WELLBENCH: AN OPEN-SOURCE BENCHMARK FOR SYNTHETIC WELL LOG GENERATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Pore pressure prediction is essential for safe well planning, yet machine learning models are limited by the scarcity of real well data. Synthetic data generation is a common approach, including interpolation methods, physics-informed generators, and data-driven models. To our knowledge, no prior work systematically compares these approaches under a matched hyperparameter-tuning budget using both distributional and downstream evaluation. We introduced WellBench, a multi-basin benchmark, and a Physics-Optimized Forward Model (POFM), a physics-based scientific simulator for synthetic well-log generation that integrates petrophysical relationships and is calibrated to real well-log distributions using Tree-structured Parzen Estimation (TPE). We compared POFM against a TPE-tuned CTGAN, SMOTE, and Smoothed Bootstrap across four regions. POFM achieved the highest Distance-to-Closest-Record (DCR), indicating the greatest separation from real training records, while the interpolation generators best reproduced the real marginals but yielded the lowest DCR. Across seven downstream models, only the interpolation generators achieved positive blind-well Train-Synthetic-Test-Real (R^2) (up to ≈ 0.98), while CTGAN and POFM yielded negative R^2. Under a depth-held-out split, Train-Real-Test-Real (TRTR) failed at every real-data fraction, highlighting the challenge posed by limited real-data coverage. Yet TSTR with interpolation-based generators outperformed TRTR at every fraction. These results reveal a fundamental tradeoff between distributional fidelity and downstream utility. They also show that interpolation-based generators can outperform complex synthetic-data approaches when real data is scarce.

1 INTRODUCTION

Pore pressure, the pressure exerted by fluids within a rock's pores, must be predicted to avoid situations such as kicks and blowouts (Khoshnaw et al., 2015). For this reason, drilling operators rely on models to convert sonic and density logs into a pressure curve. However, the calibration of these physical models requires data that can be scarce (Gao et al., 2023).

This scarcity of data has motivated machine learning approaches that aim at learning pore pressure from well logs. The main limitation lies less in the choice of machine learning model than in the lack of available data. To solve this issue, synthetic well log generation can be used to produce labeled samples (Fredere et al., 2026). If the synthetic logs are realistic, machine learning models trained on them may generalize to the unobserved deeper intervals of real wells.

Based on the literature, we consider two families of synthetic generators. The first is physics-informed: a chain of petrophysical equations whose parameters are tuned to a target basin, giving depth extrapolation and expert-interpretable parameters. The second is data-driven such as the Conditional Tabular GAN (CTGAN) (Xu et al., 2019) and two interpolation resamplers, SMOTE (Chawla et al., 2002) and Smoothed Bootstrap.

The two approaches optimize different objectives. Physics-informed generators restrict their outputs to physically valid ranges (Antonion et al., 2024), whereas data-driven generators reproduce the empirical distribution (Goodfellow et al., 2014). To our knowledge, no prior work compares both families on a unified multi-basin benchmark, tunes each with the Tree-structured Parzen Estimator (TPE), and evaluates distributional fidelity together with downstream pore-pressure R^2 under depth-held-out TSTR/TRTR splits. This study provides that comparison. The depth split mirrors look-ahead

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

drilling: having drilled the shallow interval, can we predict the deeper one? We therefore ask how much generative complexity is needed to predict unobserved deeper intervals from already-drilled shallow ones.

The primary contribution is a benchmark and its empirical findings rather than a new generator:

- WellBench, an open-source, multi-basin benchmark for synthetic well-log generation with four calibrated generators (including POFM, a physics-based generator calibrated per region by TPE), per-region parameters, and fixed seeds.
- A controlled budget comparison of physics-based, generative (CTGAN), and interpolation-based synthetic-data approaches.
- An evaluation connecting synthetic-data fidelity and novelty (DCR) with downstream ML transfer (TSTR/TRTR).
- The finding that greater physical structure and novelty do not necessarily improve downstream predictive utility under data scarcity.

2 Related Work

Machine learning approaches pore-pressure prediction as a regression task; on four Potwar Basin wells, the basin of our Zones 1–2, a meta-ensemble reached $R^2 = 0.93$ (Amjad et al., 2025). Data scarcity raises questions about whether such models generalize across basins. Synthetic data has addressed scarcity in many fields (Zhao et al., 2024), using methods such as VAEs, GANs, and interpretable data generators (Lautrup et al., 2024). The closest prior work generated synthetic well logs with a lithology-conditioned VAE and evaluated them on classification in a held-out well (Fredere et al., 2026); we instead compare physics-based, adversarial, and interpolation generators on pore-pressure regression. Earlier well-log studies assessed physics-informed and data-driven paradigms using different criteria. As such, this justifies our unified distributional and downstream evaluation (Huang & Tian, 2024). The classic workflow estimates overburden from bulk density, defines a sonic-log normal compaction trend (NCT), and uses Eaton's exponent to convert NCT deviation into pore pressure (Eaton, 1975). The exponent varies from basin to basin and is determined from direct measurements. CTGAN uses mode specific normalization and a conditional vector to reduce mode collapse in imbalanced tabular data (Xu et al., 2019). Tree-structured Parzen Estimation (TPE) (Bergstra et al., 2011) is a Bayesian optimization method designed for high-dimensional search spaces. Often, it is used through the Optuna library (Akiba et al., 2019). Train-Synthetic-Test-Real (TSTR) measures whether a model trained only on synthetic data generalizes to real test data and semi-TSTR mixes in real data (5–80% here). Train-Real-Test-Real (TRTR) trains on the same fraction of real data only and is the real-data baseline (Santangelo et al., 2025). TSTR is a stricter synthetic-data quality test than distributional metrics because it represents feature target conditional structure.

3 Method

3.1 Geological Regions

This work evaluated four geological zones (Table 1). In Zones 1 and 2, pore pressure measurements are available. In Zones 3 and 4, the pore pressure is not available. Therefore, in Zones 3 and 4, only distributional reliability was evaluated.

Table 1: Summary of geological regions used for benchmarking.

Region	Location	Wells	Log Variables
Zone 1: Missa Keswal	Eastern Potwar Basin, Pakistan	3	8
Zone 2: Joyamair–Minwal	Eastern Potwar Basin, Pakistan	3	8
Zone 3: IODP U1343E	Bering Sea, USA	1	4
Zone 4: Volve	Norwegian North Sea, Norway	5	4

Zone 1 and Zone 2 each have 3 wells. Every well in both zones includes GR, DT, RHOB, RES_DEEP, and derived HP, OB, DT_NCT, and PPP. Zones 3–4 contain just the four main logs altogether. The per-field lithologies and well identifiers are in Appendix A.

3.2 Synthetic Generators

The POFM combines six petrophysical formulas (Athy compaction, Raymer–Hunt–Gardner, density mixing, Archie, a shale-index GR map, and a hydrostatic/overburden/NCT/Eaton pressure chain) into one depth-extrapolating generator (Algorithm 1). Outputs are clipped to physical bounds (Table 2). A reduced Zones 3–4 variant outputs only five rock-property logs (Algorithm 2, Appendix C). POFM has more than 20 hyperparameters that were tuned by TPE (500 Optuna trials per region) against a softmax-weighted sum of four fidelity criteria (JSD, WD, Frobenius, ACF). POFM was then re-fit over 20 seeds.

Algorithm 1 POFM

Require: Depth vector $\mathbf{z}$, hyperparameters $\boldsymbol{\theta}$, random seed s
Ensure: Synthetic log DataFrame $\mathbf{D}$

$\phi \leftarrow \texttt{AthyPorosity}(\mathbf{z}, \boldsymbol{\theta}, s)$ ▷ Eq. equation 1
DT $\leftarrow \texttt{RHG}(\phi, \boldsymbol{\theta}, s+10)$ ▷ Eq. equation 2
RHOB $\leftarrow \texttt{DensityMix}(\phi, \mathbf{z}, \boldsymbol{\theta}, s+20)$ ▷ Eq. equation 3
RES $\leftarrow \texttt{Archie}(\phi, \boldsymbol{\theta}, s+30)$ ▷ Eq. equation 4
GR $\leftarrow \texttt{ShaleGR}(\phi, \boldsymbol{\theta}, s+40)$ ▷ Eq. equation 5
HP $\leftarrow \texttt{Hydrostatic}(\mathbf{z}, \boldsymbol{\theta})$ ▷ Eq. equation 6
OB $\leftarrow \texttt{Overburden}(\mathbf{z}, \text{RHOB}, \boldsymbol{\theta})$ ▷ Eq. equation 7
$\text{DT}_{\text{NCT}} \leftarrow \texttt{NCT}(\mathbf{z}, \text{DT}, \boldsymbol{\theta})$ ▷ Eq. equation 8
PPP $\leftarrow \texttt{Eaton}(\text{DT}, \text{DT}_{\text{NCT}}, \text{OB}, \text{HP}, \boldsymbol{\theta})$ ▷ Eq. equation 9
return $\mathbf{D} = \{\mathbf{z}, \text{GR}, \text{DT}, \text{RHOB}, \text{RES}, \text{HP}, \text{OB}, \text{DT}_{\text{NCT}}, \text{PPP}\}$

Step 1: Porosity via Athy's compaction law. Depth-dependent porosity is modeled as

$$\phi(z) = \phi_0 \exp(-c \cdot z) + A_\ell \sin\left(\frac{z}{\lambda_\ell}\right) + \epsilon_\phi, \qquad \epsilon_\phi \sim \mathcal{N}(0, \sigma_\phi^2), \tag{1}$$

where ϕ_0 is the surface porosity, c is the compaction coefficient, A_ℓ and λ_ℓ define lithological layering. ϵ_ϕ adds geological heterogeneity (for the remaining variables, ϵ has the analogous role). The result is clipped to $[0.02, 0.50]$.

Step 2: Sonic transit time via the Raymer–Hunt–Gardner relation.

$$\text{DT}(z) = \phi(z)\,\text{DT}_{\text{fluid}} + \big(1 - \phi(z)\big)\,\text{DT}_{\text{ma}} + \epsilon_{\text{DT}}, \qquad \epsilon_{\text{DT}} \sim \mathcal{N}(0,\, \sigma_{\text{DT}}^2), \tag{2}$$

where DT_{ma} is the matrix transit time in μs/ft.

Step 3: Bulk density via the density mixing law.

$$\text{RHOB}(z) = \phi(z) \cdot \rho_f + \big(1 - \phi(z)\big) \cdot \big(\rho_{\text{ma}} + 0.1 \sin(z/120)\big) + \epsilon_\rho, \tag{3}$$

where ρ_f and ρ_{ma} are the fluid and matrix densities and the sinusoidal term introduces lithological variation.

Step 4: Deep resistivity via Archie's law.

$$R_t = \frac{a \cdot R_w}{\phi^m \cdot S_w^n}, \qquad \log_{10}(\widehat{R}_t) = \log_{10}(R_t) + \epsilon_R, \quad \epsilon_R \sim \mathcal{N}(0, \sigma_R^2), \tag{4}$$

with tortuosity factor a, cementation exponent m, saturation exponent n, fluid resistivity R_w, and fluid saturation S_w.

Step 5: Gamma ray from shale index.

$$\text{GR}(z) = \text{GR}_{\text{shale}} + \frac{\phi(z) - 0.02}{0.48}\big(\text{GR}_{\text{clean}} - \text{GR}_{\text{shale}}\big) + \epsilon_{\text{GR}}, \qquad \text{GR} \in [0,\ 200]\ \text{API}. \tag{5}$$

Step 6: Derived pressure quantities. The hydrostatic pressure assumes a static fluid column of density $\rho_{f,\mathrm{HP}}$:

$$\mathrm{HP}(z) = 1.4223 \cdot \rho_{f,\mathrm{HP}} \cdot z, \tag{6}$$

where the constant 1.4223 converts g/cc·ft to psi.

The overburden stress is initialised at the shallowest measurement using an average shallow bulk density $\bar{\rho}_{\mathrm{shallow}}$ and accumulated by trapezoidal integration of the RHOB log:

$$\mathrm{OB}(z_0) = \bar{\rho}_{\mathrm{shallow}} \cdot 1.4223 \cdot z_0, \qquad \mathrm{OB}(z_i) = \mathrm{OB}(z_0) + 1.4223 \sum_{j=1}^{i} \frac{\mathrm{RHOB}_{j-1} + \mathrm{RHOB}_j}{2} \Delta z_j. \tag{7}$$

The normal compaction trend blends a smoothed version of the observed sonic log with a theoretical exponential decay, controlled by the weight w:

$$\mathrm{DT}_{\mathrm{NCT}}(z) = w \cdot \overline{\mathrm{DT}}_{\mathrm{smooth}}(z) + (1 - w)\big[\mathrm{DT}_0 + (\mathrm{DT}_s - \mathrm{DT}_0)e^{-kz}\big]. \tag{8}$$

Pore pressure is estimated via Eaton's method, which quantifies the deviation of the observed sonic from the normal compaction trend:

$$\mathrm{PPP}(z) = \mathrm{OB} - (\mathrm{OB} - \mathrm{HP}) \left(\frac{\mathrm{DT}}{\mathrm{DT}_{\mathrm{NCT}}} \right)^n. \tag{9}$$

The calibration aggregates four fidelity criteria with a softmax-weighted sum: $\mathcal{L}_{\mathrm{JS}}$, the Jensen–Shannon divergence per-variable averaged across log variables; $\mathcal{L}_{\mathrm{WD}}$, the analogous Wasserstein distance; $\mathcal{L}_{\mathrm{Frob}}$, the Frobenius distance between real and synthetic cross-log correlation matrices; and $\mathcal{L}_{\mathrm{ACF}}$, the RMSE between their autocorrelation functions. The total loss is

$$\mathcal{L} = \sum_k w_k \mathcal{L}_k, \qquad w_k = \frac{\exp(\ell_k)}{\sum_j \exp(\ell_j)}, \tag{10}$$

where the logits ℓ_k are themselves TPE-tuned search-space variables. This way allowing the optimizer to discover which criterion is best for a given basin. Each component $\mathcal{L}_k$ is divided by a fixed baseline scale (untuned-generator values: 0.55, 1.50, 3.50, 0.25 for JSD/WD / Frobenius/ACF) before softmax weighting, so the learned logits express fraction-of-baseline rather than raw magnitudes.

Table 2: Physical bounds applied to each synthetic log variable after generation. Values outside these ranges are clipped to enforce physical plausibility for hydrocarbon-bearing settings.

Variable	Min	Max	Unit
GR	0	200	API
DT	30	180	μs/ft
RHOB	1.2	2.9	g/cc
RES_DEEP	0.01	10000	Ωm
HP	500	15000	psi
OB	2000	40000	psi
DT_NCT	30	180	μs/ft
PPP	0	30000	psi

For data-driven baselines, we used CTGAN (Xu et al., 2019), trained per region on the shallow 70% of all real wells with depth conditioning and tuned via TPE (15 trials over its 5D space; 3 vs 23 trials/dim for POFM) across 4 seeds. In addition, two interpolation baselines were adopted. The first is SMOTE (Chawla et al., 2002), which interpolates between a record and its k nearest neighbours. The second is a Smoothed Bootstrap approach, selecting sample data while adding per-feature Gaussian variability. Both achieve high marginal fidelity but remain close to the training data. All four generators use a common interface and the same TSTR/TRTR pipeline. Search spaces, optimal settings, and further details appear in Appendices D, G, and A.

3.3 Performance Evaluation Metrics

We evaluated synthetic quality with marginal divergences (JSD, WD), multivariate distances (SWD, MMD with an RBF kernel, energy distance) and downstream TSTR/TRTR deep-interval R^2 over the seven regressors listed in Section 5.2.

4 Experimental Setup

The evaluation held out the deepest 30% of each well (by depth) as a blind test block; we reported R^2 on this block as deep-interval R^2. Training used the shallower interval for TRTR/semi-TSTR and synthetic logs for TSTR. Generators were fit only on the shallow 70% of each well. Hyperparameters were selected with 40 Optuna trials per setup, each re-fit over 20 fixed random seeds. Training scenarios are TRTR (5–80% of real data), pure TSTR (synthetic only) and semi-TSTR (synthetic data plus a contiguous real depth block covering 5–80% of the well, drawn from its shallow 70%; fractions exceeding the available pool are capped). Each TRTR run thus saw a single real block, whereas the generators were fit on the full shallow interval. All Optuna searches used synthetic-only validation splits to avoid real test-set leakage. The features were standardized and only rows with all values present were retained. The target variable is pore pressure (PPP), a direct downhole pressure measurement and independent of the log features. Also, the Eaton chain defines PPP only inside POFM's generator, never as the evaluation target. Depth was excluded from the downstream feature set for every model, so no model used depth as an input. Ridge, Random Forest, and KRR were accelerated with RAPIDS cuML (Raschka et al., 2020), while gradient-boosted trees and 1D-CNN used native GPU backends, all on a single NVIDIA A100. The four generators are available via an anonymized repository (link in supplementary materials). They share one interface and clip output to the physical limits (Table 2) and the outputs are replicated for a given `(region, seed, depth)` tuple.

