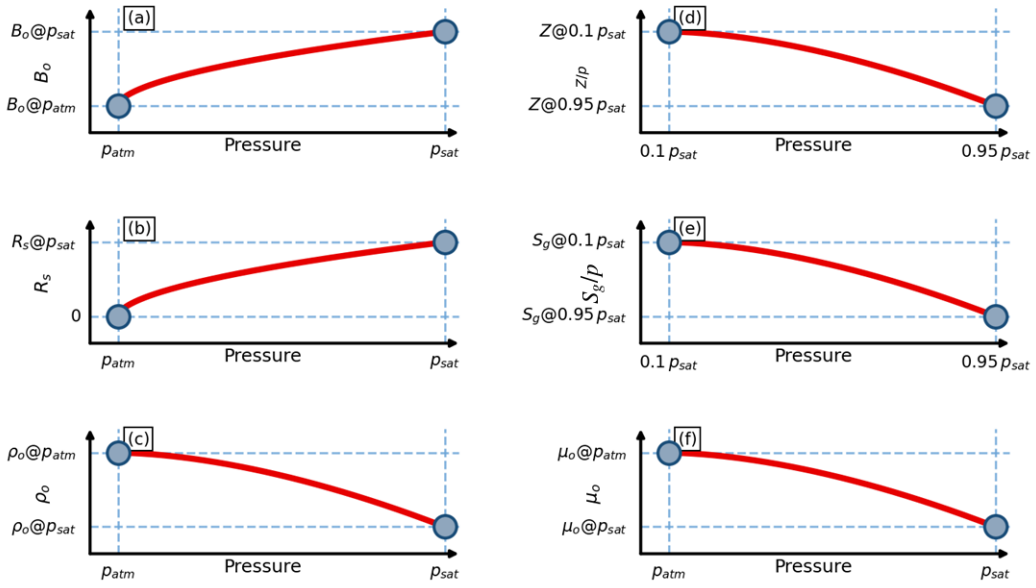


Figure 10. Endpoint anchors and diphasic curve-shape reconstruction windows used in the DLE workflow: (a) B_o, (b) R_s, (c) ρ_o, (d) Z, (e) S_g, and (f) μ_o.

4. Results

Unlike classic supervised regression modeling where a part of the dBase needs to be "sacrificed" and kept aside to evaluate later the trained model's performance, the Fluidsdata approach can

 doi:10.20944/preprints202603.0729.v1

provide results for all fluids, thanks to the local interpolation feature. Indeed, to evaluate the model performance for the DLE and viscosity test sets, each single fluid is inferred from its neighborhood in a leave-one-out manner, thus handling it as a completely unknown target. Moreover, the tiering system provides a transparent proxy for neighborhood strength, with Tier 1 being the strongest and Tier 6 the weakest. Accuracy is reported at two complementary levels. Endpoint agreement compares anchor predictions through parity plots and relative-error histograms, while curve-level agreement evaluates per-fluid curve MAPE (%) over the modeled pressure path and summarizes its distribution together with the 95th-percentile error level, which captures "worst typical" behavior without being dominated by extreme outliers. For viscosity, curve-level errors are reported separately for the saturated branch reconstructed over $p_{atm} \leq p \leq p_{sat}$. Representative Tier 1–3 examples are used to assess qualitative fidelity, while tier-resolved summary statistics across Tiers 1–6 provide broader context on how performance degrades as neighborhood support weakens.

Before examining prediction accuracy, it is useful to demonstrate that the neighbor selection procedure only retrieves fluids that are locally coherent in the space of the measured test responses. Figure 11 shows the normalized $R_s(p)$ curve of a representative target fluid against its five closest neighbors (green) and against the remainder of the archive (light gray). The test fluid is shown in red, while the neighbor set forms a tight bundle around the test curve for both properties, in clear contrast to the much broader envelope spanned by the full database. This qualitative agreement indicates that the descriptor-space distance metric and tiering system, driven by compositional/thermodynamic information such as composition, temperature, and saturation/flash descriptors, retrieves fluids that are also neighbors. In practical terms, the method mimics the first step a domain expert would follow when estimating or quality controlling PVT test behavior from an archive: identify a small set of thermodynamically similar fluids and then infer endpoints and curve shape within that locally coherent family.

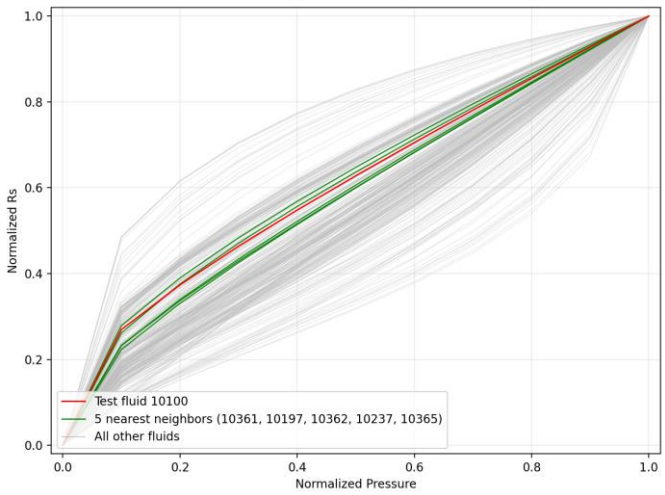


Figure 11. Qualitative validation of neighborhood selection for DLE test curve families: normalized R_s for a test fluid (red), its five nearest neighbors (green), and all other fluids (gray).

The focus of this work is on the methodology for predicting reservoir-fluid PVT properties using the LIM, which is applicable to any database that contains appropriately validated measurements. The reported matching statistics are specific to the database used here, detailed discussion of the database/validation is out of scope.

4.1. Oil Volume Factor

Starting with B_o (rb/stb), the Tier 1–3 example curves in Figure 12 demonstrate close tracking of the laboratory $B_o(p)$ trend across the depletion path, with inferred curves remaining smooth and consistent with expected DLE behavior. Figure 6 illustrates three representative cases, one from each

 doi:10.20944/preprints202603.0729.v1

tier (Tier 1, Tier 2, and Tier 3), where the predicted curve follows the measured depletion trajectory with only modest deviations. Across these examples, the increase in mismatch from Tier 1 to Tier 3 manifests primarily as gradual bias (mild offsets) rather than shape distortion or spurious oscillations.

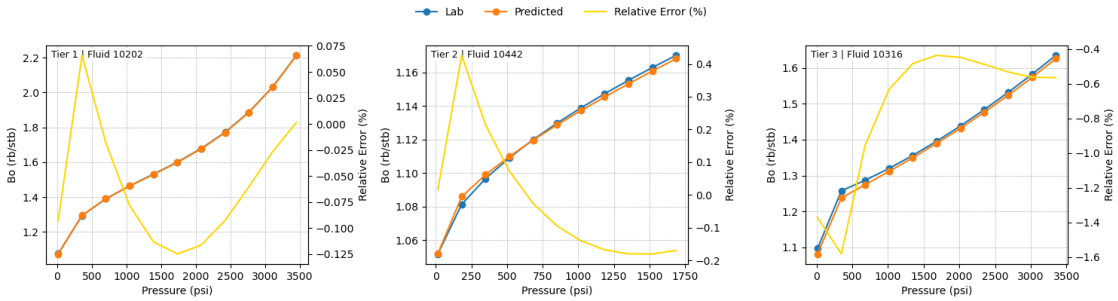


Figure 12. Example inferred B_o curves.

Endpoint prediction diagnostics in Figure 13 show that B_o predictions remain well anchored at both ends of the pressure range, with relative-error histograms centered close to zero at p_{sat} and p_{atm}. The p_{sat} endpoint exhibits a wider spread with heavier tails and a few rare outliers, whereas p_{atm} is noticeably tighter, with most cases clustered very near zero (typically within ±1%) and only occasional excursions (up to ±2%). Overall, the endpoints indicate minimal bias, with more consistent accuracy at atmospheric pressure. The latter is attributed to the limited range that $B_o@p_{atm}$ occupies compared to that of $B_o@p_{sat}$.

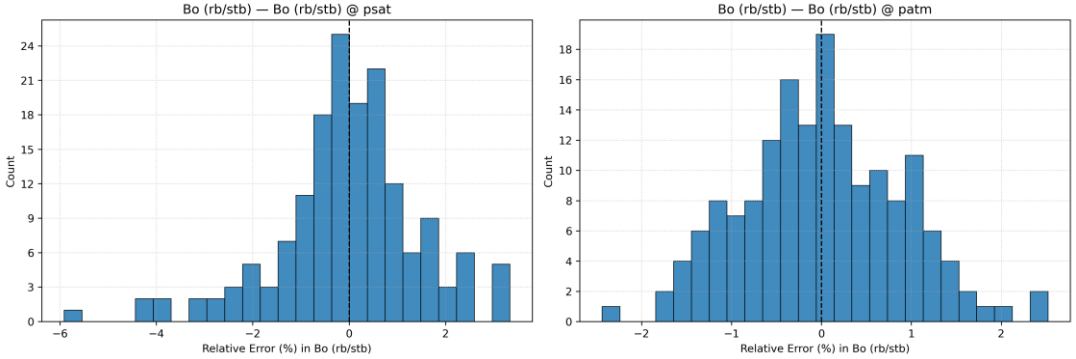


Figure 13. B_o endpoint validation at p_{sat} and p_{atm} (Tier 1–3): relative-error histograms.

Figure 14 summarizes curve-level B_o accuracy through the distribution of curve MAPE and its ECDF, where curve MAPE is defined as the average absolute relative error along the full pressure range. Most B_o curves exhibit low errors (predominantly in the 0.5–1.5% range), with the ECDF indicating a median (50th-percentile) around 0.9% and a 95th-percentile near 2.7%, meaning that nearly all strong-tier cases stay within a few percent over the full depletion path. This pattern is consistent with the tier-wise statistics in Table 2: across Tiers 1–6 the mean curve MAPE is 1.84%, while when restricting to fluids exhibiting a descent number of close neighbors, i.e. Tiers 1–3, the performance improves to 1.01%. The monotonic increase from Tier 1 (0.84%) to Tier 6 (4.13%) further supports the tiers as a practical confidence label for expected B_o-curve fidelity.

 doi:10.20944/preprints202603.0729.v1

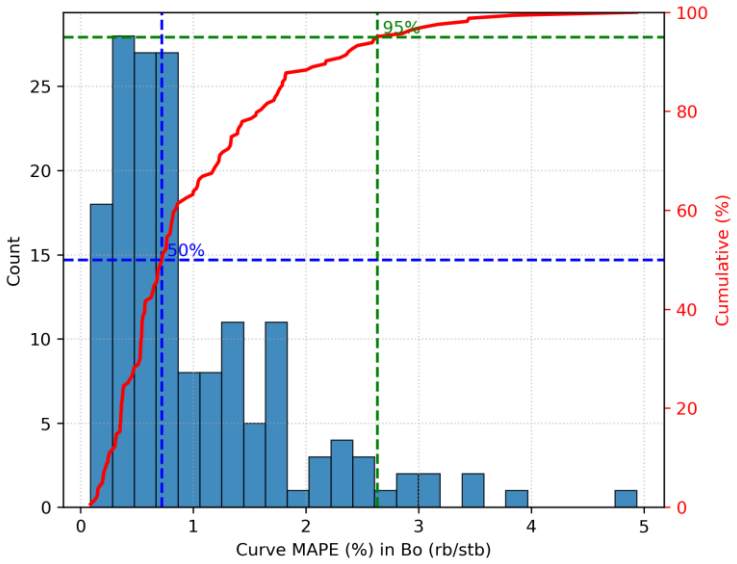


Figure 14. Database-wide accuracy in predicted monophasic B_o curves (Tiers 1-3).

Table 2. Tier-resolved error statistics for B_o curves.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**1.01**	**0.72**	**0.83**
Overall (Tier 1-6)	**323**	**1.84**	**1.04**	**2.94**
1	100	0.84	0.64	0.67
2	44	1.18	1.02	0.83
3	19	1.54	1.21	1.30
4	56	1.53	1.22	1.10
5	37	1.84	1.14	2.71
6	67	4.13	1.94	5.35

Note that B_o curve depends on the LIM prediction at p_{sat} and the relative volume predicted by the CCE LIM approach appearing in Part I of this work[15].

4.2. Oil Density

For the oil density ρ_o (g/cc), the Tier 1–3 examples in Figure 15 demonstrate density curves that remain smooth and closely aligned with laboratory curves over the full depletion path, reflecting robust curve-shape fidelity in the strongest tiers.

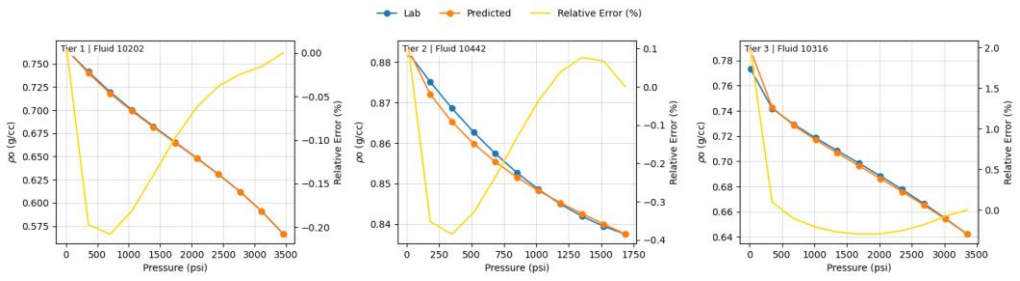


Figure 15. Example inferred ρ_o curves.

Endpoint agreement for ρ_{atm} in Figure 16 is largely tight, with the relative-error histogram peaking near 0% and most fluids falling within roughly ±2%. The distribution shows a slightly longer negative tail, with rare outliers reaching about −5%, while the positive side extends more modestly

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 10 March 2026 doi:10.20944/preprints202603.0729.v1

to around +3.7%. Overall, the endpoint errors suggest clearly minimal systematic bias for the majority of fluids at atmospheric conditions.

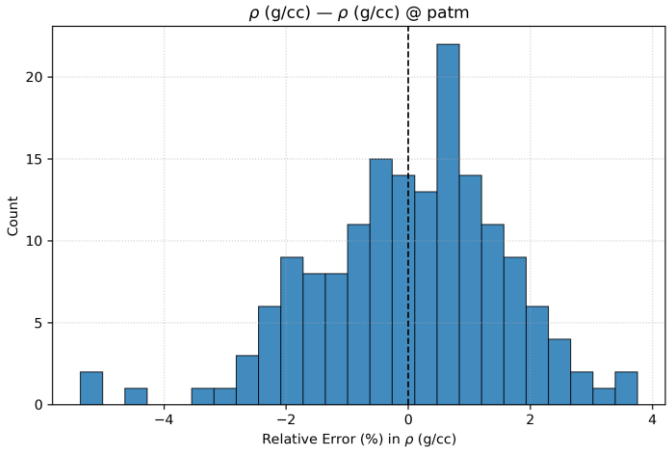


Figure 16. ρ_o endpoint validation at patm (Tier 1–3): relative-error histogram.

At the curve level, Figure 17 shows that ρ_o curve MAPE is tightly concentrated at sub-percent values, with the ECDF reaching the 50th percentile at around 0.4% and the 95th one near 1.45%. This indicates both strong typical performance and a well-controlled upper tail, i.e., only a small minority of cases exceed 1.5% over the full depletion path. The Tier summary in Table 3 is consistent with this stability: the overall mean curve MAPE is 0.79%, improving to 0.51% for Tiers 1–3, with a gradual rise toward weaker neighborhoods (up to 1.46% in Tier 6). It should be noted that the extremely low errors achieved are attributed to the fact that $p@p_{sat}$ is estimated very accurately by the CCE LIM (see Table 1 and [15]).

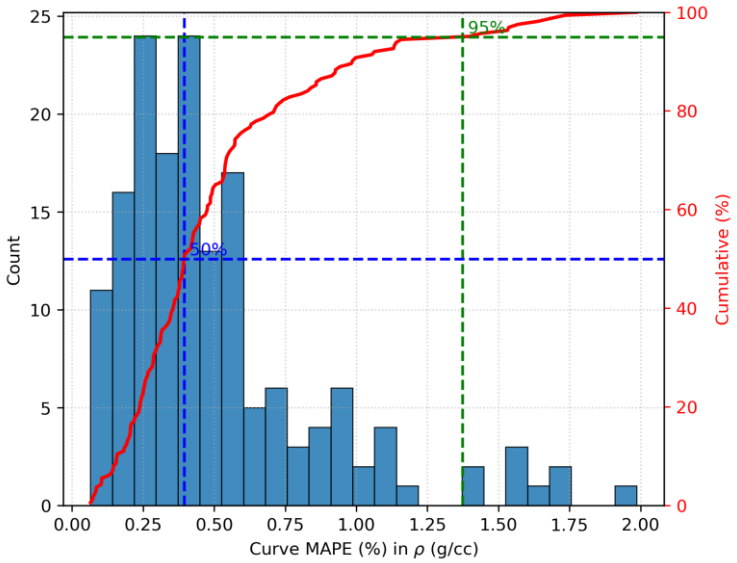


Figure 17. Database-wide accuracy in predicted ρ curves (Tiers 1-3).