5 Results

5.1 Distributional Fidelity

We compared all four generators to the same real-data references per zone using marginal (JSD, WD), multivariate, and novelty (DCR, separation from real training records) metrics. Table 3 summarizes per-zone aggregates for JSD, WD, SDMetrics (DataCebo, Inc., 2026) quality (mean of its column-shape and column-pair-trend scores), DCR, and calibration time.

Table 3: Per-zone distributional fidelity for all four generators. JSD and WD are mean ± std across the log variables of each zone (lower is better). Quality (SDMetrics column-shapes/trends) and DCR baseline protection are higher-is-better; Time is the TPE calibration wall-clock (single A100). Best generator per zone and metric in **bold**.

Zone	Generator	JSD↓	WD↓	Quality↑	DCR↑	Time (min)↓
Zone 1 Missa Keswal	POFM	0.095±0.092	110.11±146.69	0.857	**0.243**	47.1
	CTGAN	0.008±0.003	19.19±28.79	0.957	0.165	104.5
	SMOTE	**0.000 ± 0.000**	**2.12 ± 3.20**	**0.996**	0.003	15.2
	Smoothed Bootstrap	0.001±0.001	2.32±3.57	0.995	0.030	**10.0**
Zone 2 Joyamair–Minwal	POFM	0.315±0.292	346.32±742.21	0.576	**1.000**	98.6
	CTGAN	0.007±0.003	18.63±25.75	0.949	0.172	239.7
	SMOTE	**0.000 ± 0.000**	**2.87 ± 4.22**	**0.996**	0.006	33.6
	Smoothed Bootstrap	0.001±0.001	3.40±4.84	0.989	0.046	**21.7**
Zone 3 IODP U1343E	POFM	0.232±0.192	14.63±18.81	0.535	**1.000**	219.1
	CTGAN	0.006±0.002	0.21±0.28	0.975	0.565	461.1
	SMOTE	**0.003 ± 0.001**	**0.06 ± 0.06**	0.983	0.113	98.8
	Smoothed Bootstrap	0.003±0.002	0.07±0.08	**0.986**	0.114	**52.6**
Zone 4 Volve	POFM	0.087±0.065	16.55±16.04	0.698	**0.056**	215.6
	CTGAN	0.004±0.002	8.83±14.23	0.947	0.006	452.3
	SMOTE	0.000±0.000	**0.74 ± 1.23**	**0.994**	0.002	97.1
	Smoothed Bootstrap	**0.000 ± 0.000**	11.43±22.39	0.936	0.005	**51.4**

Interpolation generators (SMOTE, Smoothed Bootstrap) matched real marginals and calibrated fastest, but had the lowest DCR. POFM showed the highest DCR, but it had the weakest marginal performance, while CTGAN was intermediate. This ordering mirrored the downstream results of Section 5.2. Generators with the lowest DCR (closest to real data) transferred best under TSTR. The breakdown per-variable is given in Table 4, and the multivariate structure distances in Table 5.

Table 4: Per-variable Jensen–Shannon divergence to the real logs for each generator, with the real-to-real (R2R) within-zone baseline. Values below R2R indicate a generator whose marginals are closer to a training well than two real wells are to each other (a memorisation signature). Zone 3 (IODP, single well) has no R2R.

Zone	Variable	POFM	CTGAN	SMOTE	Smoothed Bootstrap	R2R
Zone 1 Missa Keswal	GR	0.295	0.005	0.000	0.001	0.348
	DT	0.041	0.009	0.001	0.001	0.286
	RHOB_combined	0.090	0.008	0.000	0.001	0.401
	RES_DEEP	0.031	0.005	0.001	0.000	0.171
	HP	0.063	0.013	0.000	0.001	0.304
	OB	0.044	0.007	0.000	0.001	0.315
	DT_NCT	0.165	0.006	0.000	0.002	0.457
	PPP	0.031	0.008	0.000	0.001	0.362
Zone 2 Joyamair–Minwal	GR	0.126	0.005	0.000	0.000	0.469
	DT	0.370	0.010	0.001	0.000	0.156
	RHOB_combined	0.519	0.011	0.001	0.000	0.154
	RES_DEEP	0.040	0.003	0.000	0.000	0.449
	HP	0.046	0.007	0.000	0.002	0.458
	OB	0.030	0.006	0.000	0.001	0.473
	DT_NCT	0.693	0.009	0.000	0.003	0.573
	PPP	0.693	0.010	0.000	0.002	0.471
Zone 3 IODP U1343E	GR	0.254	0.004	0.002	0.002	—
	DT	0.469	0.009	0.002	0.006	—
	RHOB_combined	0.204	0.004	0.004	0.002	—
	RES_DEEP	0.001	0.005	0.003	0.002	—
Zone 4 Volve	GR	0.120	0.004	0.000	0.000	0.392
	DT	0.078	0.004	0.000	0.000	0.365
	RHOB_combined	0.150	0.006	0.001	0.000	0.354
	RES_DEEP	0.001	0.001	0.000	0.000	0.033

Zone 3 (IODP, a well) has no R2R baseline. The R2R column is the JS divergence within a zone and serves as a reference for natural inter-well variability. Table 4 shows that CTGAN, SMOTE, and Bootstrap sat below R2R across variables. This indicated a lack of plausibility for a generator producing samples and was a repetitive signature. POFM was the generator whose JSD matched, and sometimes exceeded, R2R, aligning with its DCR and samples. The distances (Table 5) confirm the ordering beyond the marginals. SMOTE and Bootstrap were about an order of magnitude below CTGAN across zones, while POFM had the largest distances because its equation chain does not reproduce the structure across zones.

Table 5: Multivariate distances between synthetic and real samples (lower is better). Best generator per zone in **bold**. POFM's values are read from its calibration report; the others from SDMetrics.

Zone	Generator	SWD↓	MMD (RBF)↓	Energy↓
Zone 1 Missa Keswal	POFM	0.406	0.0867	0.427
	CTGAN	0.083	0.0035	0.018
	SMOTE	0.035	0.0004	0.003
	Smoothed Bootstrap	**0.034**	**0.0004**	**0.003**
Zone 2 Joyamair–Minwal	POFM	1.963	0.7379	7.622
	CTGAN	0.107	0.0057	0.029
	SMOTE	**0.032**	0.0003	**0.002**
	Smoothed Bootstrap	0.032	**0.0003**	0.002
Zone 3 IODP U1343E	POFM	10.745	0.4490	4.840
	CTGAN	0.159	0.0182	0.045
	SMOTE	0.056	0.0010	0.004
	Smoothed Bootstrap	**0.042**	**0.0005**	**0.003**
Zone 4 Volve	POFM	0.324	0.1118	0.249
	CTGAN	0.074	0.0029	0.010
	SMOTE	0.030	0.0004	0.001
	Smoothed Bootstrap	**0.027**	**0.0002**	**0.001**

5.2 Downstream Utility: TSTR and TRTR

We evaluated each generator using downstream regressors (Ridge, LightGBM (Ke et al., 2017), XGBoost (Chen & Guestrin, 2016), Random Forest (Breiman, 2001), CatBoost (Prokhorenkova et al., 2018), KRR (Saunders et al., 1998), CNN), tuned with Optuna trials on synthetic splits across seeds. The test block was excluded from training, and tuning used synthetic validation splits, preventing leakage into model selection. The pure-TSTR result and the TRTR baseline are in Tables 6 and 7.

Table 6: Pure-TSTR deep-interval R^2 (mean $\pm$ 95% CI across the six wells of Zones 1–2), by downstream model and generator. Higher is better; positive means the synthetic-trained model transfers to the real deep interval. Interpolation generators (SMOTE, Smoothed Bootstrap) transfer positively; the learned/physics generators (CTGAN, POFM) do not.

Model	POFM	CTGAN	SMOTE	Smoothed Bootstrap
Ridge	-2.63 ± 3.71	-1.62 ± 1.76	0.42 ± 0.08	0.35 ± 0.08
LightGBM	-3.37 ± 3.38	-3.17 ± 3.35	0.98 ± 0.02	0.89 ± 0.05
XGBoost	-3.36 ± 3.27	-3.21 ± 3.25	0.98 ± 0.02	0.90 ± 0.04
Random Forest	-3.09 ± 3.28	-2.93 ± 3.30	0.98 ± 0.01	0.89 ± 0.04
CatBoost	-3.18 ± 3.38	-3.25 ± 3.39	0.96 ± 0.02	0.89 ± 0.04
KRR	-1.72 ± 1.68	-1.64 ± 1.53	0.85 ± 0.05	0.84 ± 0.05
1D CNN	-2.92 ± 2.73	-2.02 ± 1.10	0.65 ± 0.05	0.62 ± 0.06

Table 7: TRTR baseline: deep-interval R^2 (mean $\pm$ 95% CI across six wells) when training on a fraction of *real* data only. TRTR is generator-independent; values are pooled across the four generator pipelines. R^2 is negative at every fraction, so the task is data-volume limited.

Model	5%	10%	20%	40%	60%	80%
Ridge	-115.2 ± 50.83	-28.12 ± 23.20	-97.30 ± 166.8	-8.15 ± 14.91	-2.14 ± 4.17	-2.05 ± 3.81
LightGBM	-2.97 ± 2.03	-2.76 ± 1.83	-2.27 ± 1.51	-1.22 ± 0.79	-1.41 ± 1.65	-1.63 ± 0.99
XGBoost	-2.95 ± 1.95	-2.69 ± 1.81	-2.18 ± 1.56	-1.28 ± 0.86	-1.44 ± 1.65	-1.57 ± 0.97
Random Forest	-3.02 ± 2.01	-2.76 ± 1.83	-2.17 ± 1.54	-1.57 ± 1.23	-1.50 ± 1.57	-1.70 ± 1.24
CatBoost	-2.87 ± 1.99	-2.49 ± 1.73	-2.12 ± 1.52	-1.13 ± 0.80	-1.24 ± 1.37	-1.15 ± 0.72
KRR	-81.30 ± 149.4	-21.88 ± 24.78	-6.59 ± 10.83	-8.29 ± 14.87	-2.14 ± 4.17	-2.74 ± 4.73
1D CNN	-160.4 ± 20.38	-29.43 ± 13.61	-11.02 ± 5.42	-4.31 ± 5.20	-1.74 ± 2.58	-1.80 ± 2.40

From the results, we conclude that there are two classes of generators. Interpolation generators (SMOTE, Bootstrap) transferred well, with tree ensembles reaching $R^2 \approx 0.98$. CTGAN and POFM were negative for all models, meaning they performed worse than predicting the mean. Adding

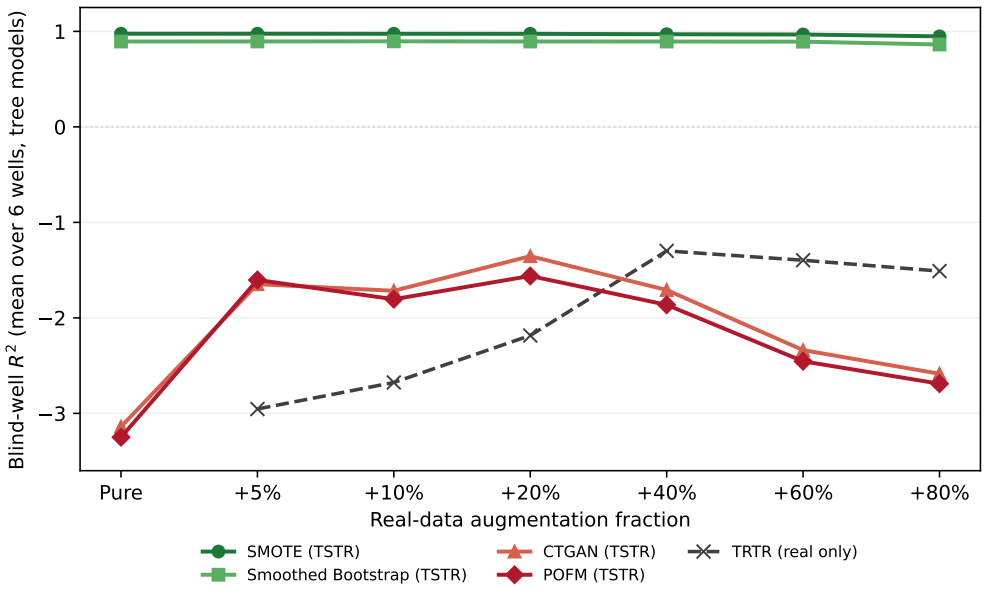

Figure 1: Deep-interval R^2 vs. real-data augmentation fraction (mean over the six Zone 1–2 wells, tree ensembles). SMOTE and Smoothed Bootstrap stay positive at every fraction; CTGAN, POFM, and the TRTR baseline stay negative—the generator, not the amount of real data, determines transfer.

data did not change this ordering. The 5–80% sweep per generator is in Appendix H, with per-well breakdowns for all four generators in Appendices I–P. To separate log-to-pressure transfer from POFM's Eaton relationship (PPP is derived from DT_NCT, OB, and HP), we re-ran TSTR with those Eaton-input features removed (Table 8). The interpolation generators dropped by half but stayed positive (tree ensembles $R^2 \approx 0.4$–0.5), so their transfer reflects a log signal rather than a recovered Eaton identity. CTGAN and POFM remained negative under feature sets, confirming their failure is distributional, not an artifact of the derived features.

Table 8: Ablation: pure-TSTR deep-interval R^2 (mean ± 95% CI across the six Zone 1–2 wells) with the Eaton-input features (DT_NCT, OB, HP) *removed*. Compare to Table 6 (all features): interpolation generators fall by about half yet stay positive, while CTGAN and POFM stay negative—so the interpolation transfer is not a recovered Eaton identity.

Model	POFM	CTGAN	SMOTE	Smoothed Bootstrap
Ridge	−1.64±1.42	−1.73±1.58	−0.48±0.65	−0.48±0.66
LightGBM	−1.56±1.35	−1.60±1.42	0.45±0.50	0.35±0.53
XGBoost	−1.55±1.40	−1.60±1.50	0.47±0.48	0.36±0.53
Random Forest	−1.46±1.32	−1.53±1.43	0.48±0.46	0.39±0.52
CatBoost	−1.34±1.35	−1.39±1.45	0.45±0.49	0.38±0.52
KRR	−1.68±1.36	−1.76±1.49	0.35±0.50	0.37±0.52
1D CNN	−4.68±2.61	−3.45±1.92	0.18±0.58	0.18±0.53

The TRTR baseline (training on real data only) was negative across all fractions and models, and interpolation-based augmentation exceeded it at every fraction. Figure 1 summarizes Tables 6–7 over the entire sweep. The compute footprint is in Appendix A.4.

6 Discussion

TRTR yielded negative R^2 at every real-data fraction (5–80%), so the generator ranking depends on the evaluation. Under data scarcity, the interpolation generators reproduced each well's local

feature pressure to cover the held-out deep interval, whereas the sparse real training block did not. This coverage explains why their TSTR reached $R^2 \approx 0.98$. POFM's high DCR reflected greater separation from training records, and its samples spanned regions the test well did not occupy and degraded predictions. CTGAN concentrated probability mass near the training wells, helping only when holdout wells mirror the training set. The visual comparison confirmed CTGAN output depth-independent bands while POFM reproduced depth-varying structure (Figure 7, Appendix Q). CTGAN's per-variable JSD fell below the real-to-real baseline, and the mean between models $\bar{R}^2$ stayed negative for both. CTGAN was negative in all 42 pure-TSTR runs (Appendix K). This dissociation is itself a finding. The three families encode different inductive biases: interpolation assumes the deep interval stays near the observed shallow manifold, a conditional generator would learn cross-well shallow-to-deep evolution, and POFM extrapolates through petrophysical laws. SMOTE is therefore likely to perform best when the deep interval remains close to the shallow one. Physical plausibility and novelty (high DCR) and downstream ML utility (TSTR) are distinct objectives. POFM maximized DCR while the interpolation generators maximized downstream ML utility. A generator that is more physically novel is, as such, not more useful downstream under data scarcity. Future work will study conditional continuation of partially observed wells, using depth-window sequence models and stochastic generators that yield uncertainty over deeper intervals.