Table 3. Tier-resolved error statistics for ρ curves.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**0.51**	**0.39**	**0.37**
Overall (Tier 1-6)	**323**	**0.79**	**0.53**	**0.86**
1	100	0.45	0.40	0.29

 doi:10.20944/preprints202603.0729.v1

2	44	0.63	0.39	0.50
3	19	0.53	0.39	0.30
4	56	0.81	0.73	0.53
5	37	0.77	0.59	0.72
6	67	1.46	1.00	1.45

4.3. Solution Gas

Moving to R_s (scf/stb), the representative Tier 1–3 curves in Figure 18 preserve the correct qualitative depletion behavior and remain smooth across the pressure range, supporting recovery of the dominant $R_s(p)$ curve family under strong neighborhood support. Note that relative errors are higher than those of B_o due to low values of R_s appearing at low pressures which in turn are related to the reducing solution gas left with reducing cell pressure. For example, if R_s at some low-pressure step is reported at 10 scf/stb whereas its predicted counterpart is 12 scf/stb, which corresponds to a relative error of 20%, although the deviation, when evaluated in absolute terms (i.e. scf/stb) is within the accuracy range of the measurements taken at the PVT lab.

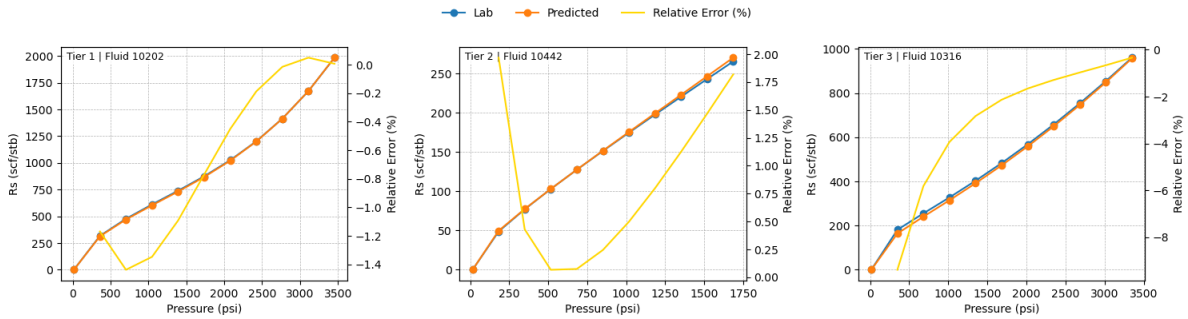


Figure 18. Example inferred R_s curves.

Endpoint agreement for R_s at p_{sat} in Figure 19 is quite strong: most fluids lie close to the identity line, and the relative-error distribution is heavily concentrated at small values (predominantly within 0–10%, with many cases in the 0–5% range). The histogram is right-skewed, with a long but sparse tail of 7 high tier fluids, indicating that a small subset of fluids drives the largest endpoint deviations.

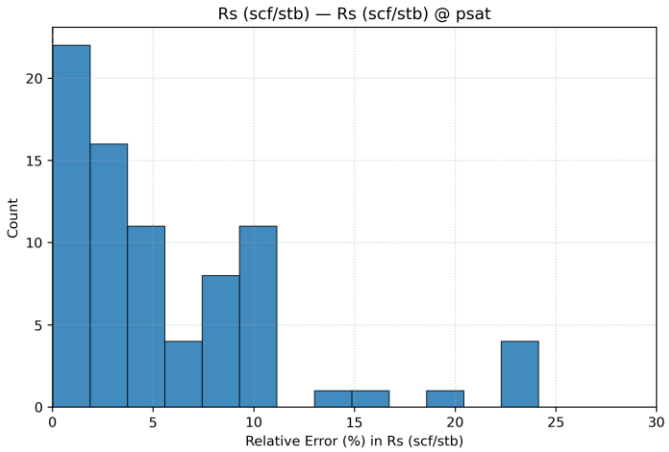


Figure 19. R_s endpoint validation at psat (Tier 1–3): relative-error histogram.

Table 4 reports the mean absolute relative error for the predicted endpoint $R_s@p_{sat}$ across tiers. The results confirm strong endpoint prediction performance (overall mean 6.37%), with Tier 1–3 averaging to 7.03%.

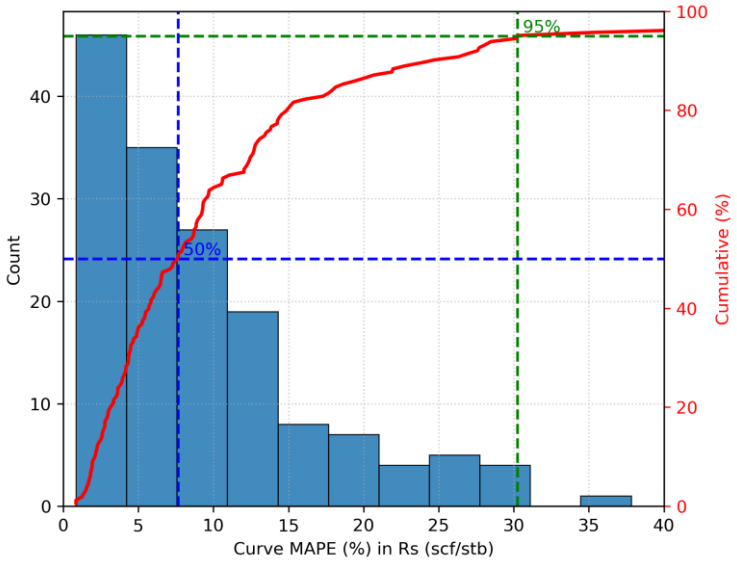
 doi:10.20944/preprints202603.0729.v1

Table 4. Tier-resolved error statistics for R_s at p_{sat}.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**6.37**	**4.23**	**7.20**
Overall (Tier 1-6)	**323**	**7.03**	**4.72**	**8.14**
1	100	6.54	4.71	7.18
2	44	7.04	4.59	7.86
3	19	9.58	6.44	12.54
4	56	5.34	3.86	5.75
5	37	5.47	3.85	6.52
6	67	6.13	4.14	6.18

At the curve level, Figure 20 indicates that most fluids' full R_s curve MAPE fall in the low-error regime with the 50th-percentile located at ~7.5%, while a long upper tail drives the 95th-percentile to ~30%. Overall, this still reflects good performance for the majority of cases, the large MAPE values are largely a scaling effect: when true R_s is small, even modest absolute deviations translate into very large relative errors, so the metric can "skyrocket" for a subset of low-R_s stages/fluids. This "strong majority + tail" structure is consistent with the practical difficulty of gas-associated quantities in DLE workflows, where R_s is sensitive to stagewise gas removal and metering practices, small liberated-gas volumes early in depletion, and uncertainty propagation in derived quantities. Importantly, the reconstructed $R_s(p)$ trends remain smooth and physically plausible; the tail indicates that a minority of fluids dominates the upper envelope rather than broad instability.

Tier summaries in Table 5 align with this interpretation: the overall mean curve MAPE is 11.48%, the Tier 1–3 mean is 10.16%, and mean errors increase toward weaker neighborhoods (Tier 1: 10.67% to Tier 6: 11.23%), confirming that neighborhood strength remains informative for anticipating R_s reliability. An important nuance in Table 5 is that Tiers 4–5 can exhibit lower mean curve MAPE than Tiers 1–3, which should not be interpreted as weaker neighborhoods yielding intrinsically better predictions. This behavior is largely metric-driven: Tiers 4–5 in the PVT repository utilized are dominated by medium-to-high volatility fluids whose R_s values remain higher across much of the pressure range, as opposed to Tiers 1-3 which correspond mostly to low volatility oils. In other words, for more volatile fluids the same absolute mismatch (in scf/stb) produces a smaller percentage error because the denominator is larger; conversely, less volatile fluids with very low late-stage R_s are penalized more strongly by MAPE even when absolute deviations are modest.



 doi:10.20944/preprints202603.0729.v1

Figure 20. Database-wide accuracy in predicted R_s curves (Tiers 1-3). R_s values predictions at low pressures are included.

Table 5. Tier-resolved error statistics for R_s curves.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**10.16**	**6.49**	**11.76**
Overall (Tier 1-6)	**323**	**11.48**	**7.65**	**13.29**
1	100	10.67	7.58	12.44
2	44	12.04	7.66	13.32
3	19	14.45	6.55	17.33
4	56	7.23	5.16	5.77
5	37	6.84	4.81	6.85
6	67	11.23	7.18	13.00

By comparing point to full-curve results, it is shown that the LIM predicted endpoints at p_{sat} are more reliable than the whole curve. It is demonstrated that the larger curve-level errors are not primarily driven by poor anchor placement, but by the behavior of relative-error metrics when R_s becomes small at low pressures, where even modest absolute deviations (within practical lab uncertainty in scf/stb) can translate into superficially inflated percentages. Conclusively, MAPE is a commonly used, still highly biased indicator which needs to be studied with caution to avoid misinterpretation.

4.4. Liberated Gas Z Factor

For the Z factor, the Tier 1–3 examples in Figure 21 show smooth $Z(p)$ curves with close agreement to laboratory trends, indicating strong curve-shape fidelity in the strongest tiers.

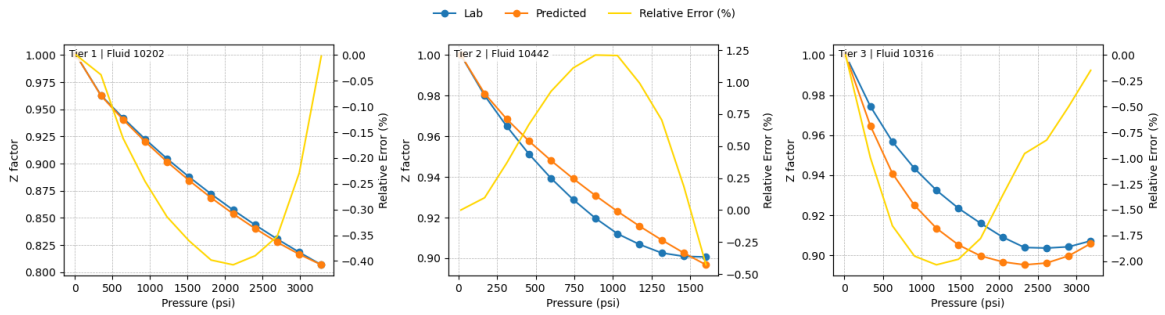


Figure 21. Example inferred Z factor curves.

Endpoint diagnostics in Figure 22 for the Z factor at $0.95p_{sat}$ and $0.1p_{sat}$ show strong overall anchoring, with the error distributions centered close to 0%. The vast majority of the $0.95\,p_{sat}$ endpoint performance is accumulated within the ±5% range, with a few rare excursions reaching roughly −9.5% on the low side and 11.5% on the high side (for some Tier 6 fluids), even though most cases remain within a few percent of zero. In contrast, the $0.1p_{sat}$ endpoint, being closer to the fixed value of 1 at atmospheric pressure, is tighter, with the bulk of errors concentrated near zero (typically within ±1%) and only occasional outliers extending to about −2.2% and +4%.

 doi:10.20944/preprints202603.0729.v1

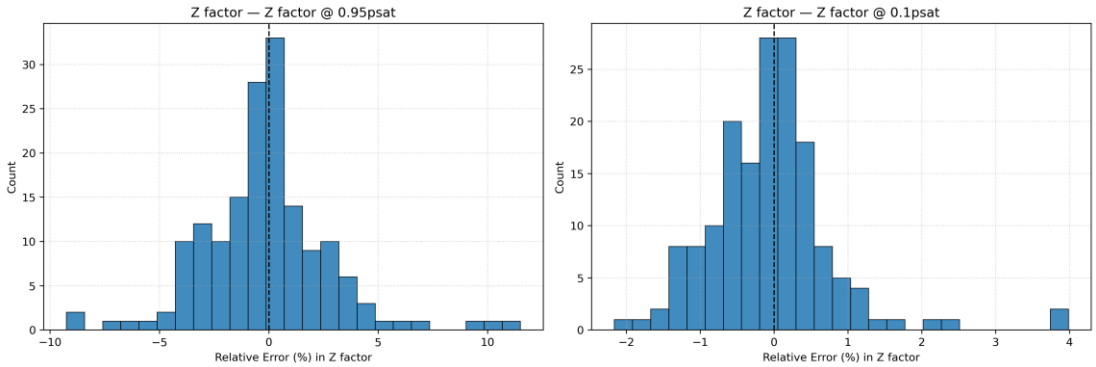


Figure 22. Z endpoint validation at p_{sat} and p_{atm}- (Tier 1–3): relative-error histograms.

After compiling the endpoints to the predicted curve shapes, Figure 23 demonstrates at the curve level that the Z-factor curve MAPE is concentrated at low values with a well-controlled right tail. Most curves lie within the 0–2% range with the ECDF reaching 0.7% for half of the samples while the 95th percentile arrives at 3.9–4.0%. The results indicate consistently good agreement over the full depletion path, while only a single Tier 6 fluid contributing to the upper envelope (extending toward ~10%). The full Tier summary in Table 6 confirms this stability: the overall mean curve MAPE is 1.58%, improving to 1.32% for Tier 1–3, with a gradual rise toward weaker neighborhoods (up to 2.04% in Tier 6).

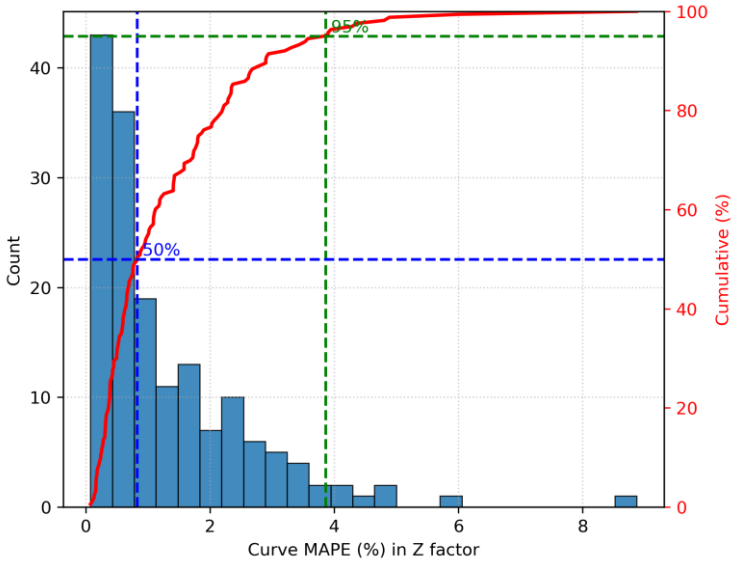


Figure 23. Database-wide accuracy in predicted Z factor curves (Tiers 1-3).

Table 6. Tier-resolved error statistics for Z factor curves.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**1.32**	**0.82**	**1.29**
Overall (Tier 1-6)	**323**	**1.58**	**1.19**	**1.36**
1	100	1.06	0.67	1.09
2	44	1.76	1.50	1.59
3	19	1.63	1.43	1.19
4	56	1.47	1.20	1.08
5	37	2.10	1.55	1.73
6	67	2.04	2.03	1.32

 doi:10.20944/preprints202603.0729.v1

4.5. Gas Volume Factor

B_g (rcf/scf) in Figure 24 shows smooth inferred curves that closely track laboratory behavior across Tier 1–3, with deviations directly comparable to those observed for Z thanks to the linear relationship of the two liberated gas properties.

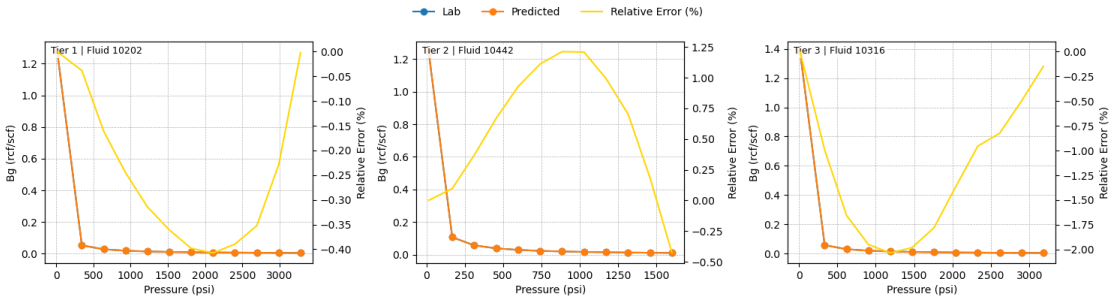


Figure 24. Example inferred B_g curves.