Our evaluation has limitations. The blind test block is the deepest interval of each well, so training and test share the same wells. This mirrors look-ahead prediction while drilling, but does not test transfer to wells from other fields. POFM relies on a softmax-weighted aggregation of four fidelity criteria, which introduces additional hyperparameters and complicates comparisons between generators. The TPE budgets differ in absolute terms (500 vs 15 trials) but were scaled by search-space dimensionality ($\approx$23 vs 3 trials/dim). POFM assumes single-fluid Archie behavior and a single-lithology compaction model, limiting transferability beyond the benchmarked regimes. CTGAN models the regional joint distribution rather than $P(\text{deep logs} \mid \text{shallow logs of the same well})$. Downstream models also treat depth samples as independent rows, so the 1D CNN convolves over features, not depth. Finally, this study does not evaluate intermediate targets (e.g. DT, RHOB) or per-well PPP recalibration. These are extensions that may help isolate synthetic-data quality from labeling-protocol variability.

7 Conclusion

We presented WellBench, a multi-basin benchmark for synthetic well-log generation. The interpolation generators led in the distributional metrics and had a positive TSTR R^2. CTGAN and POFM had negative TSTR R^2. POFM led in DCR and exposes interpretable parameters, while CTGAN's marginal scores reflected memorization. TRTR was negative at every real-data fraction, whereas the choice of generator determines the sign of TSTR R^2. Promising directions are conditional shallow-to-deep continuation and hybrid interpolation methods constrained by petrophysical equations.

Reproducibility Statement

We provide full implementation details in the appendices: the POFM algorithm and petrophysical equations (Section 3.2, Algorithm 1), the hyperparameter search spaces and per-region optimised values for POFM and CTGAN (Appendices D, E), and the physical bounds and dataset details (Appendices A, B). All four generators, the calibrated per-region parameters, and the fixed random seeds are released in an anonymized repository: `https://anonymous.4open.science/r/wellbench-5A68/README.md`

Ethics Statement

This work complies with the ICLR Code of Ethics. The models were trained on a non-public well-log dataset containing no human subjects or sensitive personal data. The data used in this study can be made available by the authors upon request.

REFERENCES

Halliburton Chart Book Sep | PDF | Flow Measurement | Electrical Resistivity And Conductivity, a. URL https://www.scribd.com/document/376115477/66027-Halliburton-Chart-Book-Sep.

KGS–Geological Log Analysis–The Gamma Ray Log, b. URL https://www.kgs.ku.edu/Publications/Bulletins/LA/03_gamma.html.

(PDF) Electrical Resistivity Structure of the Altiplano-Puna Magma Body and Volcan Uturuncu from Magnetotelluric Data, c. URL https://www.researchgate.net/publication/312210761_Electrical_Resistivity_Structure_of_the_Altiplano-Puna_Magma_Body_and_Volcan_Uturuncu_from_Magnetotelluric_Data.

Well Log Suites - SEPM Strata, d. URL https://sepmstrata.org/page.aspx?pageid=167.

The Why and Basic Principles of Managed Well-Bore Pressure. In *Managed Pressure Drilling*, pp. 1–38. Gulf Publishing Company, January 2008. doi: 10.1016/B978-1-933762-24-1.50007-3. URL https://www.sciencedirect.com/science/chapter/edited-volume/pii/B9781933762241500073.

Environmentally assisted cracking (EAC) in oil and gas production. In *Stress Corrosion Cracking*, pp. 570–607. Woodhead Publishing, January 2011. ISBN 978-0-85709-376-9. doi: 10.1533/9780857093769.4.570. URL https://www.sciencedirect.com/science/chapter/edited-volume/abs/pii/B9781845696733500158.

Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A Next-generation Hyperparameter Optimization Framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD '19, pp. 2623–2631, New York, NY, USA, July 2019. Association for Computing Machinery. ISBN 978-1-4503-6201-6. doi: 10.1145/3292500.3330701. URL https://dl.acm.org/doi/10.1145/3292500.3330701.

Muhammad Raiees Amjad, Rohan Benjamin Varghese, and Tehmina Amjad. Machine Learning Models for Subsurface Pressure Prediction: A Data Mining Approach. *Computers*, 14(11):499, November 2025. ISSN 2073-431X. doi: 10.3390/computers14110499. URL https://www.mdpi.com/2073-431X/14/11/499.

Klapa Antonion, Xiao Wang, Maziar Raissi, and Laurn Joshie. Machine Learning Through Physics–Informed Neural Networks: Progress and Challenges. *Academic Journal of Science and Technology*, 9(1):46–49, January 2024. ISSN 2771-3032. doi: 10.54097/b1d21816. URL https://drpress.org/ojs/index.php/ajst/article/view/16292.

George Asquith and Daniel Krygowski. *Basic Well Log Analysis*. Number 16 in AAPG Methods in Exploration. American Association of Petroleum Geologists, Tulsa, OK, 2 edition, 2004. doi: 10.1306/Mth16823.

James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for Hyper-Parameter Optimization. In *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011. URL https://papers.nips.cc/paper_files/paper/2011/hash/86e8f7ab32cfd12577bc2619bc635690-Abstract.html.

Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

Tianqi Chen and Carlos Guestrin. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, New York, NY, USA, 2016. Association for Computing Machinery. doi: 10.1145/2939672.2939785.

DataCebo, Inc. SDMetrics: Metrics for synthetic data generation projects. https://github.com/sdv-dev/SDMetrics, 2026. Accessed: 2026-09-24.

Ben A. Eaton. The Equation for Geopressure Prediction from Well Logs. OnePetro, September 1975. doi: 10.2118/5544-MS. URL https://dx.doi.org/10.2118/5544-MS.

Aline Cambri Fredere, Gabriel De Oliveira Ramos, Luciano Garim Garcia, Mateus da Rocha Simionato, José Manuel Marques Teixeira de Oliveira, and Ariane Santos da Silveira. Lithology-Aware Conditional Variational Autoencoder for Synthetic Well Log Generation in Petroleum Reservoirs (Student Abstract). *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(48):41209–41211, March 2026. ISSN 2374-3468. doi: 10.1609/aaai.v40i48.42216. URL https://ojs.aaai.org/index.php/AAAI/article/view/42216.

Guozhong Gao, Omid Hazbeh, Meysam Rajabi, Somayeh Tabasi, Hamzeh Ghorbani, Reza Seyedkamali, Milad Shayanmanesh, Ahmed E. Radwan, and Amir H. Mosavi. Application of GMDH model to predict pore pressure. *Frontiers in Earth Science*, 10, January 2023. ISSN 2296-6463. doi: 10.3389/feart.2022.1043719. URL https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2022.1043719/full.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html.

Wancheng Huang and Yuan Tian. Reservoir parameters prediction based on spatially transferred long short-term memory network. *PLOS ONE*, 19(1):e0296506, January 2024. ISSN 1932-6203. doi: 10.1371/journal.pone.0296506. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0296506.

Ali Samer Hussien and Abdalratha Mohammed Sahaab. Well Logs Analysis and Petrophysical Properties of Shiranish Formation in the East Baghdad Oil Field, Central Iraq. *IOP Conference Series: Earth and Environmental Science*, 1300(1):012035, February 2024. ISSN 1755-1315. doi: 10.1088/1755-1315/1300/1/012035. URL https://doi.org/10.1088/1755-1315/1300/1/012035.

Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017.

F. Khoshnaw, P. Jaf, and S. Farkha. Pore, Abnormal Formation and Fracture Pressure Prediction. *WIT Transactions on Ecology and the Environment*, 186, 2015. ISSN 1743-3541. URL https://trid.trb.org/View/1377421. ISBN: 9781845648381.

Anton Danholt Lautrup, Tobias Hyrup, Arthur Zimek, and Peter Schneider-Kamp. Systematic Review of Generative Modelling Tools and Utility Metrics for Fully Synthetic Tabular Data. *ACM Comput. Surv.*, 57(4):90:1–90:38, December 2024. ISSN 0360-0300. doi: 10.1145/3704437. URL https://dl.acm.org/doi/10.1145/3704437.

Weiji Liu, Pengjie Wang, and Xiaohua Zhu. Brittle-ductile transition and fragmentation mechanism of rocks under ultra-deep conditions. *Geoenergy Science and Engineering*, 257:214272, February 2026. ISSN 2949-8910. doi: 10.1016/j.geoen.2025.214272. URL https://www.sciencedirect.com/science/article/pii/S294989102500630X.

Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

Sebastian Raschka, Joshua Patterson, and Corey Nolet. Machine learning in Python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, 11(4):193, 2020. doi: 10.3390/info11040193.

Gabriele Santangelo, Giovanna Nicora, Riccardo Bellazzi, and Arianna Dagliati. How good is your synthetic data? SynthRO, a dashboard to evaluate and benchmark synthetic tabular data. *BMC Medical Informatics and Decision Making*, 25(1):89, February 2025. ISSN 1472-6947. doi: 10.1186/s12911-024-02731-9. URL https://doi.org/10.1186/s12911-024-02731-9.

Craig Saunders, Alexander Gammerman, and Volodya Vovk. Ridge regression learning algorithm in dual variables. In *Proceedings of the 15th International Conference on Machine Learning*, pp. 515–521. Morgan Kaufmann, 1998.

Pradeep Kumar Shukla, David Lall, and Vikram Vishal. A Robust Mechanistic Model for Pore Pressure Prediction from Petrophysical Logs Aided by Machine Learning in the Gas Hydrate-Bearing Sediments over the Offshore Krishna–Godavari Basin, India. *Natural Resources Research*, 32(6):2727–2752, December 2023. ISSN 1573-8981. doi: 10.1007/s11053-023-10262-9. URL https://doi.org/10.1007/s11053-023-10262-9.

Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Modeling Tabular data using Conditional GAN. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/254ed7d2de3b23ab10936522dd547b78-Abstract.html.

Jincai Zhang. Effective stress, porosity, velocity and abnormal pore pressure prediction accounting for compaction disequilibrium and unloading. *Marine and Petroleum Geology*, 45:2–11, August 2013. ISSN 0264-8172. doi: 10.1016/j.marpetgeo.2013.04.007. URL https://www.sciencedirect.com/science/article/pii/S0264817213000913.

Yushuai Zhang, Dawei Lv, Yong Wang, Haiyan Liu, Guangzeng Song, and Jie Gao. Geological characteristics and abnormal pore pressure prediction in shale oil formations of the Dongying depression, China. *Energy Science & Engineering*, 8(6):1962–1979, 2020. ISSN 2050-0505. doi: 10.1002/ese3.641. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/ese3.641. _eprint: https://scijournals.onlinelibrary.wiley.com/doi/pdf/10.1002/ese3.641.

Tianjie Zhao, Sheng Wang, Chaojun Ouyang, Min Chen, Chenying Liu, Jin Zhang, Long Yu, Fei Wang, Yong Xie, Jun Li, Fang Wang, Sabine Grunwald, Bryan M. Wong, Fan Zhang, Zhen Qian, Yongjun Xu, Chengqing Yu, Wei Han, Tao Sun, Zezhi Shao, Tangwen Qian, Zhao Chen, Jiangyuan Zeng, Huai Zhang, Husi Letu, Bing Zhang, Li Wang, Lei Luo, Chong Shi, Hongjun Su, Hongsheng Zhang, Shuai Yin, Ni Huang, Wei Zhao, Nan Li, Chaolei Zheng, Yang Zhou, Changping Huang, Defeng Feng, Qingsong Xu, Yan Wu, Danfeng Hong, Zhenyu Wang, Yinyi Lin, Tangtang Zhang, Prashant Kumar, Antonio Plaza, Jocelyn Chanussot, Jiabao Zhang, Jiancheng Shi, and Lizhe Wang. Artificial intelligence for geoscience: Progress, challenges, and perspectives. *The Innovation*, 5(5), September 2024. ISSN 2666-6758. doi: 10.1016/j.xinn.2024.100691. URL https://www.cell.com/the-innovation/abstract/S2666-6758(24)00129-2.

A Extended Method Details

This appendix provides granular details on dataset sources, physical depth constraints, and baseline model configurations referenced in the main text.

A.1 Dataset Overview

We evaluated our methodology in four regions (Section 3.1), detailed in Table 9.

Table 9: Summary of well log datasets and available curve features.

Region / Field	Geologic Setting	Wells	Available Curves
Zones 1 & 2	Eastern Potwar Basin	6	GR, DT, RHOB, RES_DEEP, HP, OB, DT_NCT, PPP
Expedition 323 (U1343E)	Bering Sea Continental Slope	1	GR, DT, RHOB, RES_DEEP
Volve Field	Norwegian North Sea	5	GR, DT, RHOB, RES_DEEP

A.2 Depth Range Justification

Although the Earth's crust extends to approximately 165,000 feet, hydrocarbon instances are concentrated in upper intervals:

- **10,000 to 21,000 ft:** main zone for the majority of commercial hydrocarbon discoveries (noa, 2011).
- **21,000 to 33,000 ft:** Infrequent occurrences.
- **Over 33,000 ft:** Rare occurrences.

Below the brittle–ductile transition, rock behavior departs from POFM's assumptions, leading to model degradation (Liu et al., 2026). Therefore, calibration depth ranges are limited to intervals above 4,500 ft (Zones 1–2) or 14,000 ft (Zone 4).

A.3 Baseline Model Specifications

CTGAN models numeric values as multimodal distributions using conditional vectors to preserve categories. Models are trained per region on real wells, using depth as a conditioning feature normalized with per-well min–max scaling to capture depth-dependent trends. Trial budgets for CTGAN (15) and POFM (500) also scale with search-space dimensionality to avoid disadvantaging CTGAN. In contrast, held-out validation wells are excluded from the sampling pool for both interpolation baselines. SMOTE creates synthetic records by interpolating between an instance and its k-nearest neighbors, sampling along the observed data hull (Chawla et al., 2002). Smoothed Bootstrap samples records with replacement and adds per-feature Gaussian noise, yielding resamples around each point.

A.4 Compute Footprint

Table 10 summarizes the study's compute footprint. All models use under 5 GB of host RAM; only the 1D CNN materially uses the GPU ($\approx$191 MB vs. $\approx$16 MB for the classical models).

Table 10: Downstream compute footprint: mean peak host RAM (GB) and peak GPU memory (MB) per model, averaged over all generators, wells, and augmentation fractions. Only the 1D CNN makes material use of the GPU.

Model	**Peak host RAM (GB)**	**Peak GPU (MB)**
Ridge	3.67	15.8
LightGBM	3.80	15.7
XGBoost	3.68	15.6
Random Forest	3.76	15.6
CatBoost	3.83	15.5
KRR	4.48	15.8
1D CNN	3.71	190.7

B WELL LOGGING PARAMETER RANGES

Table 11: Definition and parameter bounds are set to physical ranges.

Mnemonic	Range	Description
GR	[0; 200]	Gamma Ray logging involves measuring natural radiation. It's a physical impossibility to have negative radiation, establishing the minimal value at 0 API. It's common practice to set the max values at 150 API, as Uranium-rich sands tend to hover around these values. This extension to 200 API is made to accommodate organic-rich soils. (noa, b; Asquith & Krygowski, 2004; noa, d)
DT	[30; 180]	Sonic transit time logging is the measurement of how fast sonic waves go through the material. Dolomite (43,5 us/ft) tends to be the lower value as it is the fastest common rock, however 30 is used to accommodate uncommon materials, typically mafic rocks, to prevent overlooking anomalies. As for the max value, 180 us/t tends to be used as it is the value typically registered on water/mud. (noa, d; Hussien & Sahaab, 2024)
RHOB	[1.2; 2.9]	Bulk density measures the overall density of a rock. A lower value of 1.2 was chosen as it is the values of the coal-rich lithologies, as any value lower that than tends to need a different log with different tools. The upper value of 2.9 corresponds to Anhydrite the common upper most limit of sedimentary logs. (noa, a;d)
RES_DEEP	[0.01; 10000]	Deep resistivity measures how easily electricity flows through the formation. A value of 0 would mean infinity conductivity, so the chosen lower value represents a practical maximum, commonly found in massive sulfides. The upper most value of 10000 however represent the values found in very tight lithologies like quartzite (noa, c;d).
HP	[500; 15000]	Hydrostatic Pressure logging is the measurement of how much pressure a column of water inflicts on the borehole. 500 was chosen as the lower value as it is a common value for safety drilling in hydrocarbon boreholes. 15000 was chosen as the higher value as it a value that tends to be measured in ultra deep drilling operations (Zhang, 2013; noa, 2008; Shukla et al., 2023; Zhang et al., 2020).
OB	[2000; 40000]	Overburden Pressure logging is closely tied to HP as it measures the pressure of the rock and fluid column. The lower value was chosen as assuming a standard value of 1 psi/ft at 2000 feet. On the other hand the upper most value of 40000 was chosen as a safety net as an above average pressure until 32000 feet (Zhang, 2013; noa, 2008; Shukla et al., 2023; Zhang et al., 2020).
DT_NCT	[30; 180]	Normal Compaction Trend is the theoretical value DT on a normal pressure situation. Typically the lower value found in Dolomites is 40, however 30 was chosen as a safety net to compensate the earlier safety net in DT. The upper most value of 180 corresponds to the maximum DT value (Zhang, 2013; noa, 2008; Shukla et al., 2023; Zhang et al., 2020).
PPP	[0; 30000]	Pore pressure prediction is the calculation of the fluid pressure in a formation using the relationship between DT and DT_NCT. 0 was chosen as it is mathematically impossible to have negative values. The upper most limit was chosen 30000 as PPP tends to be half of the OB in normal conditions, the added 10000 is a safety net for any anomalies (Zhang, 2013; noa, 2008; Shukla et al., 2023; Zhang et all,2020).