The B_g curve-MAPE distribution in Figure 25 closely mirrors the Z factor distribution, with the ECDF reaching the 95th percentile at 3.9%, which is expected because B_g is computed deterministically from the real-gas relation $B_g = ZRT/p$ on the same pressure grid, so Z deviations propagate directly into B_g. This is also reflected in Table 7, where the overall mean curve MAPE is 1.58%, improving to 1.32% for Tier 1–3 and increasing to 2.04% in Tier 6, showing a consistent monotone trend with neighborhood strength.

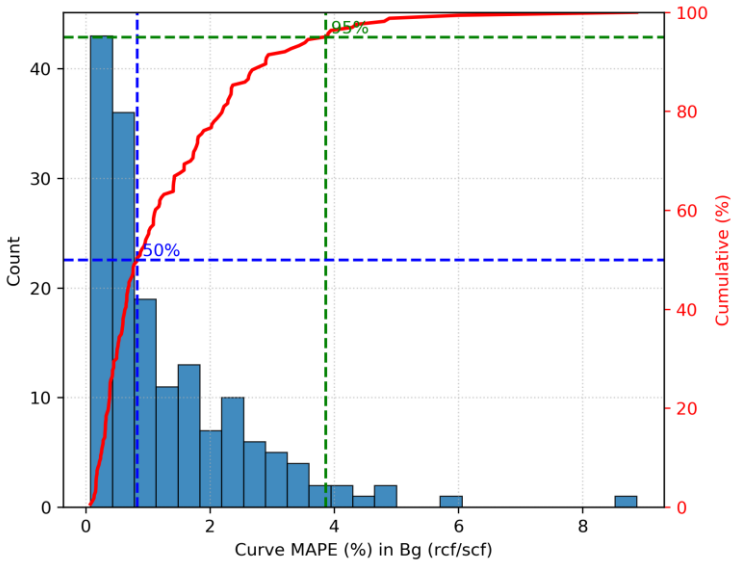


Figure 25. Database-wide accuracy in predicted B_g curves (Tiers 1-3).

Table 7. Tier-resolved error statistics for B_g curves.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**1.32**	**0.82**	**1.29**
Overall (Tier 1-6)	**323**	**1.58**	**1.19**	**1.36**
1	100	1.06	0.67	1.09
2	44	1.76	1.50	1.59
3	19	1.63	1.43	1.19
4	56	1.47	1.20	1.08
5	37	2.10	1.55	1.73

 doi:10.20944/preprints202603.0729.v1

6	67	2.04	2.03	1.32

4.6. Liberated Gas Gravity

The S_g results exhibit intermediate behavior due to the dominating effect of the liberated gas composition rather than that of its volumetric behavior. The Tier 1–3 example curves in Figure 26 reproduce the overall DLE-consistent shape and remain smooth, indicating good qualitative reconstruction.

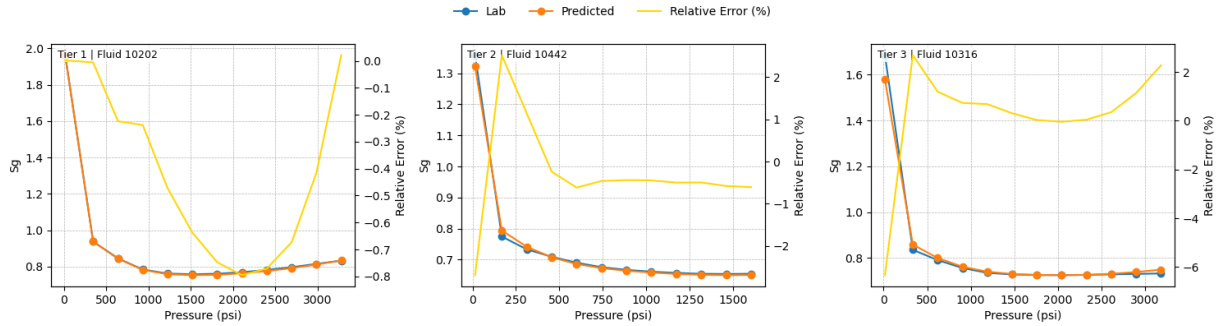


Figure 26. Example inferred S_g curves.

Focusing only at the endpoint diagnostics in Figure 27 for S_g at $0.95p_{sat}$ and S_g at $0.1p_{sat}$, results indicate that errors remain broadly centered near 0%, supporting good overall endpoint anchoring, but with noticeably wider dispersion than the oil phase properties endpoints. At $0.95p_{sat}$, the distribution is strongly concentrated around zero, with the dominant mass within roughly ±10% and only a small number of fluids forming asymmetric tails (down to about −27% and up to ~52%). At $0.1p_{sat}$, the histo-gram remains similarly centered, with most cases falling within approximately −15% to +15%, while with some rare outliers extend to about −35% and up to ~68%. Overall, the endpoints show near-zero systematic bias for the majority of fluids, and the broader range is driven by a limited minority rather than widespread instability.

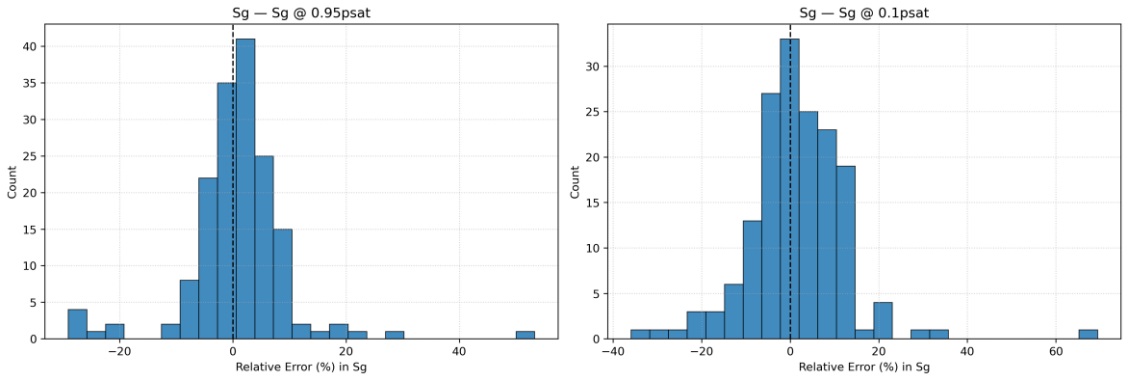


Figure 27. S_g endpoint validation at psat and patm (Tier 1–3): relative-error histograms.

When it comes to the S_g curve-MAPE distribution, shown in Figure 28, the plot is right-skewed, with most curves concentrated at low single-digit errors. Half of the dBase fluids are predicted within less than 5% and while the ECDF reaching reaches the 95th percentile at ~16% whereas half of the dBase fluids are predicted within an error of less than 5%. The remaining spread is driven by a relatively small subset of fluids that form the upper tail. Overall, this still reflects good curve-level behavior for the majority of cases—the larger percentages are not necessarily large absolute mismatches, but often occur when: S_g is small or varies over a narrow range, where even modest absolute deviations translate into inflated relative-error metrics. The Tier statistics in Table 8 are consistent with this interpretation: the overall mean curve MAPE is 5.46%.

 doi:10.20944/preprints202603.0729.v1

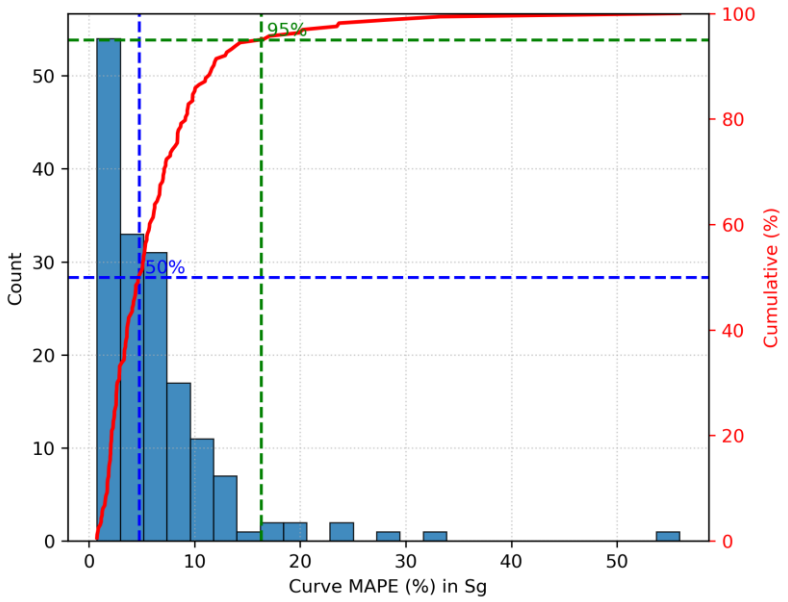


Figure 28. Database-wide accuracy in predicted S_g curves (Tiers 1-3).