Most hydrocarbon instances are found at depths between 10,000-21,000 ft. Fewer instances are found beyond 21,000 ft, and single-digit cases are found beyond 33,000 ft (noa, 2011). The brittle-ductile transition (typically 26,000–46,000 ft) limits the behavior of the rocks, which deviates from the assumptions of our POFM, and reliability degrades accordingly (Liu et al., 2026), meaning calibration is restricted to shallower depths (in Section 3.1 to depths above 4,500 ft (Z1–Z2) or 14,000 ft (Z4)). For Zone 3 (Bering Sea) and Zone 4 (Volve), pore pressure measurements are not available. As such, the study applies a reduced POFM (Algorithm 2) that generates only the five primary logs of the rock properties.

C REDUCED POFM

This algorithm is a simplified version of POFM as presented in Section 3.2 (Algorithm 1), covering steps 1 through 5.

Algorithm 2 POFM with no pore pressure

Require: Depth vector $\mathbf{z}$, hyperparameters $\boldsymbol{\theta}$, random seed s
Ensure: Petrophysical Log DataFrame $\mathbf{D}$
1: $\phi \leftarrow \texttt{AthyPorosity}(\mathbf{z}, \boldsymbol{\theta}, s)$
2: DT $\leftarrow \texttt{RHG}(\phi, \boldsymbol{\theta}, s + 10)$
3: RHOB $\leftarrow \texttt{DensityMix}(\phi, \mathbf{z}, \boldsymbol{\theta}, s + 20)$
4: RES $\leftarrow \texttt{Archie}(\phi, \boldsymbol{\theta}, s + 30)$
5: GR $\leftarrow \texttt{ShaleGR}(\phi, \boldsymbol{\theta}, s + 40)$
6: **return** $\mathbf{D} = \{\mathbf{z}, \phi, \text{GR}, \text{DT}, \text{RHOB}, \text{RES}\}$

D HYPERPARAMETER SEARCH SPACE OF POFM

This section describes the physics-parameter search space and the equations used.

Table 12: Hyperparameter search space with physically motivated bounds. Parameters are grouped by the petrophysical law they govern.

Law	Parameter	Min	Max	Scale
Athy's Compaction	ϕ_0 (surface porosity)	0.15	0.75	linear
	c (compaction coeff., 1/ft)	10^{-6}	10^{-2}	log
	σ_ϕ (porosity noise)	0.001	0.15	linear
	A_ℓ (layering amplitude)	0.001	0.25	linear
	λ_ℓ (layering period, ft)	2.0	300	linear
RHG Sonic	$\mathrm{DT}_{\mathrm{ma}}$ (μs/ft)	38	95	linear
	DT_f (μs/ft)	140	300	linear
	σ_{DT} (μs/ft)	0.1	30	linear
Density Mixing	ρ_{ma} (g/cc)	2.40	2.95	linear
	ρ_f (g/cc)	0.70	1.30	linear
	σ_ρ (g/cc)	0.001	0.20	linear
Archie's Law	a (tortuosity)	0.3	3.5	linear
	m (cementation exp.)	1.0	4.5	linear
	n (saturation exp.)	1.0	4.5	linear
	R_w (Ωm)	0.001	2.0	log
	S_w (fraction)	0.15	1.0	linear
	σ_R	0.001	0.50	linear
GR Model	$\mathrm{GR}_{\mathrm{clean}}$ (API)	0	80	linear
	$\mathrm{GR}_{\mathrm{shale}}$ (API)	50	300	linear
	σ_{GR} (API)	0.5	35	linear
Eaton's PPP	n (Eaton exponent)	0.5	10.0	linear

E Optimised Hyperparameters per Region

Tables 13–16 report the full set of optimised hyperparameters obtained via TPE calibration for each of the four geological regions. These values correspond to the best trial identified during the Optuna search (Section 3.2). They are used for all final evaluations reported in the main text.

Table 13: Optimised hyperparameters for Zone 1: Missa Keswal.

Law / Group	**Parameter**	**Value**
Athy's Compaction	ϕ_0 (surface porosity)	0.5454
	c (compaction coeff., 1/ft)	4.914×10^{-4}
	σ_ϕ (porosity noise)	0.0085
	A_ℓ (layering amplitude)	0.1946
	λ_ℓ (layering period, ft)	7.55
RHG Sonic	$\mathrm{DT}_{\mathrm{ma}}$ (μs/ft)	67.17
	DT_f (μs/ft)	152.27
	σ_{DT} (μs/ft)	7.00
Density Mixing	ρ_{ma} (g/cc)	2.854
	ρ_f (g/cc)	1.291
	σ_ρ (g/cc)	0.131
Archie's Law	a (tortuosity)	1.490
	m (cementation exp.)	2.335
	n (saturation exp.)	1.596
	R_w (Ωm)	0.103
	S_w (fraction)	0.514
	σ_R (log-scale)	0.137
GR Model	$\mathrm{GR}_{\mathrm{clean}}$ (API)	41.97
	$\mathrm{GR}_{\mathrm{shale}}$ (API)	220.03
	σ_{GR} (API)	23.38
Pressure[†]	$\rho_{f,\mathrm{HP}}$ (g/cc)	0.917
	$\bar{\rho}_{\mathrm{shallow}}$ (g/cc)	1.854
	DT_0 (NCT intercept, μs/ft)	40.89
	DT_s (NCT surface, μs/ft)	297.62
	k (NCT decay, 1/ft)	1.966×10^{-4}
	NCT smooth window (samples)	932
	w (NCT blend weight)	0.769
	n_{Eaton} (Eaton exponent)	0.981

Table 14: Optimised hyperparameters for Zone 2: Joyamair–Minwal.

Law / Group	Parameter	Value
Athy's Compaction	ϕ_0 (surface porosity)	0.3898
	c (compaction coeff., 1/ft)	3.375×10^{-4}
	σ_ϕ (porosity noise)	0.0770
	A_ℓ (layering amplitude)	0.2408
	λ_ℓ (layering period, ft)	13.45
RHG Sonic	$\mathrm{DT}_{\mathrm{ma}}$ (μs/ft)	41.57
	DT_f (μs/ft)	182.42
	σ_{DT} (μs/ft)	8.34
Density Mixing	ρ_{ma} (g/cc)	2.921
	ρ_f (g/cc)	0.903
	σ_ρ (g/cc)	0.183
Archie's Law	a (tortuosity)	3.123
	m (cementation exp.)	1.714
	n (saturation exp.)	3.750
	R_w (Ωm)	0.335
	S_w (fraction)	0.963
	σ_R (log-scale)	0.182
GR Model	$\mathrm{GR}_{\mathrm{clean}}$ (API)	30.01
	$\mathrm{GR}_{\mathrm{shale}}$ (API)	98.81
	σ_{GR} (API)	15.89
Pressure†	$\rho_{f,\mathrm{HP}}$ (g/cc)	1.050
	$\bar{\rho}_{\mathrm{shallow}}$ (g/cc)	1.642
	DT_0 (NCT intercept, μs/ft)	59.95
	DT_s (NCT surface, μs/ft)	307.86
	k (NCT decay, 1/ft)	4.823×10^{-4}
	NCT smooth window (samples)	1494
	w (NCT blend weight)	0.921
	n_{Eaton} (Eaton exponent)	1.368

Table 15: Optimised hyperparameters for Zone 3: IODP U1343E.

Law / Group	Parameter	Value
Athy's Compaction	ϕ_0 (surface porosity)	0.3710
	c (compaction coeff., 1/m)	4.280×10^{-5}
	σ_ϕ (porosity noise)	0.0669
	A_ℓ (layering amplitude)	0.1475
	λ_ℓ (layering period, m)	22.59
RHG Sonic	$\mathrm{DT}_{\mathrm{ma}}$ (μs/ft)	64.53
	DT_f (μs/ft)	214.12
	σ_{DT} (μs/ft)	12.34
Density Mixing	ρ_{ma} (g/cc)	2.411
	ρ_f (g/cc)	0.890
	σ_ρ (g/cc)	0.173
Archie's Law	a (tortuosity)	1.003
	m (cementation exp.)	3.494
	n (saturation exp.)	2.480
	R_w (Ωm)	8.289×10^{-3}
	S_w (fraction)	0.911
	σ_R (log-scale)	0.499
GR Model	$\mathrm{GR}_{\mathrm{clean}}$ (API)	55.24
	$\mathrm{GR}_{\mathrm{shale}}$ (API)	99.73
	σ_{GR} (API)	19.07

Table 16: Optimised hyperparameters for Zone 4: Volve.

Law / Group	Parameter	Value
Athy's Compaction	ϕ_0 (surface porosity)	0.7271
	c (compaction coeff., 1/m)	5.805×10^{-4}
	σ_ϕ (porosity noise)	0.0764
	A_ℓ (layering amplitude)	0.1599
	λ_ℓ (layering period, m)	51.65
RHG Sonic	$\mathrm{DT}_{\mathrm{ma}}$ (μs/ft)	67.53
	DT_f (μs/ft)	189.58
	σ_{DT} (μs/ft)	17.24
Density Mixing	ρ_{ma} (g/cc)	2.553
	ρ_f (g/cc)	0.923
	σ_ρ (g/cc)	0.187
Archie's Law	a (tortuosity)	2.358
	m (cementation exp.)	1.348
	n (saturation exp.)	1.961
	R_w (Ωm)	0.014
	S_w (fraction)	0.857
	σ_R (log-scale)	0.498
GR Model	$\mathrm{GR}_{\mathrm{clean}}$ (API)	78.20
	$\mathrm{GR}_{\mathrm{shale}}$ (API)	50.34
	σ_{GR} (API)	33.46

F HYPERPARAMETER SEARCH SPACE OF CTGAN

In this section, we describe the search space of the CTGAN baseline hyperparameters.

Table 17: CTGAN hyperparameter search space.

Group	**Parameter**	**Min**	**Max**	**Scale**
Architecture	Embedding dim	32	256	log
	Generator hidden dims	64	512	log
Optimisation	Learning rate	10^{-5}	10^{-3}	log
	Batch size	100	1000	log
Training	Epochs	100	1000	log

G Optimised Hyperparameters of CTGAN

In this section, we present the optimal CTGAN hyperparameters identified using the tree-structured Parzen estimator.

Table 18: Best CTGAN hyperparameters identified by TPE per region

Parameter	**Zone 1**	**Zone 2**	**Zone 3**	**Zone 4**
Embedding dim	256	256	256	256
Generator hidden dims	128,128	128,128	128,128	128,128
Learning rate	3.07×10^{-4}	1.29×10^{-4}	1.29×10^{-4}	1.29×10^{-4}
Batch size	1000	500	500	500
Epochs	500	500	500	500
Best JSD (Optuna)	0.0765	0.0769	0.0920	0.0369
Wall time (min)	106.5	176.0	17.5	492.9

H TSTR BY AUGMENTATION FRACTION (PER GENERATOR)

Tables 19–22 report the full augmentation sweep for each generator: deep-interval R^2 (mean $\pm$ 95% CI across the six wells of Zones 1–2) from pure synthetic training (Pure) to $+80\%$ real data. SMOTE and Smoothed Bootstrap remain positive and roughly flat across fractions; CTGAN and POFM remain negative throughout, confirming that mixing in real data does not rescue the learned/physics generators.

Table 19: TSTR deep-interval R^2 (mean $\pm$ 95% CI across six wells) for SMOTE across augmentation fractions. Pure = 100% synthetic training data; $+x\%$ adds that fraction of real data.

Model	Pure	+5%	+10%	+20%	+40%	+60%	+80%
Ridge	0.42±0.08	0.38±0.09	0.37±0.09	0.37±0.09	0.36±0.08	0.31±0.10	0.34±0.12
LightGBM	0.98±0.02	0.98±0.02	0.98±0.02	0.98±0.02	0.98±0.01	0.98±0.01	0.96±0.04
XGBoost	0.98±0.02	0.98±0.01	0.98±0.02	0.98±0.01	0.97±0.02	0.97±0.02	0.96±0.03
Random Forest	0.98±0.01	0.98±0.01	0.98±0.01	0.98±0.01	0.98±0.01	0.98±0.01	0.98±0.01
CatBoost	0.96±0.02	0.96±0.02	0.96±0.02	0.96±0.02	0.95±0.02	0.94±0.03	0.89±0.07
KRR	0.85±0.05	0.86±0.05	0.86±0.05	0.86±0.05	0.86±0.05	0.86±0.05	0.61±0.28
1D CNN	0.65±0.05	0.62±0.05	0.57±0.09	0.56±0.10	0.53±0.09	0.49±0.11	0.06±0.27

Table 20: TSTR deep-interval R^2 (mean $\pm$ 95% CI across six wells) for Smoothed Bootstrap across augmentation fractions. Pure = 100% synthetic training data; $+x\%$ adds that fraction of real data.

Model	Pure	+5%	+10%	+20%	+40%	+60%	+80%
Ridge	0.35±0.08	0.34±0.09	0.34±0.09	0.33±0.08	0.31±0.08	0.29±0.09	0.29±0.10
LightGBM	0.89±0.05	0.89±0.04	0.90±0.04	0.89±0.04	0.89±0.04	0.89±0.04	0.86±0.06
XGBoost	0.90±0.04	0.90±0.04	0.90±0.04	0.90±0.04	0.90±0.04	0.90±0.04	0.86±0.08
Random Forest	0.89±0.04	0.89±0.04	0.89±0.04	0.89±0.04	0.89±0.04	0.89±0.05	0.88±0.06
CatBoost	0.89±0.04	0.89±0.04	0.89±0.04	0.89±0.04	0.89±0.04	0.89±0.04	0.85±0.07
KRR	0.84±0.05	0.84±0.05	0.85±0.05	0.85±0.05	0.85±0.05	0.80±0.11	0.53±0.39
1D CNN	0.62±0.06	0.58±0.07	0.54±0.11	0.57±0.06	0.51±0.06	0.50±0.11	0.22±0.19

Table 21: TSTR deep-interval R^2 (mean $\pm$ 95% CI across six wells) for CTGAN across augmentation fractions. Pure = 100% synthetic training data; $+x\%$ adds that fraction of real data.

Model	Pure	+5%	+10%	+20%	+40%	+60%	+80%
Ridge	−1.62±1.76	−2.55±3.04	−2.07±2.67	−1.42±1.83	−0.74±0.83	−2.26±3.93	−1.62±2.71
LightGBM	−3.17±3.35	−1.33±0.81	−2.11±2.36	−1.05±0.65	−1.86±2.13	−2.45±3.52	−2.67±3.04
XGBoost	−3.21±3.25	−1.87±1.74	−1.78±1.53	−1.79±1.49	−2.06±2.63	−2.44±3.41	−2.72±3.07
Random Forest	−2.93±3.30	−1.62±1.47	−1.36±0.76	−1.17±0.77	−1.98±2.39	−2.58±3.40	−2.90±2.85
CatBoost	−3.25±3.39	−1.77±1.50	−1.61±1.74	−1.40±1.24	−0.93±0.59	−1.88±2.38	−2.05±2.11
KRR	−1.64±1.53	−1.18±1.37	−1.00±0.86	−0.43±0.32	−0.68±0.75	−2.14±3.96	−7.42±10.79
1D CNN	−2.02±1.10	−1.54±1.42	−1.74±1.06	−1.79±2.00	−1.18±0.91	−1.16±1.50	−1.14±1.50

Table 22: TSTR deep-interval R^2 (mean $\pm$ 95% CI across six wells) for POFM across augmentation fractions. Pure = 100% synthetic training data; $+x\%$ adds that fraction of real data.