Table 8. Tier-resolved error statistics for S_g curves.

Tier	Size	Mean MAPE (%)	Median MAPE (%)	Std MAPE (%)
Overall (Tier 1-3)	**163**	**6.39**	**4.75**	**6.45**
Overall (Tier 1-6)	**323**	**5.46**	**3.95**	**5.31**
1	100	6.37	5.07	6.59
2	44	6.49	5.32	5.39
3	19	6.25	2.87	8.11
4	56	3.34	2.67	2.24
5	37	4.29	2.97	3.20
6	67	5.61	4.25	4.39

4.7. Oil Viscosity

The viscosity workflow described in Section 3 is assessed under a practical, conditioned setting in which μ_{atm} is available and a monophasic viscosity measurement is provided in the undersaturated region above p_{sat}. This monophasic point is used to back-calculate the bubble-point anchor μ_{sat} by inverting Beal's correlation (Eq. 3.1), and this configuration is denoted Variation 0. Table 9 reports endpoint errors at p_{sat} and p_{atm}, together with curve-level errors for the undersaturated ($p \geq p_{sat}$)and saturated ($p \leq p_{sat}$) reconstruction regions.

To quantify sensitivity to data availability, Table 9 also reports three reduced-support configurations. In Variation 1, the monophasic viscosity point is retained but it is provided at a higher pressure in the undersaturated region, increasing the extrapolation distance when inverting Beal to estimate μ_{sat}, while μ_{atm} is no longer imposed and is instead inferred by LIM. In Variation 2, μ_{atm} is again treated as known and imposed as a fixed endpoint for saturated-branch normalization, but no monophasic viscosity point is available, so μ_{sat} is inferred by LIM without the additional p_s constraint. In Variation 3, no auxiliary viscosity information is assumed available, so both μ_{sat} and μ_{atm} are inferred by LIM, yielding the least conditioned case. As expected, performance deteriorates as inputs are removed. Importantly, removing the monophasic point above p_{sat} has a stronger impact than removing the μ_{atm} constraint, because the p_s support directly constrains μ_{sat} and propagates through Beal across the undersaturated branch, whereas μ_{atm} primarily stabilizes the normalization of the saturated-branch reconstruction.

 doi:10.20944/preprints202603.0729.v1

Table 9. Viscosity model variations and performance metrics (mean/median MAPE) for endpoints and curves.

	Available input (μ)			μ_{sat} (MAPE %)		Monophasic Curve (MAPE %)		μ_{atm} (MAPE %)		Diphasic curve (MAPE %)	
Var #	p_s	p_{atm}	p_{sat}	Mean	Median	Mean	Median	Mean	Median	Mean	Median
0	Low p_s	Known	Beal	2.13	1.28	2.18	1.47	0.00	0.00	7.75	5.23
1	High p_s	LIM	Beal	5.38	3.37	2.45	1.60	27.20	16.77	15.82	10.50
2	-	Known	LIM	29.65	17.86	29.15	18.62	0.00	0.00	20.33	13.65
3	-	LIM	LIM	29.65	17.86	29.15	18.62	27.20	16.77	29.19	19.25

To complement the summary metrics, the following figures focus on Variations 0–1 and provide diagnostics at two levels: endpoint fidelity at the bubble-point anchor μ_{sat}, and saturated-branch (diphasic) curve fidelity over the full pressure grid. We omit a p_{atm} endpoint diagnostic and monophasic-branch curve plots because they are not independently discriminative for this comparison: viscosities for $p \geq p_{sat}$are generated through Beal's correlation and therefore largely track the inferred μ_{sat}. The plots are restricted to Tier 1–3 fluids to emphasize differences attributable to conditioning rather than limited neighborhood support.

The signed endpoint relative-error histograms in Figure 29 provide a compact, distribution-level confirmation of the conditioning effect, showing both dispersion and bias in the μ_{sat} estimate (Tier 1–3). In Variation 0, where the workflow leverages the full set of available viscosity inputs, the errors are strongly concentrated around zero with relatively short tails. This tight, near-unbiased clustering indicates that the added conditioning information effectively stabilizes the bubble-point anchor and delivers consistently reliable μ_{sat} estimates across the population. In Variation 1, reducing the available inputs weakens this stabilization: the distribution becomes broader and the tails extend further, with a small subset of fluids driving the largest deviations. Importantly, the dominant mass of the histogram remains centered close to zero, suggesting that the method remains well-behaved for most fluids even under reduced support, while the additional inputs in Variation 0 primarily act to suppress outliers and improve robustness.

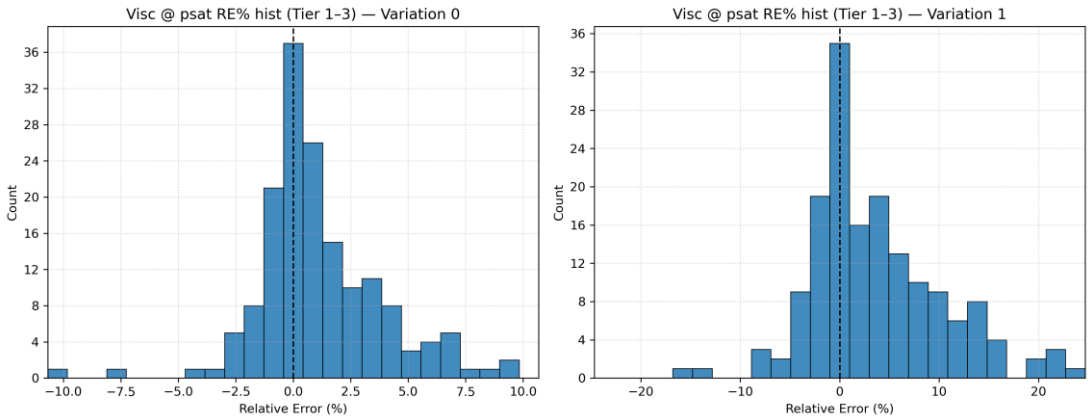


Figure 29. μ_o endpoint validation at p_{sat} (Tier 1–3): relative-error histograms.

 doi:10.20944/preprints202603.0729.v1

The diphasic-curve MAPE histograms with ECDF in Figure 30 show how the viscosity conditioning translates into saturated-branch reconstruction quality for Variations 0–1 (Tier 1–3). In Variation 0, the distribution is strongly weighted toward low MAPE values and the ECDF climbs steeply, meaning that the majority of fluids are reconstructed with small saturated-branch errors. This is also reflected by the percentile markers: the median occurs at a low MAPE (≈ a few percent) and even the 95th percentile remains relatively contained, indicating that the conditioning not only improves the typical case but also suppresses the error tail, i.e., it increases robustness.

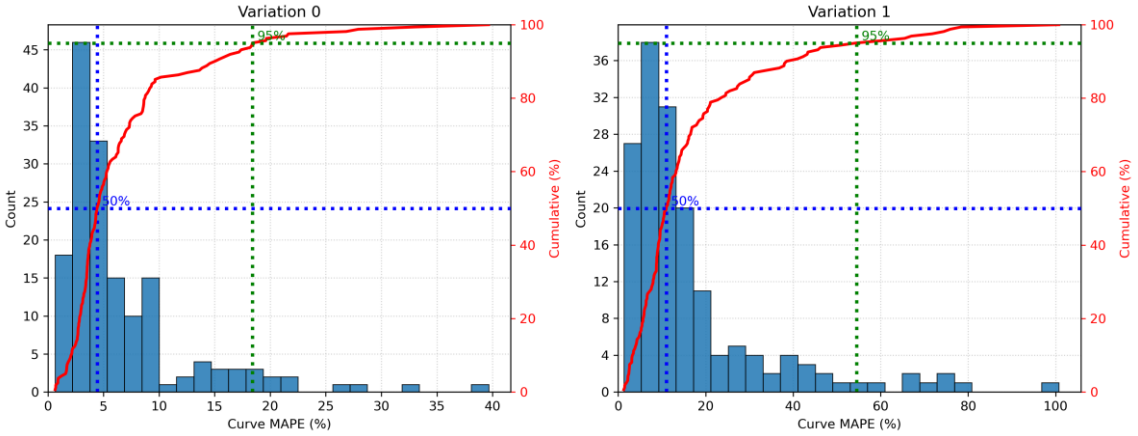


Figure 30. Database-wide accuracy in predicted μ_o curves (Tiers 1-3).