Model	Pure	+5%	+10%	+20%	+40%	+60%	+80%
Ridge	−2.63±3.71	−2.62±3.02	−2.12±2.65	−1.47±1.81	−0.76±0.82	−2.28±3.92	−1.65±2.70
LightGBM	−3.37±3.38	−1.46±0.89	−2.41±2.73	−1.71±1.64	−2.16±2.38	−2.42±3.09	−2.78±3.03
XGBoost	−3.36±3.27	−1.83±1.74	−1.75±1.52	−1.61±1.37	−2.10±2.63	−2.49±3.40	−2.84±3.02
Random Forest	−3.09±3.28	−1.35±0.99	−1.47±0.85	−1.62±1.31	−2.10±2.40	−2.70±3.38	−3.02±2.83
CatBoost	−3.18±3.38	−1.78±1.51	−1.59±1.73	−1.32±1.17	−1.08±0.77	−2.20±2.87	−2.12±2.09
KRR	−1.72±1.68	−1.24±1.37	−1.06±0.86	−0.54±0.49	−0.64±0.75	−2.54±3.83	−7.41±10.80
1D CNN	−2.92±2.73	−2.01±1.24	−1.61±1.20	−1.75±1.46	−1.33±0.91	−1.45±2.21	−1.53±1.92

I PER-WELL TSTR RESULTS (POFM)

Table 23 provides the full per-well breakdown of the physics-informed TSTR results summarised in Table 22. Each cell reports the mean test-well R^2 over 20 random seeds, the variance σ^2 across seeds, and the wall-clock time (in seconds) for the Optuna search plus the final 20-seed evaluation.

Table 23: Per-well TSTR results using POFM generator. σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		Pure TSTR			**+5%**			**+10%**			**+20%**		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	−1.46	0.001	0.4	−10.21	113.1	0.4	−8.74	100.4	0.4	−5.93	60.575	0.4
	LightGBM	−11.65	0.018	161.8	−3.18	0.438	92.8	−9.18	6.184	94.1	−5.48	1.277	115.7
	XGBoost	−11.24	0.008	115.6	−6.14	3.197	51.7	−5.45	5.120	57.4	−4.76	4.875	48.3
	Random Forest	−11.29	0.005	289.2	−3.18	1.404	88.0	−3.04	1.927	146.5	−4.32	3.012	179.0
	CatBoost	−11.56	0.023	262.0	−5.22	0.226	452.9	−5.61	2.655	884.4	−3.65	0.530	495.2
	KRR	−1.27	0.001	68.3	−4.59	17.417	83.5	−2.46	6.663	86.6	−0.29	0.051	105.6
	1D CNN	−5.54	70.020	124.7	−4.91	32.875	145.9	−4.22	25.990	147.5	−5.06	13.685	301.9
Z1-W2	Ridge	−0.13	0.001	0.4	−0.66	1.793	0.4	−0.17	0.314	0.4	0.03	0.015	0.4
	LightGBM	−1.50	0.002	97.7	−0.62	0.233	97.0	−0.45	0.213	87.7	−0.39	0.165	97.2
	XGBoost	−1.39	0.001	121.7	−0.55	0.206	81.3	−0.59	0.646	55.8	−0.16	0.061	62.9
	Random Forest	−1.32	0.000	180.2	−0.26	0.097	62.8	−0.95	0.879	163.5	−0.14	0.104	166.5
	CatBoost	−1.22	0.009	253.2	−0.33	0.249	1186.0	−0.19	0.219	379.2	−0.12	0.213	465.4
	KRR	−0.04	0.000	45.2	−0.00	0.000	53.9	−0.01	0.000	57.6	0.04	0.014	52.1
	1D CNN	−0.60	0.057	324.5	−1.41	20.959	211.4	−0.67	1.479	140.9	−0.18	0.370	99.2
Z1-W3	Ridge	−0.25	0.001	0.5	−0.58	0.572	0.5	−0.33	0.032	0.5	−0.24	0.015	0.5
	LightGBM	−1.11	0.001	122.5	−0.86	0.475	102.0	−0.75	0.106	142.8	−0.68	0.290	98.2
	XGBoost	−1.19	0.002	96.4	−0.77	0.373	77.5	−0.85	0.144	113.1	−0.87	0.545	120.7
	Random Forest	−1.07	0.000	286.8	−0.26	0.119	100.3	−0.41	0.100	110.2	−0.55	0.235	128.4
	CatBoost	−1.04	0.001	1047.6	−0.66	0.166	434.3	−0.23	0.038	864.2	−0.53	0.613	570.0
	KRR	−0.98	0.001	81.2	−0.30	0.200	95.7	−0.27	0.104	106.5	−0.24	0.015	105.3
	1D CNN	−0.16	0.007	214.3	−1.09	6.256	296.6	−0.37	0.175	144.0	−0.90	1.798	674.2
Z2-W1	Ridge	−0.36	0.000	0.4	−0.70	0.364	0.4	−0.49	0.131	0.4	−0.30	0.071	0.4
	LightGBM	−0.74	0.002	125.7	−0.73	0.025	71.8	−0.70	0.042	66.7	−0.43	0.009	88.1
	XGBoost	−0.70	0.001	107.7	−0.65	0.010	107.8	−0.61	0.031	64.2	−0.66	0.031	52.0
	Random Forest	−0.76	0.000	178.7	−0.78	0.010	157.3	−0.88	0.029	182.7	−0.79	0.014	201.4
	CatBoost	−0.76	0.001	563.6	−0.71	0.030	384.5	−0.53	0.032	426.0	−0.35	0.013	339.0
	KRR	−0.79	0.000	46.6	−1.02	1.138	58.6	−2.20	10.405	69.1	−0.97	1.781	83.4
	1D CNN	−0.96	0.013	123.1	−0.71	0.348	249.1	−0.52	0.624	188.6	−0.51	0.203	378.4
Z2-W2	Ridge	−12.01	2.359	0.4	−1.38	0.340	0.4	−1.01	0.163	0.4	−0.69	0.043	0.5
	LightGBM	−4.07	0.019	111.4	−2.55	0.044	131.2	−2.56	0.131	105.5	−2.74	0.080	131.4
	XGBoost	−4.44	0.007	106.6	−1.96	0.091	62.1	−2.14	0.052	71.3	−2.37	0.053	87.6
	Random Forest	−3.04	0.000	203.8	−2.55	0.058	192.8	−2.55	0.084	215.4	−2.94	0.358	139.4
	CatBoost	−3.47	0.006	354.1	−2.71	0.035	835.5	−2.57	0.106	571.0	−2.63	0.084	477.1
	KRR	−5.91	0.005	159.7	−1.24	0.007	175.6	−1.22	0.013	191.1	−1.60	0.042	224.9
	1D CNN	−8.65	1.064	342.5	−2.58	1.036	298.8	−2.46	1.821	731.9	−2.63	1.441	323.9
Z2-W3	Ridge	−1.56	0.126	0.4	−2.20	1.125	0.4	−2.01	0.611	0.4	−1.67	0.106	0.5
	LightGBM	−1.14	0.004	122.4	−0.81	0.004	151.4	−0.79	0.010	162.9	−0.53	0.004	152.3
	XGBoost	−1.20	0.001	143.8	−0.91	0.009	45.3	−0.85	0.016	52.6	−0.81	0.017	52.8
	Random Forest	−1.05	0.000	345.8	−1.04	0.066	94.1	−1.03	0.043	110.7	−0.96	0.077	98.6
	CatBoost	−1.03	0.001	256.8	−1.03	0.008	878.0	−0.42	0.021	326.7	−0.62	0.024	425.0
	KRR	−1.31	0.000	86.0	−0.27	0.000	113.2	−0.20	0.000	121.3	−0.21	0.000	147.3
	1D CNN	−1.61	0.007	146.7	−1.34	0.128	159.1	−1.39	0.473	608.7	−1.24	0.361	246.7

J PER-WELL TRTR RESULTS (POFM)

Table 24 provides the per-well TRTR breakdown summarised in Table 7. Note the catastrophic Ridge instability on individual wells in the 5%/10% regime (e.g. Z1-W1 5%: $R^2 \approx -7338$); these dominate the across-well aggregate mean.

Table 24: Per-well TRTR results using POFM. σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		5% Real			10% Real			20% Real		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	−115.24	22801	0.3	−84.80	12482	0.3	−34.62	846.9	0.3
	LightGBM	−5.49	12.717	38.8	−5.11	8.953	46.9	−4.05	5.020	79.7
	XGBoost	−5.27	10.004	33.8	−5.09	9.315	78.3	−3.85	8.051	51.1
	Random Forest	−5.51	12.925	111.0	−5.26	10.668	51.2	−3.95	8.069	70.0
	CatBoost	−5.17	8.695	799.7	−4.83	9.794	337.2	−3.60	7.226	334.5
	KRR	−5.16	8.535	4.7	−83.35	12253	11.2	−34.21	823.0	16.6
	1D CNN	−154.67	61535	37.9	−46.15	7139	23.6	−22.54	762.3	33.8
Z1-W2	Ridge	−88.92	32660	0.3	−30.78	3747	0.3	−521.94	1969794	0.3
	LightGBM	−0.84	2.035	51.5	−0.85	1.654	55.9	−0.78	2.068	71.1
	XGBoost	−0.98	2.409	25.3	−0.76	1.697	57.4	−0.60	1.546	71.8
	Random Forest	−0.97	2.587	108.9	−0.79	1.380	107.3	−0.51	1.120	74.7
	CatBoost	−0.75	1.639	463.5	−0.54	0.970	479.6	−0.47	1.137	308.0
	KRR	−12.42	239.9	2.0	−0.92	2.573	9.1	−0.32	0.132	19.3
	1D CNN	−194.25	158715	25.3	−47.01	4663	49.0	−11.04	649.8	99.4
Z1-W3	Ridge	−135.25	48276	0.3	−21.91	835.1	0.3	−3.40	16.519	0.3
	LightGBM	−1.93	3.473	29.8	−2.18	4.706	46.9	−1.87	2.978	56.3
	XGBoost	−2.02	2.865	31.1	−2.07	4.232	51.1	−1.70	1.919	46.7
	Random Forest	−2.10	3.878	50.0	−2.35	5.977	115.5	−1.88	3.576	113.3
	CatBoost	−1.96	2.640	335.9	−1.94	4.265	1114.2	−1.72	2.974	425.8
	KRR	−462.32	923329	4.8	−21.73	815.1	11.2	−1.74	1.468	19.7
	1D CNN	−185.16	91847	29.2	−10.57	141.6	68.0	−8.63	208.4	135.7
Z2-W1	Ridge	−225.96	138284	0.3	−9.25	137.7	0.3	−1.32	3.161	0.3
	LightGBM	−2.00	2.226	37.4	−1.87	1.542	53.9	−1.56	1.062	128.5
	XGBoost	−1.83	2.081	45.1	−1.76	1.309	89.4	−1.44	0.936	66.3
	Random Forest	−1.81	2.157	113.2	−1.77	1.236	108.4	−1.44	0.837	63.0
	CatBoost	−1.74	1.349	419.2	−1.58	0.893	351.1	−1.42	0.990	291.9
	KRR	−1.72	1.116	3.8	−9.23	137.1	9.1	−1.32	3.160	15.1
	1D CNN	−103.03	47534	34.0	−18.33	956.6	32.8	−5.29	71.926	72.1
Z2-W2	Ridge	−37.02	4763	0.3	−9.99	147.2	0.3	−20.19	848.8	0.3
	LightGBM	−6.78	32.779	61.7	−6.06	15.987	41.4	−5.08	10.318	81.1
	XGBoost	−6.70	31.294	53.5	−5.91	14.819	56.8	−5.21	10.628	47.3
	Random Forest	−6.83	31.395	87.7	−5.90	15.818	39.3	−5.04	9.723	121.9
	CatBoost	−6.74	32.342	335.6	−5.52	13.719	250.4	−5.18	9.787	337.7
	KRR	−5.48	42.779	8.3	−4.05	20.256	17.6	−1.24	2.436	20.7
	1D CNN	−318.98	127833	47.1	−6.45	39.667	30.5	−3.90	20.241	40.0
Z2-W3	Ridge	−88.74	14640	0.3	−12.01	311.2	0.3	−2.31	9.296	0.3
	LightGBM	−0.72	0.895	35.8	−0.53	0.874	48.0	−0.31	0.325	85.4
	XGBoost	−0.91	1.677	60.5	−0.53	0.865	64.2	−0.25	0.303	57.5
	Random Forest	−0.91	1.819	88.6	−0.51	0.867	54.6	−0.23	0.262	127.4
	CatBoost	−0.89	1.714	616.8	−0.54	0.802	393.7	−0.32	0.279	348.0
	KRR	−0.69	0.697	8.3	−12.00	310.9	14.0	−0.69	0.684	18.7
	1D CNN	−91.04	15994	34.9	−57.65	8134	89.8	−3.56	23.814	107.2

K PER-WELL TSTR RESULTS (CTGAN)

Table 25 provides the TSTR breakdown based on CTGAN per-well summarized in Table 21.

Table 25: Per-well TSTR results using CTGAN generator. σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		Pure TSTR			**+5%**			**+10%**			**+20%**		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	−1.46	0.001	0.4	−10.21	113.1	0.4	−8.74	100.4	0.4	−5.93	60.575	0.4
	LightGBM	−11.56	0.015	265.6	−3.18	0.437	92.9	−8.04	9.209	92.4	−2.11	0.526	129.5
	XGBoost	−11.24	0.008	115.9	−6.14	3.197	52.8	−5.45	5.120	58.3	−4.76	4.875	49.4
	Random Forest	−11.29	0.005	289.5	−5.06	1.814	72.2	−3.04	1.927	146.7	−2.75	1.745	130.1
	CatBoost	−11.56	0.023	326.3	−5.22	0.226	516.9	−5.61	2.655	955.6	−3.65	0.530	563.2
	KRR	−1.27	0.001	77.8	−4.59	17.417	92.0	−2.46	6.663	95.9	−0.29	0.051	119.1
	1D CNN	−4.24	44.857	131.0	−4.91	32.875	146.8	−4.22	25.990	149.6	−6.80	17.364	469.6
Z1-W2	Ridge	−0.13	0.001	0.4	−0.66	1.793	0.4	−0.17	0.314	0.4	0.03	0.015	0.4
	LightGBM	−1.49	0.001	96.9	−0.54	0.182	79.0	−0.45	0.209	89.5	−0.39	0.164	89.6
	XGBoost	−1.39	0.001	121.0	−0.55	0.206	81.5	−0.59	0.646	56.7	−0.16	0.061	63.1
	Random Forest	−1.32	0.000	179.7	−0.26	0.097	62.6	−0.95	0.879	164.2	−0.14	0.104	167.0
	CatBoost	−1.22	0.009	318.9	−0.33	0.249	1252.6	−0.19	0.219	443.8	−0.12	0.213	533.7
	KRR	−0.04	0.000	50.8	−0.00	0.000	59.6	−0.01	0.000	65.1	0.04	0.014	60.2
	1D CNN	−0.60	0.057	323.5	−0.15	0.339	298.4	−1.61	11.062	173.7	−0.18	0.370	99.5
Z1-W3	Ridge	−0.25	0.001	0.4	−0.58	0.572	0.4	−0.33	0.032	0.5	−0.24	0.015	0.5
	LightGBM	−1.11	0.001	119.2	−0.86	0.476	99.6	−0.84	0.198	129.7	−0.59	0.236	114.5
	XGBoost	−1.19	0.002	95.7	−0.77	0.373	76.2	−0.85	0.144	109.7	−0.87	0.545	117.7
	Random Forest	−1.07	0.000	280.6	−0.26	0.119	97.1	−0.41	0.100	107.8	−0.55	0.235	124.9
	CatBoost	−1.04	0.001	1086.8	−0.66	0.166	489.6	−0.23	0.038	910.3	−0.53	0.613	617.0
	KRR	−0.98	0.001	90.9	−0.30	0.200	107.6	−0.27	0.104	117.0	−0.24	0.015	113.2
	1D CNN	−1.87	0.482	145.9	−0.46	0.540	366.6	−0.68	0.478	160.7	−0.63	0.124	317.3
Z2-W1	Ridge	−0.36	0.000	0.4	−0.70	0.364	0.4	−0.49	0.131	0.5	−0.30	0.071	0.4
	LightGBM	−0.74	0.002	125.3	−0.73	0.025	85.4	−0.70	0.042	66.6	−0.61	0.034	84.4
	XGBoost	−0.70	0.001	108.8	−0.65	0.010	108.5	−0.61	0.031	65.2	−0.66	0.031	52.5
	Random Forest	−0.76	0.000	179.2	−0.78	0.010	157.6	−0.88	0.029	182.4	−0.79	0.014	202.0
	CatBoost	−0.76	0.001	631.6	−0.71	0.030	450.3	−0.53	0.032	494.7	−0.35	0.013	404.9
	KRR	−0.79	0.000	54.1	−1.02	1.138	65.3	−2.20	10.405	78.3	−0.97	1.781	94.1
	1D CNN	−0.96	0.013	123.8	−0.48	0.347	155.8	−0.60	0.586	173.8	−0.45	0.207	282.7
Z2-W2	Ridge	−5.93	2.711	0.4	−0.92	0.264	0.4	−0.66	0.110	0.5	−0.43	0.025	0.5
	LightGBM	−2.96	0.057	104.4	−1.86	0.090	161.7	−1.87	0.111	154.9	−2.09	0.159	108.0
	XGBoost	−3.53	0.010	124.1	−2.24	0.240	48.6	−2.33	0.087	68.2	−3.47	0.309	42.5
	Random Forest	−2.11	0.001	235.0	−2.25	0.289	181.7	−1.86	0.112	173.7	−1.87	0.137	194.2
	CatBoost	−3.86	0.042	327.7	−2.64	0.097	1431.3	−2.68	0.079	441.9	−3.11	0.166	309.6
	KRR	−5.43	0.000	111.6	−0.91	0.007	131.5	−0.87	0.009	134.6	−0.89	0.011	170.0
	1D CNN	−3.04	0.101	405.5	−1.92	1.128	180.7	−1.93	2.156	579.4	−1.58	0.902	210.8
Z2-W3	Ridge	−1.56	0.126	0.4	−2.20	1.125	0.5	−2.01	0.611	0.4	−1.67	0.106	0.4
	LightGBM	−1.15	0.003	124.5	−0.81	0.004	150.3	−0.79	0.010	163.1	−0.53	0.004	153.1
	XGBoost	−1.20	0.001	143.3	−0.91	0.009	45.1	−0.85	0.016	52.7	−0.81	0.017	53.1
	Random Forest	−1.05	0.000	345.4	−1.10	0.043	103.5	−1.03	0.043	111.4	−0.95	0.077	98.0
	CatBoost	−1.03	0.001	321.8	−1.03	0.008	949.9	−0.42	0.021	390.6	−0.62	0.024	493.4
	KRR	−1.31	0.000	92.8	−0.27	0.000	129.9	−0.20	0.000	141.2	−0.21	0.000	161.6
	1D CNN	−1.40	0.005	205.8	−1.34	0.128	161.0	−1.39	0.473	612.4	−1.12	0.688	320.2