In Variation 1, reducing the available inputs weakens this stabilization. The ECDF rises more gradually and the histogram develops a pronounced right tail, with the median shifting to higher MAPE and the 95th-percentile extending to substantially larger values. Importantly, a sizable portion of the population still falls in the low-to-moderate error range, but the reduced conditioning makes the outcome more sensitive for a subset of fluids, which is what drives the tail expansion.

Overall, the plots reinforce the same physically interpretable conclusion as the summary metrics: the dominant lever is the quality of the μ_{sat} anchor. When μ_{sat} is tightly constrained (Variation 0, and to a lesser extent Variation 1), it both anchors the Beal-based undersaturated reconstruction and stabilizes the saturated-branch unit-square normalization, yielding compact error distributions and a strongly contained tail. By contrast, constraining μ_{atm} alone can help stabilize the saturated-branch scaling, but it cannot recover the robustness lost when μ_{sat} is weakly conditioned. The resulting deterioration is driven primarily by a subset of fluids with large anchor misestimation, which then propagates through both branches.

5. Discussion

The results in Section 4 demonstrate that a proximity-informed, endpoint-constrained curve-reconstruction strategy can reproduce the dominant DLE signatures with engineering-grade fidelity when the target fluid is supported by close neighbors in the archive. For the strongest-support regime (Tiers 1–3), the oil-phase volumetric response is recovered with very low curve-level error: B_o curves remain tightly concentrated in the low-single-digit regime (Tier 1–3 mean curve MAPE ≈ 1.01%; Table 2) and oil-density curves are even tighter (Tier 1–3 mean curve MAPE ≈ 0.51%; Table 3), with representative examples preserving smoothness and the expected monotonic trends across the depletion path. This behavior is consistent with the method design: anchors constrain boundary behavior, while unit-square shape learning borrows only locally coherent curve families, reducing the risk of spurious oscillations and nonphysical curvature.

Liberated-gas volumetrics are similarly stable over the modeled pressure window. Z-factor curves exhibit low curve-level error (Tier 1–3 mean curve MAPE ≈ 1.32%; Table 5) and a controlled upper tail (95th-percentile level near 4% in the strong tiers), indicating that the local neighborhood

 doi:10.20944/preprints202603.0729.v1

contains sufficiently consistent gas-volumetric behavior for reliable reconstruction. Because B_g is computed deterministically from Z through $B_g = ZRT/p$ on the same grid, B_g accuracy mirrors Z accuracy by construction (Table 7), which is operationally useful: it enforces internal consistency of liberated-gas volumetrics without requiring B_g to be learned as an independent target and prevents incoherent Z–B_g combinations from propagating into downstream workflows.

R_s and S_g exhibit broader, right-tailed error distributions, and this is expected for both physical and metric-related reasons. In DLE depletion tests, regardless of its value at saturation conditions, R_s approaches small values at lower pressures. When the denominator is small, even modest absolute deviations translate into inflated relative errors, so curve MAPE can increase sharply for a subset of stages and fluids even when the mismatch remains moderate in scf/stb. This mechanism explains why most R_s curves cluster in a "good" regime while a minority forms a long tail (Figure 14; Table 5). A similar amplification can occur for S_g when the property varies over a narrow range. In addition, S_g is driven primarily by the liberated-gas composition (GC-derived) rather than by volumetric behavior alone, making it inherently more sensitive to compositional variability and stagewise gas-handling effects. Even so, the S_g distribution indicates that the majority of curves remain at low single-digit errors and the 95th percentile is on the order of 16% in the strong tiers (Figure 22; Table 8), while reconstructed trends remain smooth and DLE-consistent.

Oil viscosity is the most demanding target because viscosity endpoints are highly sensitive to experimental conditions, sample handling, and compositional effects. The adopted formulation preserves the same shape-preserving logic used throughout the work by treating $\mu_o(p)$ as two monotonic branches meeting at p_{sat}: the saturated (two-phase) branch below p_{sat} is reconstructed in the unit-square space, while the undersaturated (single-phase) branch above p_{sat} is generated with Beal correlation anchored at $\mu_o(p_{sat})$ (Section 3). In the practical conditioned baseline (Variation 0), a low-pressure reference $\mu_o(p_{atm})$ is available (in practice reported as a dead/stock-tank oil viscosity near atmospheric pressure) and a monophasic viscosity point just above p_{sat} is used to back-calculate $\mu_o(p_{sat})$ via inverse Beal (short extrapolation distance). Under this setting, the bubble-point endpoint is recovered with low error ($\mu_o(p_{sat})$ mean 2.13%, median 1.28%), and the saturated-branch reconstruction is correspondingly tight (diphasic curve mean 7.75%, median 5.23%; Table 9). These levels are encouraging given the intrinsic noise floor of viscosity measurements, and the reconstructed curves remain smooth and physics-consistent, capturing the expected viscosity minimum at p_{sat} and the post-bubble-point increase.

Table 9 also quantifies sensitivity when this conditioning is relaxed. In Variation 1, the monophasic viscosity point is retained but provided at a higher undersaturated pressure, increasing the extrapolation distance in the inverse-Beal step used to estimate $\mu_o(p_{sat})$, while $\mu_o(p_{atm})$ is no longer imposed and is instead inferred by LIM. Consistent with this reduced support, the $\mu_o(p_{sat})$ endpoint error increases (mean 5.38%, median 3.37%) and the saturated-branch curve errors rise (diphasic curve mean 15.82%, median 10.50%; Table 9). Notably, the dominant deterioration is not a systematic bias shift but an increase in dispersion driven by the less constrained low-pressure anchor ($\mu_o(p_{atm})$ mean 27.20%, median 16.77%). Importantly, however, the availability of any monophasic point above p_{sat} still keeps the bubble-point anchor within a relatively controlled range compared with the fully unconditioned cases. The comparison between Variations 0 and 1 therefore highlights a practically actionable point: placing even a single monophasic viscosity measurement close to p_{sat} is the most effective lever, because it directly stabilizes $\mu_o(p_{sat})$ and propagates through the Beal-based undersaturated branch and the saturated-branch normalization over the monophasic region. Additional low-pressure information further improves robustness by suppressing tail behavior.

A central practical feature of the framework is that applicability is tied to neighborhood support via tiering. Across properties, performance degrades gradually as similarity support weakens, rather than failing abruptly, and this behavior is particularly clear for oil volumetrics (e.g., B_o mean curve MAPE increasing from Tier 1 toward Tier 6; Table 2) and for density (Table 3). This makes tiers a usable confidence label in an engineering workflow: strong tiers support direct curve generation for

 doi:10.20944/preprints202603.0729.v1

simulation-ready inputs, while weaker tiers serve as a flag for additional scrutiny (e.g., manual review, request for new lab data, or EoS tuning supported by additional measurements).

It has been demonstrated how the method's "local" design mirrors how PVT experts work in practice, identify the most relevant analogs first, then infer behavior within a thermodynamically coherent family, while retaining the reproducibility and scalability of an automated pipeline. The Taylor step on top of the nearest neighbors selection acts the way an expert could "correct" the neighbor PVT value while moving towards the test fluid. For example, increasing temperature and volatility both contribute to the increase of the B_o and R_s values at saturation conditions, whereas heavier API and lab flash GOR values act oppositely. Moreover, the "distancing criteria" derived replicate the expertise of a fluid engineer when deciding which fluids in the PVT reports repository do exhibit sufficient similarity to the test fluid.

Finally, the key limitation of any proximity-informed AI model is dependence on archive coverage. Prediction quality is fundamentally bounded by whether representative analogs exist for the target fluid and depletion regime. Tiering makes this limitation explicit but does not remove it. Continued growth and QA/QC of the PVT library therefore remains important for robust deployment across diverse fluids.

6. Conclusions

This study extends a domain-driven, physics-backed, proximity-informed modeling framework from Constant Composition Expansion[15] (Part I) to Differential Liberation (DLE) and viscosity tests, explicitly addressing the curve-structured and path-dependent nature of depletion below saturation pressure. DLE test prediction is formulated as a multi-level problem that combines local interpolation for physically meaningful anchor values (endpoints at reference pressures) and shape learning of normalized intermediate trajectories in a unit-square representation, followed by endpoint-constrained reconstruction of full pressure-dependent curves. Gas properties are handled on pressure windows consistent with gas presence and metering, and B_g is derived deterministically from predicted Z to ensure internal volumetric consistency by construction.