L PER-WELL TRTR RESULTS (CTGAN)

Table 26 provides the TRTR breakdown based on CTGAN per-well summarized in Table 7.

Table 26: Per-well TRTR results using CTGAN. σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		5% Real			10% Real			20% Real		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	−115.24	22801	0.3	−84.80	12482	0.3	−34.62	846.9	0.3
	LightGBM	−5.49	12.717	39.0	−5.08	8.814	52.8	−4.05	5.012	81.3
	XGBoost	−5.27	10.004	33.8	−5.09	9.315	78.6	−3.85	8.051	51.6
	Random Forest	−5.51	12.925	114.8	−5.26	10.668	52.4	−3.95	8.069	72.5
	CatBoost	−5.17	8.695	812.3	−4.83	9.794	345.4	−3.60	7.226	339.3
	KRR	−5.16	8.535	3.8	−83.35	12253	11.2	−34.21	823.0	15.7
	1D CNN	−154.43	61345	40.9	−46.15	7138	24.3	−22.54	762.4	35.9
Z1-W2	Ridge	−88.92	32660	0.3	−30.78	3747	0.3	−521.94	1969794	0.3
	LightGBM	−0.84	2.037	51.7	−0.85	1.653	56.8	−0.78	2.068	72.1
	XGBoost	−0.98	2.409	25.5	−0.76	1.697	57.2	−0.60	1.546	72.3
	Random Forest	−0.97	2.587	111.4	−0.79	1.380	108.8	−0.51	1.120	76.0
	CatBoost	−0.75	1.639	467.0	−0.54	0.970	487.5	−0.47	1.137	310.9
	KRR	−12.42	239.9	2.4	−0.92	2.573	10.3	−0.32	0.132	15.7
	1D CNN	−196.69	162163	26.8	−47.01	4663	51.5	−11.24	262.4	69.3
Z1-W3	Ridge	−135.25	48276	0.3	−21.91	835.1	0.3	−3.40	16.519	0.3
	LightGBM	−1.99	3.920	32.5	−2.18	4.706	48.5	−1.83	2.907	68.4
	XGBoost	−2.02	2.865	31.4	−2.07	4.232	53.4	−1.70	1.919	47.7
	Random Forest	−2.10	3.878	51.9	−2.35	5.977	121.5	−1.88	3.576	118.0
	CatBoost	−1.96	2.640	339.9	−1.94	4.265	1145.7	−1.72	2.974	434.0
	KRR	−462.32	923329	4.5	−21.73	815.1	11.3	−1.74	1.468	17.8
	1D CNN	−139.47	38858	35.7	−26.99	2444	70.4	−9.76	223.2	129.0
Z2-W1	Ridge	−225.96	138284	0.3	−9.25	137.7	0.3	−1.32	3.161	0.3
	LightGBM	−2.00	2.226	38.2	−1.87	1.543	54.5	−1.56	1.062	131.3
	XGBoost	−1.83	2.081	45.6	−1.76	1.309	90.6	−1.44	0.936	67.9
	Random Forest	−1.81	2.157	117.3	−1.77	1.236	111.9	−1.44	0.837	65.3
	CatBoost	−1.74	1.349	426.7	−1.58	0.893	355.5	−1.42	0.990	295.4
	KRR	−1.72	1.116	3.0	−9.23	137.1	10.8	−1.32	3.160	14.8
	1D CNN	−143.51	66347	27.6	−15.74	822.2	60.2	−7.87	182.5	46.9
Z2-W2	Ridge	−37.02	4763	0.3	−9.99	147.2	0.3	−20.19	848.8	0.3
	LightGBM	−6.78	32.779	63.0	−6.06	15.987	42.8	−5.08	10.317	83.6
	XGBoost	−6.70	31.294	54.1	−5.91	14.819	57.2	−5.21	10.628	48.1
	Random Forest	−6.83	31.395	89.0	−5.90	15.818	39.6	−5.04	9.723	123.5
	CatBoost	−6.74	32.342	333.8	−5.52	13.719	251.4	−5.18	9.787	341.2
	KRR	−5.48	42.779	8.3	−4.05	20.256	17.3	−1.24	2.436	18.8
	1D CNN	−150.24	34137	39.6	−14.09	256.3	22.9	−2.08	8.841	46.7
Z2-W3	Ridge	−88.74	14640	0.3	−12.01	311.2	0.3	−2.31	9.296	0.4
	LightGBM	−0.72	0.895	36.4	−0.53	0.874	48.9	−0.30	0.320	68.5
	XGBoost	−0.91	1.677	61.6	−0.53	0.865	66.1	−0.25	0.303	58.5
	Random Forest	−0.91	1.819	92.5	−0.51	0.867	57.0	−0.23	0.262	132.8
	CatBoost	−0.89	1.714	634.7	−0.54	0.802	401.7	−0.32	0.279	354.4
	KRR	−0.69	0.697	8.6	−12.00	310.9	14.4	−0.69	0.684	21.1
	1D CNN	−231.26	95957	50.1	−30.99	2992	99.6	−7.29	86.506	75.9

M PER-WELL TSTR RESULTS (SMOTE)

Table 27 gives the per-well TSTR breakdown for SMOTE. Each cell reports the mean test-well R^2 over 20 seeds, the within-well variance σ^2, and wall-clock time (s).

Table 27: Per-well TSTR results using SMOTE. σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		Pure TSTR			**+5%**			**+10%**			**+20%**		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	0.29	0.000	0.4	0.19	0.000	0.4	0.18	0.001	0.4	0.18	0.002	0.4
	LightGBM	1.00	0.000	178.9	1.00	0.000	187.4	1.00	0.000	241.5	1.00	0.000	227.5
	XGBoost	1.00	0.000	75.3	1.00	0.000	119.4	1.00	0.000	114.9	0.99	0.000	105.8
	Random Forest	1.00	0.000	90.3	1.00	0.000	174.9	1.00	0.000	186.6	1.00	0.000	179.7
	CatBoost	0.99	0.000	722.6	0.99	0.000	670.6	0.98	0.000	622.9	0.98	0.000	750.1
	KRR	0.93	0.000	62.7	0.95	0.000	84.8	0.95	0.000	89.1	0.95	0.000	109.9
	1D CNN	0.70	0.000	368.9	0.67	0.000	363.9	0.69	0.000	381.0	0.68	0.000	386.4
Z1-W2	Ridge	0.43	0.000	0.4	0.42	0.000	0.4	0.41	0.000	0.4	0.40	0.001	0.4
	LightGBM	0.99	0.000	228.1	0.99	0.000	188.7	0.99	0.000	192.9	0.99	0.000	166.7
	XGBoost	0.99	0.000	213.4	0.99	0.000	133.7	0.99	0.000	202.2	0.99	0.000	187.4
	Random Forest	0.99	0.000	78.4	0.99	0.000	100.4	0.99	0.000	149.6	0.99	0.000	128.9
	CatBoost	0.98	0.000	630.2	0.98	0.000	366.9	0.98	0.000	734.3	0.98	0.000	583.9
	KRR	0.93	0.000	44.3	0.93	0.000	51.7	0.93	0.000	57.6	0.93	0.000	69.6
	1D CNN	0.72	0.001	238.9	0.63	0.003	318.8	0.58	0.004	300.4	0.60	0.004	284.2
Z1-W3	Ridge	0.33	0.000	0.4	0.32	0.000	0.4	0.32	0.000	0.4	0.32	0.000	0.4
	LightGBM	0.98	0.000	190.1	0.98	0.000	208.8	0.98	0.000	191.9	0.98	0.000	192.3
	XGBoost	0.97	0.000	163.1	0.97	0.000	125.6	0.97	0.000	115.9	0.97	0.000	142.5
	Random Forest	0.98	0.000	89.2	0.97	0.000	161.8	0.97	0.000	188.9	0.97	0.000	84.8
	CatBoost	0.95	0.000	723.6	0.95	0.000	931.9	0.95	0.000	1348.6	0.94	0.000	1650.4
	KRR	0.79	0.000	76.8	0.80	0.001	92.5	0.81	0.000	100.0	0.82	0.001	123.1
	1D CNN	0.55	0.000	206.3	0.49	0.011	401.0	0.38	0.022	217.5	0.43	0.012	476.4
Z2-W1	Ridge	0.48	0.000	0.4	0.38	0.000	0.4	0.40	0.000	0.4	0.42	0.001	0.4
	LightGBM	0.98	0.000	141.5	0.99	0.000	326.1	0.98	0.000	197.3	0.98	0.000	170.5
	XGBoost	0.98	0.000	143.7	0.98	0.000	184.1	0.98	0.000	149.1	0.98	0.000	165.5
	Random Forest	0.99	0.000	84.8	0.99	0.000	133.1	0.99	0.000	142.5	0.99	0.000	123.5
	CatBoost	0.98	0.000	700.5	0.97	0.000	955.1	0.97	0.000	590.7	0.97	0.000	870.2
	KRR	0.80	0.000	52.8	0.81	0.000	61.7	0.81	0.000	64.9	0.81	0.000	78.6
	1D CNN	0.65	0.001	312.9	0.64	0.002	349.4	0.65	0.001	312.8	0.60	0.001	364.5
Z2-W2	Ridge	0.42	0.000	0.5	0.41	0.001	0.5	0.40	0.001	0.5	0.37	0.001	0.5
	LightGBM	0.94	0.000	178.3	0.94	0.000	241.6	0.94	0.000	206.1	0.94	0.000	224.4
	XGBoost	0.93	0.000	136.7	0.94	0.000	138.8	0.94	0.000	225.8	0.94	0.000	269.2
	Random Forest	0.95	0.000	194.9	0.95	0.000	108.2	0.95	0.000	98.4	0.95	0.000	101.0
	CatBoost	0.91	0.001	1302.6	0.92	0.001	1246.3	0.93	0.001	764.9	0.93	0.000	1288.1
	KRR	0.82	0.000	133.2	0.82	0.000	173.4	0.82	0.000	191.1	0.82	0.000	209.2
	1D CNN	0.60	0.002	509.5	0.62	0.001	611.7	0.53	0.009	583.1	0.38	0.027	441.9
Z2-W3	Ridge	0.58	0.000	0.4	0.54	0.000	0.5	0.52	0.000	0.5	0.52	0.000	0.5
	LightGBM	0.99	0.000	78.8	0.99	0.000	189.9	0.98	0.000	167.3	0.98	0.000	210.2
	XGBoost	0.98	0.000	222.5	0.98	0.000	256.5	0.98	0.000	231.6	0.98	0.000	128.8
	Random Forest	0.99	0.000	106.4	0.99	0.000	169.7	0.99	0.000	179.5	0.99	0.000	200.4
	CatBoost	0.97	0.000	383.8	0.96	0.000	353.2	0.95	0.000	337.7	0.96	0.000	422.5
	KRR	0.83	0.000	100.1	0.83	0.000	112.3	0.83	0.000	135.2	0.83	0.000	166.6
	1D CNN	0.70	0.002	433.6	0.68	0.000	374.7	0.61	0.001	196.4	0.68	0.000	461.1

N PER-WELL TRTR RESULTS (SMOTE)

Table 28 gives the per-well TRTR breakdown for the SMOTE pipeline. TRTR trains on real data only.

Table 28: Per-well TRTR results (SMOTE pipeline). σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		5% Real			10% Real			20% Real		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	−115.24	22801	0.3	−84.80	12482	0.3	−34.62	846.9	0.4
	LightGBM	−5.49	12.717	38.9	−5.11	8.952	46.9	−4.05	5.026	80.0
	XGBoost	−5.27	10.004	33.3	−5.09	9.315	77.9	−3.85	8.051	51.2
	Random Forest	−5.51	12.925	112.3	−5.26	10.668	51.6	−3.95	8.069	70.6
	CatBoost	−5.17	8.695	798.4	−4.83	9.794	337.3	−3.60	7.226	335.6
	KRR	−5.16	8.535	3.7	−83.35	12253	10.9	−34.21	823.0	16.9
	1D CNN	−126.09	39069	37.6	−46.15	7139	23.5	−22.54	762.3	33.6
Z1-W2	Ridge	−88.92	32660	0.3	−30.78	3747	0.3	−521.94	1969794	0.3
	LightGBM	−0.84	2.036	51.5	−0.85	1.653	56.4	−0.78	2.068	72.0
	XGBoost	−0.98	2.409	24.8	−0.76	1.697	56.5	−0.60	1.546	71.2
	Random Forest	−0.97	2.587	109.9	−0.79	1.380	107.7	−0.51	1.120	75.3
	CatBoost	−0.75	1.639	458.8	−0.54	0.970	477.5	−0.47	1.137	307.0
	KRR	−12.42	239.9	2.3	−0.92	2.573	8.5	−0.32	0.132	15.9
	1D CNN	−194.61	159972	25.4	−47.01	4663	49.4	−14.94	584.8	63.8
Z1-W3	Ridge	−135.25	48276	0.3	−21.91	835.1	0.3	−3.40	16.519	0.3
	LightGBM	−2.01	3.996	29.9	−2.18	4.706	46.8	−1.87	2.975	57.0
	XGBoost	−2.02	2.865	30.7	−2.07	4.232	51.1	−1.70	1.919	46.8
	Random Forest	−2.10	3.878	50.1	−2.35	5.977	117.7	−1.88	3.576	115.1
	CatBoost	−1.96	2.640	335.2	−1.94	4.265	1118.2	−1.72	2.974	426.2
	KRR	−462.32	923329	5.5	−21.73	815.1	10.9	−1.74	1.468	18.7
	1D CNN	−135.63	32661	33.3	−21.04	1000	59.2	−8.20	172.3	137.7
Z2-W1	Ridge	−225.96	138284	0.3	−9.25	137.7	0.3	−1.32	3.161	0.3
	LightGBM	−2.00	2.226	37.5	−1.87	1.542	54.0	−1.56	1.062	129.5
	XGBoost	−1.83	2.081	44.8	−1.76	1.309	88.6	−1.44	0.936	66.6
	Random Forest	−1.81	2.157	115.1	−1.77	1.236	110.0	−1.44	0.837	63.6
	CatBoost	−1.74	1.349	420.2	−1.58	0.893	351.1	−1.42	0.990	292.3
	KRR	−1.72	1.116	3.5	−9.23	137.1	10.0	−1.32	3.160	15.0
	1D CNN	−165.75	97765	21.9	−16.53	944.4	58.4	−5.16	40.534	46.7
Z2-W2	Ridge	−37.02	4763	0.3	−9.99	147.2	0.3	−20.19	848.8	0.3
	LightGBM	−6.78	32.779	62.7	−6.06	15.987	41.9	−5.08	10.318	83.4
	XGBoost	−6.70	31.294	53.7	−5.91	14.819	57.6	−5.21	10.628	47.4
	Random Forest	−6.83	31.395	88.7	−5.90	15.818	39.9	−5.04	9.723	124.7
	CatBoost	−6.74	32.342	334.4	−5.52	13.719	250.6	−5.18	9.787	339.8
	KRR	−5.48	42.779	8.6	−4.05	20.256	17.9	−1.24	2.436	20.4
	1D CNN	−120.20	32392	40.1	−6.45	39.667	30.9	−3.90	20.241	40.0
Z2-W3	Ridge	−88.74	14640	0.3	−12.01	311.2	0.3	−2.31	9.296	0.3
	LightGBM	−0.72	0.895	35.6	−0.53	0.874	47.6	−0.30	0.319	67.3
	XGBoost	−0.91	1.677	61.1	−0.53	0.865	64.9	−0.25	0.303	58.2
	Random Forest	−0.91	1.819	90.5	−0.51	0.867	55.7	−0.23	0.262	129.9
	CatBoost	−0.89	1.714	624.8	−0.54	0.802	398.8	−0.32	0.279	351.2
	KRR	−0.69	0.697	7.7	−12.00	310.9	14.3	−0.69	0.684	21.6
	1D CNN	−91.55	12538	48.3	−41.93	3501	95.2	−21.73	631.9	96.7

O PER-WELL TSTR RESULTS (SMOOTHED BOOTSTRAP)

Table 29 gives the per-well TSTR breakdown for Smoothed Bootstrap. Each cell reports the mean test-well R^2 over 20 seeds, the within-well variance σ^2, and wall-clock time (s).

Table 29: Per-well TSTR results using Smoothed Bootstrap. σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		Pure TSTR			+5%			+10%			+20%		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	0.20	0.000	0.4	0.17	0.000	0.5	0.17	0.001	0.5	0.16	0.002	0.5
	LightGBM	0.94	0.000	251.9	0.95	0.000	211.0	0.95	0.000	240.3	0.94	0.000	249.7
	XGBoost	0.95	0.000	240.5	0.95	0.000	110.4	0.95	0.000	196.0	0.95	0.000	290.9
	Random Forest	0.95	0.000	91.0	0.94	0.000	316.7	0.95	0.000	173.8	0.95	0.000	257.9
	CatBoost	0.94	0.000	1481.3	0.95	0.000	655.4	0.94	0.000	1418.2	0.95	0.000	386.4
	KRR	0.93	0.000	65.6	0.92	0.000	82.1	0.93	0.000	89.0	0.93	0.000	109.7
	1D CNN	0.66	0.001	372.9	0.62	0.001	431.0	0.65	0.001	380.7	0.58	0.003	214.1
Z1-W2	Ridge	0.40	0.000	0.4	0.40	0.000	0.4	0.40	0.000	0.4	0.38	0.001	0.4
	LightGBM	0.95	0.000	217.2	0.95	0.000	253.0	0.95	0.000	296.7	0.95	0.000	136.9
	XGBoost	0.95	0.000	170.7	0.95	0.000	110.0	0.96	0.000	132.6	0.96	0.000	170.2
	Random Forest	0.95	0.000	168.8	0.95	0.000	99.6	0.95	0.000	101.5	0.95	0.000	87.1
	CatBoost	0.95	0.000	1094.5	0.95	0.000	1510.0	0.95	0.000	1624.3	0.95	0.000	1243.2
	KRR	0.92	0.000	45.1	0.91	0.000	54.6	0.91	0.000	57.0	0.92	0.000	65.9
	1D CNN	0.69	0.001	220.5	0.63	0.002	255.1	0.63	0.004	286.6	0.63	0.002	330.0
Z1-W3	Ridge	0.32	0.000	0.4	0.31	0.000	0.4	0.31	0.000	0.4	0.30	0.000	0.4
	LightGBM	0.80	0.000	155.3	0.82	0.000	325.4	0.83	0.000	507.8	0.81	0.000	339.3
	XGBoost	0.83	0.000	262.5	0.83	0.000	131.2	0.83	0.000	136.4	0.82	0.000	116.6
	Random Forest	0.81	0.000	132.9	0.80	0.000	414.1	0.81	0.000	253.2	0.80	0.000	256.7
	CatBoost	0.83	0.000	1484.0	0.82	0.000	897.6	0.83	0.000	342.6	0.83	0.000	418.2
	KRR	0.80	0.000	79.2	0.80	0.000	92.2	0.80	0.000	102.7	0.80	0.000	122.1
	1D CNN	0.49	0.002	148.9	0.42	0.012	157.7	0.42	0.010	176.6	0.43	0.010	389.3
Z2-W1	Ridge	0.31	0.000	0.4	0.31	0.000	0.4	0.32	0.000	0.4	0.32	0.000	0.4
	LightGBM	0.90	0.000	253.2	0.90	0.000	273.3	0.91	0.000	241.9	0.90	0.000	248.8
	XGBoost	0.91	0.000	117.4	0.91	0.000	112.4	0.91	0.000	129.8	0.91	0.000	148.1
	Random Forest	0.91	0.000	80.3	0.91	0.000	319.5	0.91	0.000	215.9	0.91	0.000	371.1
	CatBoost	0.90	0.000	1400.3	0.90	0.000	361.1	0.90	0.000	375.2	0.89	0.000	1133.0
	KRR	0.80	0.000	49.9	0.80	0.000	59.6	0.80	0.000	68.5	0.80	0.000	82.2
	1D CNN	0.61	0.001	322.3	0.57	0.002	335.7	0.49	0.001	233.9	0.56	0.001	412.6
Z2-W2	Ridge	0.37	0.000	0.4	0.36	0.001	0.5	0.34	0.001	0.5	0.33	0.001	0.5
	LightGBM	0.88	0.000	425.4	0.88	0.000	356.3	0.88	0.000	436.7	0.88	0.000	374.4
	XGBoost	0.89	0.000	186.3	0.89	0.000	372.0	0.89	0.000	269.8	0.88	0.000	202.9
	Random Forest	0.87	0.000	176.8	0.88	0.000	139.2	0.88	0.000	236.3	0.88	0.000	560.3
	CatBoost	0.87	0.000	1259.0	0.87	0.000	391.4	0.87	0.000	383.7	0.87	0.000	1463.6
	KRR	0.79	0.000	143.5	0.81	0.000	180.2	0.81	0.000	198.3	0.82	0.000	230.2
	1D CNN	0.57	0.003	322.2	0.61	0.001	274.3	0.35	0.012	544.3	0.56	0.002	354.7
Z2-W3	Ridge	0.50	0.000	0.4	0.49	0.000	0.5	0.49	0.000	0.5	0.48	0.000	0.5
	LightGBM	0.86	0.000	133.2	0.87	0.000	222.3	0.87	0.000	249.4	0.86	0.000	277.3
	XGBoost	0.88	0.000	257.4	0.88	0.000	274.0	0.87	0.000	205.0	0.87	0.000	310.7
	Random Forest	0.86	0.000	180.0	0.87	0.000	384.9	0.87	0.000	295.1	0.87	0.000	339.8
	CatBoost	0.86	0.000	1485.9	0.86	0.000	896.9	0.87	0.000	357.3	0.87	0.000	380.3
	KRR	0.81	0.000	105.1	0.82	0.000	123.4	0.82	0.000	146.9	0.82	0.000	179.5
	1D CNN	0.71	0.001	421.5	0.65	0.000	183.1	0.70	0.000	312.1	0.67	0.001	443.1

P PER-WELL TRTR RESULTS (SMOOTHED BOOTSTRAP)

Table 30 gives the per-well TRTR breakdown for the Smoothed Bootstrap pipeline. TRTR trains on real data only.

Table 30: Per-well TRTR results (Smoothed Bootstrap pipeline). σ^2 = within-well variance across 20 random seeds. "s" = wall-clock time in seconds.

		5% Real			10% Real			20% Real		
Zone/Well	**Algorithm**	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s	$R^2\uparrow$	σ^2	s
Z1-W1	Ridge	−115.24	22801	0.3	−84.80	12482	0.3	−34.62	846.9	0.3
	LightGBM	−5.49	12.717	39.1	−5.08	8.815	53.2	−4.05	5.024	79.8
	XGBoost	−5.27	10.004	33.8	−5.09	9.315	78.3	−3.85	8.051	50.5
	Random Forest	−5.51	12.925	111.1	−5.26	10.668	51.1	−3.95	8.069	69.9
	CatBoost	−5.17	8.695	801.5	−4.83	9.794	338.3	−3.60	7.226	336.3
	KRR	−5.16	8.535	3.8	−83.35	12253	11.3	−34.21	823.0	15.2
	1D CNN	−154.56	61620	38.5	−46.15	7137	23.7	−22.54	762.0	34.1
Z1-W2	Ridge	−88.92	32660	0.3	−30.78	3747	0.3	−521.94	1969794	0.3
	LightGBM	−0.84	2.035	51.8	−0.85	1.653	56.1	−0.78	2.068	69.0
	XGBoost	−0.98	2.409	24.8	−0.76	1.697	56.4	−0.60	1.546	71.3
	Random Forest	−0.97	2.587	109.3	−0.79	1.380	107.8	−0.51	1.120	75.0
	CatBoost	−0.75	1.639	461.3	−0.54	0.970	478.9	−0.47	1.137	308.0
	KRR	−12.42	239.9	1.9	−0.92	2.573	9.1	−0.32	0.132	15.7
	1D CNN	−195.43	161133	26.1	−47.01	4663	50.5	−14.50	475.3	117.9
Z1-W3	Ridge	−135.25	48276	0.3	−21.91	835.1	0.3	−3.40	16.519	0.3
	LightGBM	−2.01	3.996	30.1	−2.18	4.706	47.2	−1.83	2.908	66.3
	XGBoost	−2.02	2.865	30.9	−2.07	4.232	52.1	−1.70	1.919	46.7
	Random Forest	−2.10	3.878	50.1	−2.35	5.977	117.0	−1.88	3.576	114.1
	CatBoost	−1.96	2.640	335.4	−1.94	4.265	1122.1	−1.72	2.974	425.9
	KRR	−462.32	923329	4.5	−21.73	815.1	11.1	−1.74	1.468	16.9
	1D CNN	−134.73	35069	35.5	−11.38	188.8	73.1	−7.49	184.5	139.8
Z2-W1	Ridge	−225.96	138284	0.3	−9.25	137.7	0.3	−1.32	3.161	0.3
	LightGBM	−2.00	2.226	38.1	−1.87	1.542	54.6	−1.56	1.062	129.2
	XGBoost	−1.83	2.081	44.7	−1.76	1.309	89.1	−1.44	0.936	66.1
	Random Forest	−1.81	2.157	114.4	−1.77	1.236	109.5	−1.44	0.837	63.4
	CatBoost	−1.74	1.349	421.6	−1.58	0.893	350.2	−1.42	0.990	291.5
	KRR	−1.72	1.116	2.9	−9.23	137.1	10.4	−1.32	3.160	14.9
	1D CNN	−141.28	61775	26.8	−18.00	909.6	32.4	−5.16	40.534	46.1
Z2-W2	Ridge	−37.02	4763	0.3	−9.99	147.2	0.3	−20.19	848.8	0.3
	LightGBM	−6.78	32.779	60.5	−6.06	15.987	41.4	−5.08	10.317	81.5
	XGBoost	−6.70	31.294	53.1	−5.91	14.819	55.9	−5.21	10.628	46.4
	Random Forest	−6.83	31.395	87.9	−5.90	15.818	39.3	−5.04	9.723	122.1
	CatBoost	−6.74	32.342	331.1	−5.52	13.719	249.8	−5.18	9.787	341.5
	KRR	−5.48	42.779	7.6	−4.05	20.256	18.3	−1.24	2.436	20.7
	1D CNN	−172.24	99071	48.1	−6.45	39.667	31.1	−3.90	20.241	40.4
Z2-W3	Ridge	−88.74	14640	0.3	−12.01	311.2	0.3	−2.31	9.296	0.3
	LightGBM	−0.72	0.895	35.5	−0.53	0.874	47.9	−0.30	0.319	66.5
	XGBoost	−0.91	1.677	60.1	−0.53	0.865	64.6	−0.25	0.303	57.0
	Random Forest	−0.91	1.819	89.4	−0.51	0.867	56.0	−0.23	0.262	133.2
	CatBoost	−0.89	1.714	620.8	−0.54	0.802	394.7	−0.32	0.279	348.9
	KRR	−0.69	0.697	8.1	−12.00	310.9	14.6	−0.69	0.684	19.0
	1D CNN	−154.08	45756	44.8	−31.06	2765	95.5	−18.62	706.4	98.0

Q PER-WELL SYNTHETIC LOG COMPARISONS

Each figure compares, for one well, the real logs against all four generators (POFM, CTGAN, SMOTE, and Smoothed Bootstrap) across the eight log variables, sharing the depth axis and per-log x-ranges. Each panel shows the eight logs, arranged two generators per row for legibility.

Q.1 ZONE 1: MISSA KESWAL

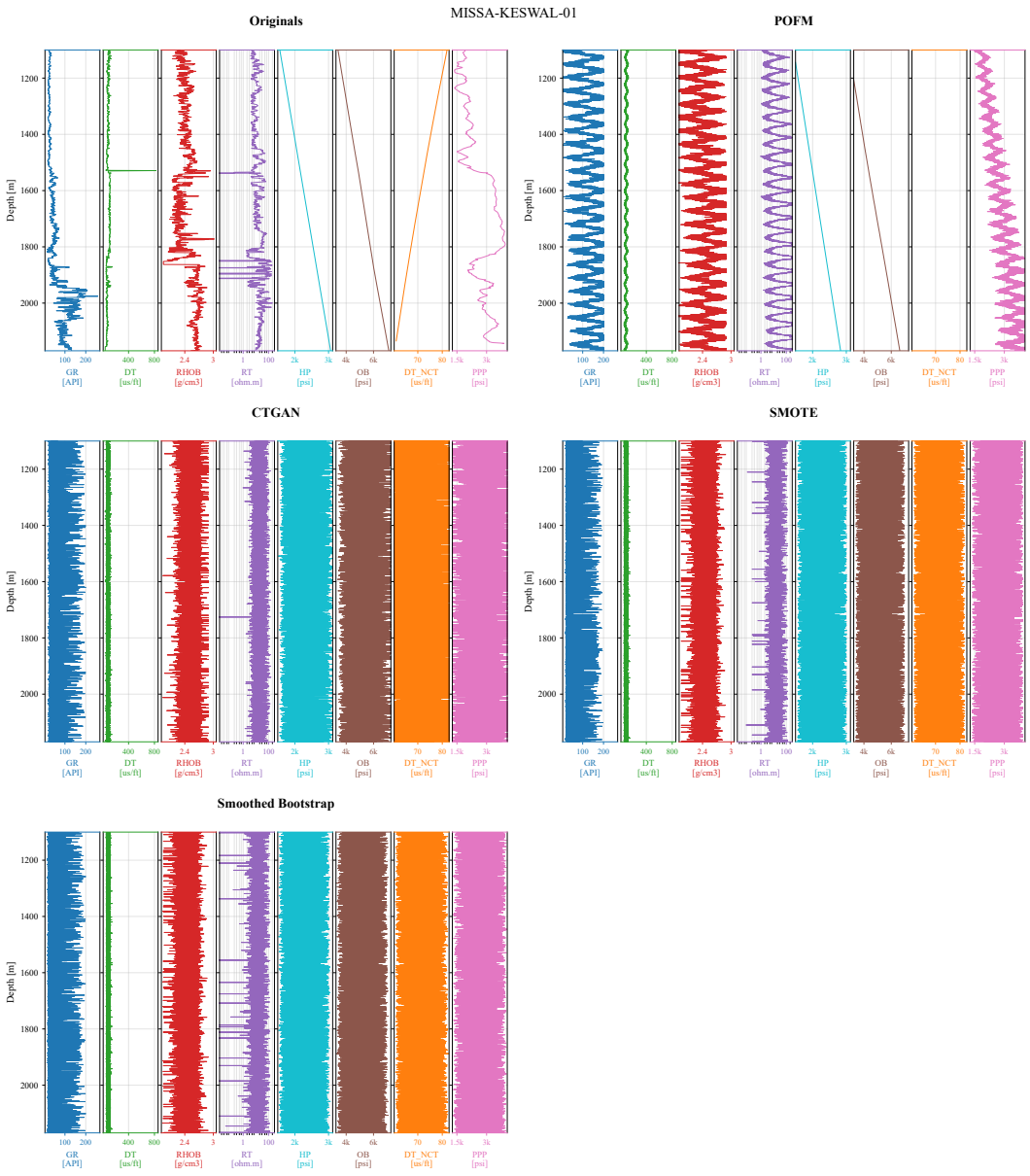

Figure 2: MISSA-KESWAL-01: real, POFM, CTGAN, SMOTE, and Smoothed Bootstrap logs.