Evaluation on a curated DLE database using a leave-one-out framework shows that the method reproduces the principal DLE signatures with very good agreement to laboratory data, particularly under strong neighborhood support. Oil volumetrics (B_o) and oil density achieve low curve-level errors for Tiers 1–3, liberated-gas volumetrics (Z and derived B_g) remain stable with controlled tails, and more sensitive quantities such as R_s and S_g retain smooth, physically plausible depletion trends while exhibiting broader right-tailed relative-error distributions consistent with low-magnitude stages and compositional sensitivity. Overall, the results support proximity-informed DLE tests surrogates as a practical way to reduce reliance on new DLE test measurements to save the cost of fluid sampling and analysis, while providing an interpretable confidence signal through neighborhood strength that aligns with engineering decision-making.

For viscosity, accuracy is governed primarily by how well the bubble-point anchor $\mu_o(p_{sat})$ is conditioned. In the practical conditioned baseline (Variation 0), a low-pressure reference $\mu_o(p_{atm})$ is available and a monophasic viscosity point just above p_{sat} enables a short-range inverse-Beal estimate of $\mu_o(p_{sat})$, yielding low single-digit errors for the bubble-point endpoint and tight saturated-branch (diphasic) curve fidelity in the strong tiers. In Variation 1, conditioning is weakened because the monophasic point is moved farther above p_{sat} and $\mu_o(p_{atm})$ is no longer imposed (it is inferred by LIM), increasing the effective degrees of freedom in the reconstruction. As a result, errors broaden and the diphasic-curve tail becomes heavier, while the dominant mass of the distribution remains centered near moderate error levels. Overall, the results support proximity-informed DLE and viscosity AI models as a practical way to reduce reliance on extensive laboratory programs while maintaining engineering-grade fidelity, with tiering providing an interpretable confidence signal that aligns with engineering decision-making.

 doi:10.20944/preprints202603.0729.v1

Author Contributions: Conceptualization, S.P.F. and V.G.; methodology, S.P.F., E.M.K., and V.G.; software, S.P.F., E.M.K., J.K. and A.M.; validation, all; resources, A.M.; data synthesis, E.M.K. with supervision from V.G. and A.M.; data QA-QC, A.S. with supervision from V.G. and A.M. writing—original draft preparation, E.M.K.; writing—review and editing, V.G., S.P.F., J.N. and A.M.; visualization, E.M.K.; supervision, V.G., J.N. and A.M.; project administration, A.M.; funding acquisition, A.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research activities in Canada received funding from National Research Council (NRC) Industrial Research Assistance Program (IRAP). The work in Greece and US was carried out using internal Fluidsdata resources as part of the company's research and development activities.

Data Availability Statement: The data presented in this study are not publicly available because they are proprietary to Fluidsdata and are subject to commercial confidentiality restrictions.

Conflicts of Interest: Authors Sofianos Panagiotis Fotias, Eirini Maria Kanakaki, Afzal Memon, Anna Samnioti, Jahir Khan, John Nighswander, and Vassilis Gaganis are employed by Fluidsdata.

Nomenclature

The following abbreviations are used in this manuscript:

LIM	Local Interpolation Method
DLE	Differential Liberation Expansion
CCE	Constant Composition Expansion
B_o	Differential oil formation volume factor (rb/stb)
R_s	Differential solution gas-oil ratio (scf/stb)
ρ_o	Oil density (g/cc)
Z	Compressibility factor
B_g	Differential gas formation volume factor (rcf/scf)
S_g	Gas specific gravity
p_{atm}	Atmospheric pressure (psi)
p_{sat}	Saturation pressure (psi)
p_b	Bubble-point pressure (psi)
μ_o	Oil viscosity (cP)
MAPE	Mean Absolute Percentage Error
ECDF	Empirical Cumulative Distribution Function

References

1. Amyx, J. W.; Bass, D. M.; Bass, D. M.; Whiting, R. L. *Petroleum Reservoir Engineering: Physical Properties*; McGraw-Hill, 1960.
2. *Petroleum Reservoir Engineering*; Cameron, J., Ed.; Syrawood Publishing House: New York, 2019.
3. Chierici, G. L. *Principles of Petroleum Reservoir Engineering*; Springer: Berlin, Heidelberg, 1995. https://doi.org/10.1007/978-3-642-78243-5.
4. Dandekar, A. Y. *Petroleum Reservoir Rock and Fluid Properties*, 2nd ed.; CRC Press: Boca Raton, 2013. https://doi.org/10.1201/b15255.
5. McCain, W. D.; Spivey, J. P.; Lenn, C. P. *Petroleum Reservoir Fluid Property Correlations*; PennWell Corporation, 2011.
6. Whitson, C. H.; Brulé, M. R. *Phase Behavior*; Society of Petroleum Engineers, 2000. https://doi.org/10.2118/9781555630874.
7. Pedersen, K. S.; Christensen, P. L.; Shaikh, J. A. *Phase Behavior of Petroleum Reservoir Fluids*, 2nd ed.; CRC Press: Boca Raton, 2014. https://doi.org/10.1201/b17887.
8. Tewari, R. D.; Dandekar, A. Y.; Ortiz, J. M. *Petroleum Fluid Phase Behavior: Characterization, Processes, and Applications*; CRC Press: Boca Raton, 2018. https://doi.org/10.1201/9781315228808.

 doi:10.20944/preprints202603.0729.v1

9. Moses, P. L. Engineering Applications of Phase Behavior of Crude Oil and Condensate Systems (Includes Associated Papers 16046, 16177, 16390, 16440, 19214 and 19893). *J. Pet. Technol.* **1986**, *38* (07), 715–723. https://doi.org/10.2118/15835-PA.
10. Di Primio, R.; Dieckmann, V.; Mills, N. *PVT* and Phase Behaviour Analysis in Petroleum Exploration. *Org. Geochem.* **1998**, *29* (1), 207–222. https://doi.org/10.1016/S0146-6380(98)00102-8.
11. Chapter 2 Pvt Analysis for Oil. In *Developments in Petroleum Science*; Elsevier, 1978; Vol. 8, pp 45–71. https://doi.org/10.1016/S0376-7361(08)70008-6.
12. Kanakaki, E. M.; Gaganis, V. Automated Equations of State Tuning Workflow Using Global Optimization and Physical Constraints. *Liquids* **2024**, *4* (1), 261–277. https://doi.org/10.3390/liquids4010012.
13. Jacoby, R.; Yarborough, L. PVT MEASUREMENTS ON PETROLEUM RESERVOIR FLUIDS AND THEIR USES. *Ind. Eng. Chem.* **1967**, *59* (10), 48–62. https://doi.org/10.1021/ie50694a010.
14. Papanikolaou, P.; Kanakaki, E. M.; Lempesis, S.; Gaganis, V. Mass Balance-Based Quality Control of PVT Results of Reservoir Oil DL Studies. *Energies* **2024**, *17* (13), 3301. https://doi.org/10.3390/en17133301.
15. Fotias, S. P.; Kanakaki, E. M.; Gaganis, V.; Samnioti, A.; Khan, J.; Nighswander, J.; Memon, A. A Domain-Driven, Physics-Backed, Proximity-Informed AI Model for PVT Predictions—Part I: Constant Composition Expansion. Preprints January 20, 2026. https://doi.org/10.20944/preprints202601.1522.v1.
16. Zhao, H.; Memon, A.; Gao, J.; Taylor, S. D.; Sieben, D.; Ratulowski, J.; Alboudwarej, H.; Pappas, J.; Creek, J. Heavy Oil Viscosity Measurements: Best Practices and Guidelines. *Energy Fuels* **2016**, *30* (7), 5277–5290. https://doi.org/10.1021/acs.energyfuels.6b00300.
17. Lohrenz, J.; Bray, B. G.; Clark, C. R. Calculating Viscosities of Reservoir Fluids From Their Compositions. *J. Pet. Technol.* **1964**, *16* (10), 1171–1176. https://doi.org/10.2118/915-PA.
18. Beal, C. The Viscosity of Air, Water, Natural Gas, Crude Oil and Its Associated Gases at Oil Field Temperatures and Pressures. *Trans. AIME* **1946**, *165* (01), 94–115. https://doi.org/10.2118/946094-G.
19. Craft, B. C.; Hawkins, M.; Terry, R. E. *Applied Petroleum Reservoir Engineering*; Pearson College Div: Englewood Cliffs, N.J, 1991.