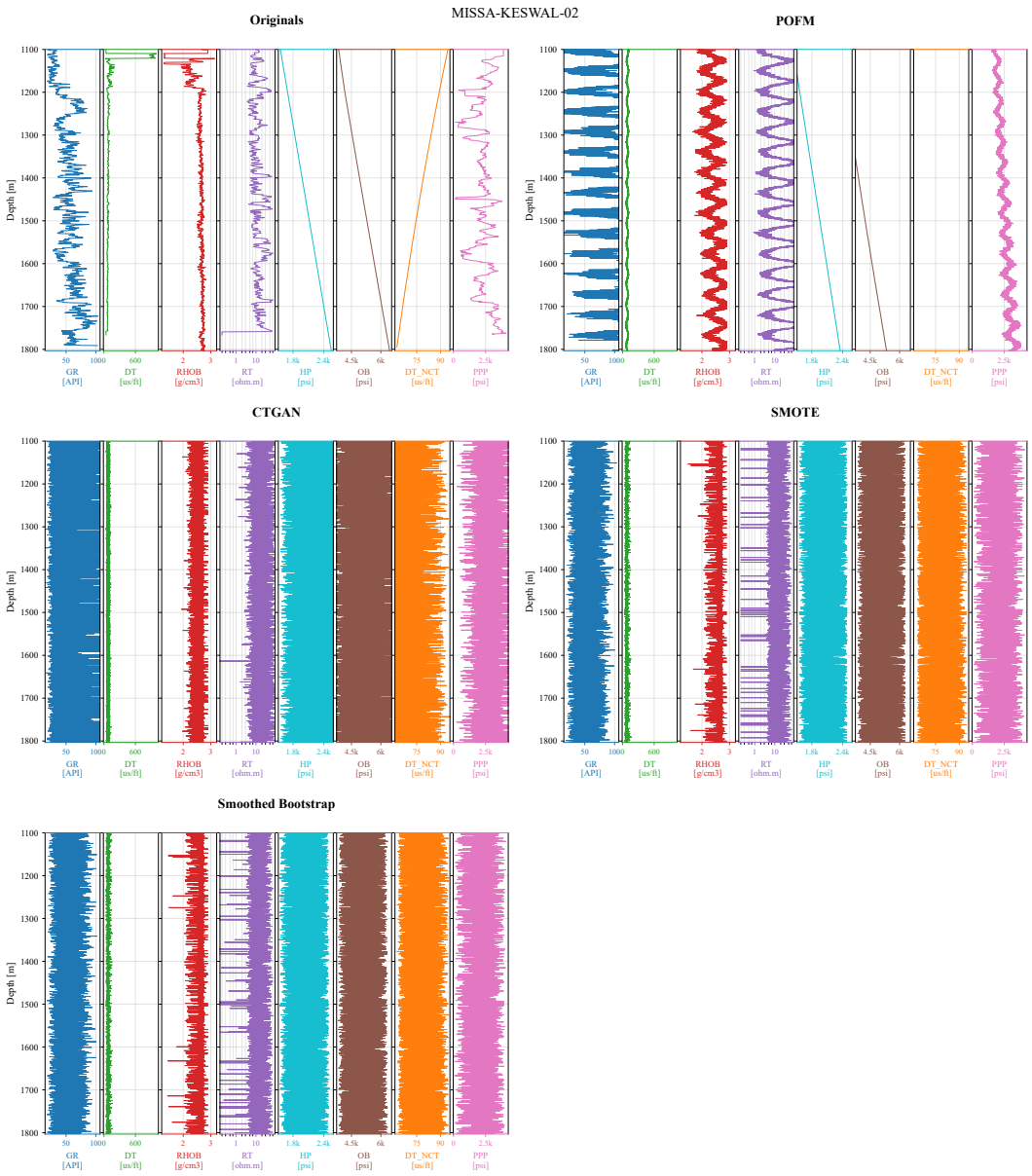

Figure 3: MISSA-KESWAL-02: real, POFM, CTGAN, SMOTE, and Smoothed Bootstrap logs.

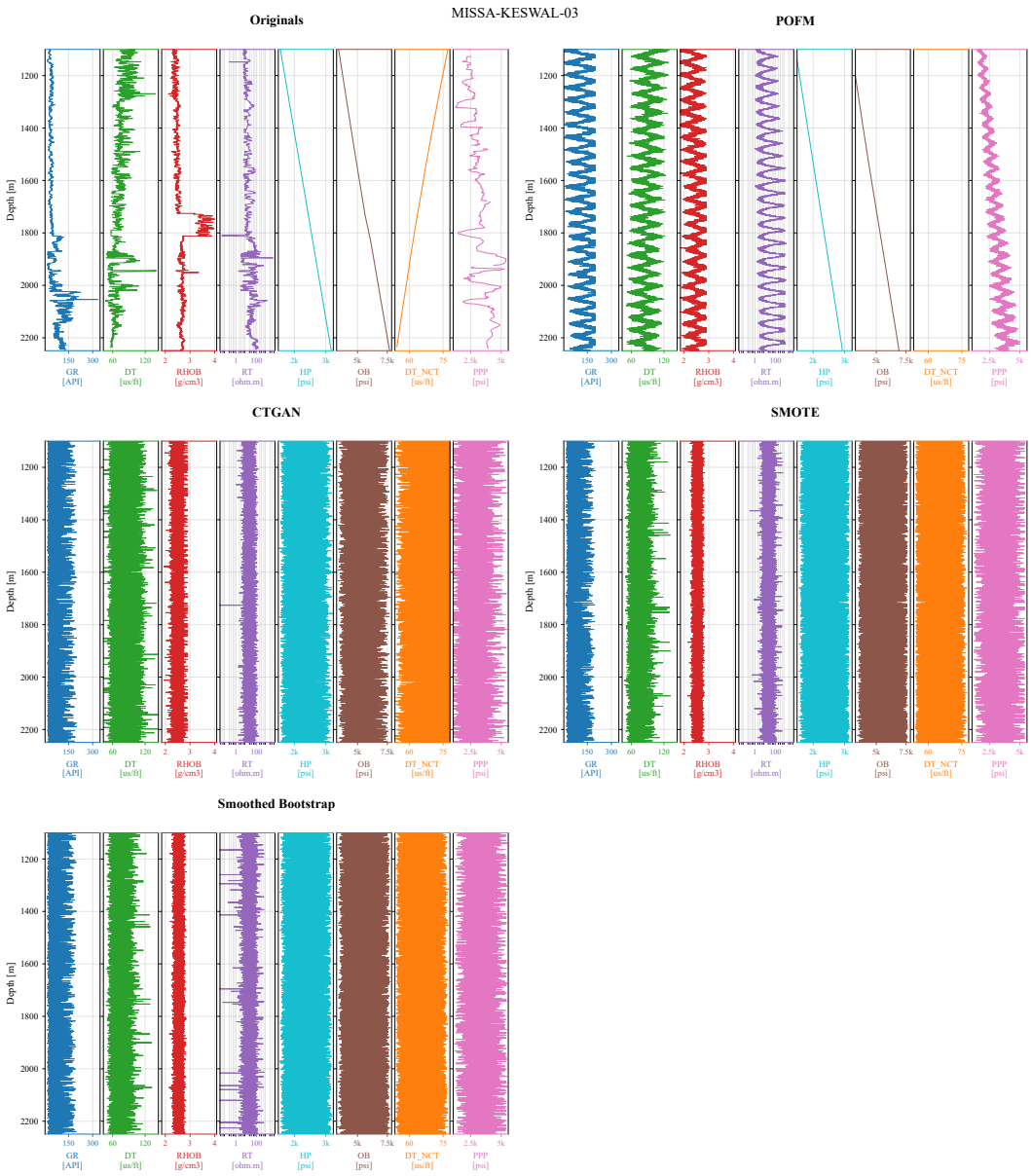

Figure 4: MISSA-KESWAL-03: real, POFM, CTGAN, SMOTE, and Smoothed Bootstrap logs.

Q.2 ZONE 2: JOYAMAIR–MINWAL

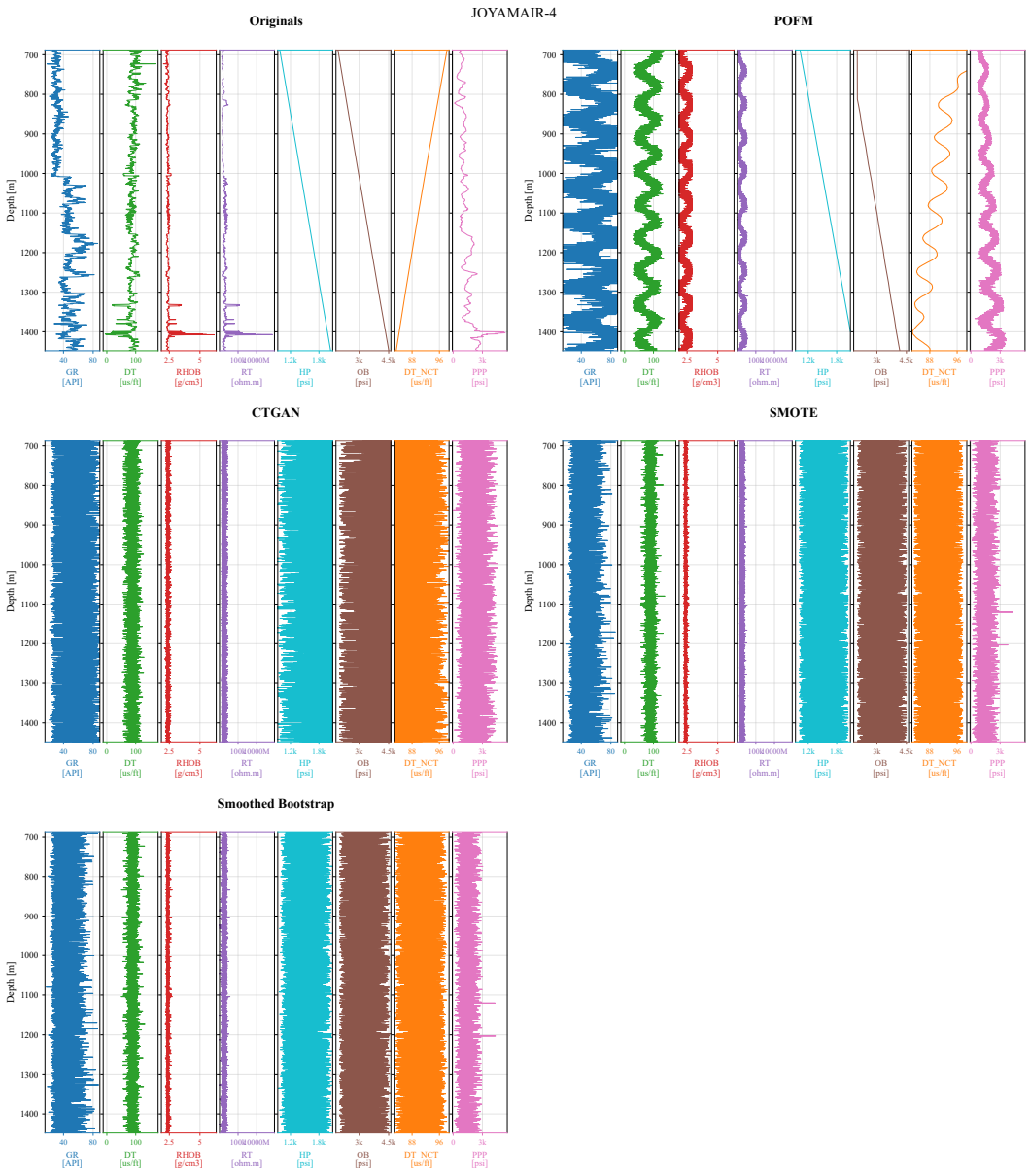

Figure 5: JOYAMAIR-4: real, POFM, CTGAN, SMOTE, and Smoothed Bootstrap logs.

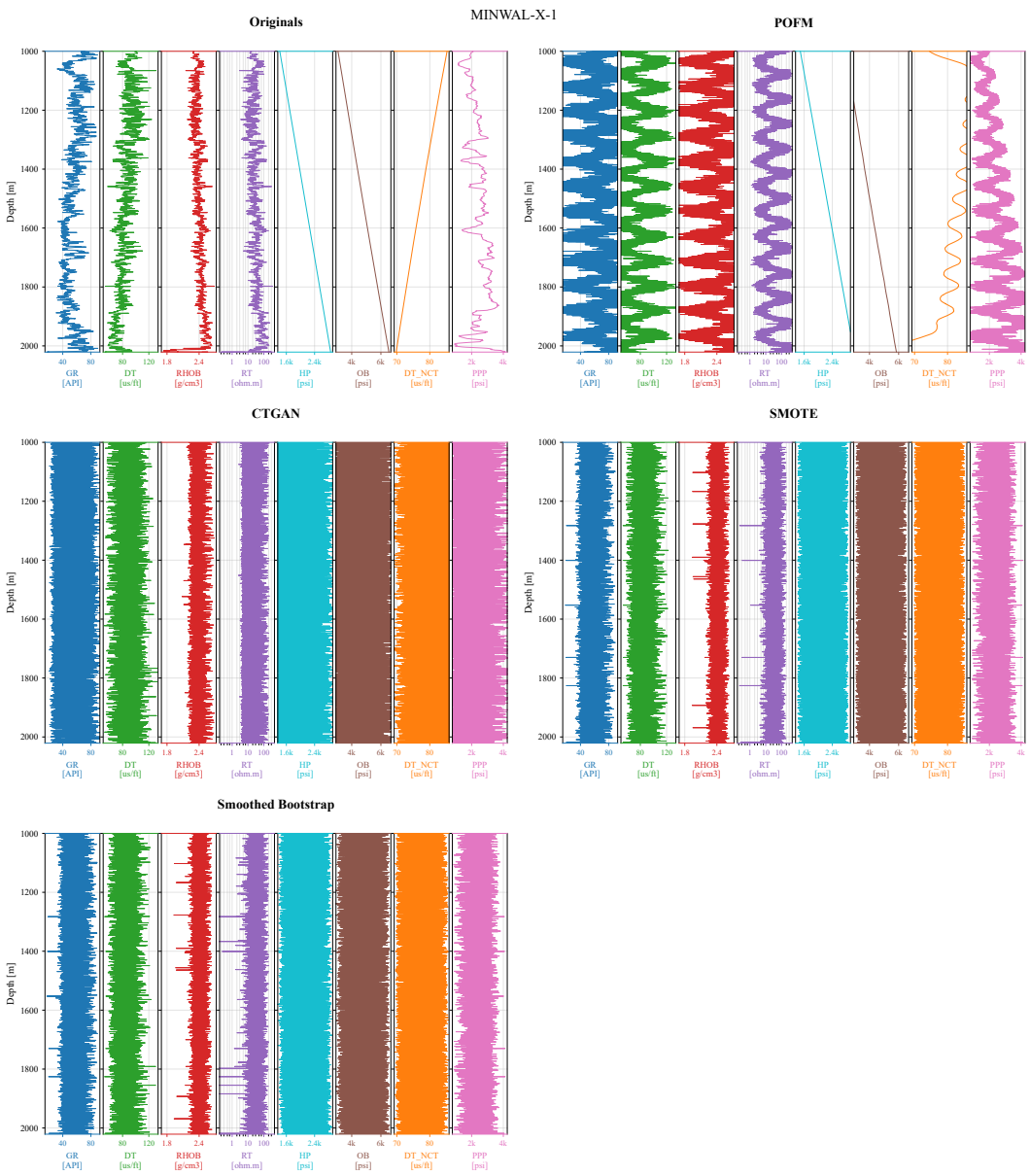

Figure 6: MINWAL-X-1: real, POFM, CTGAN, SMOTE, and Smoothed Bootstrap logs.

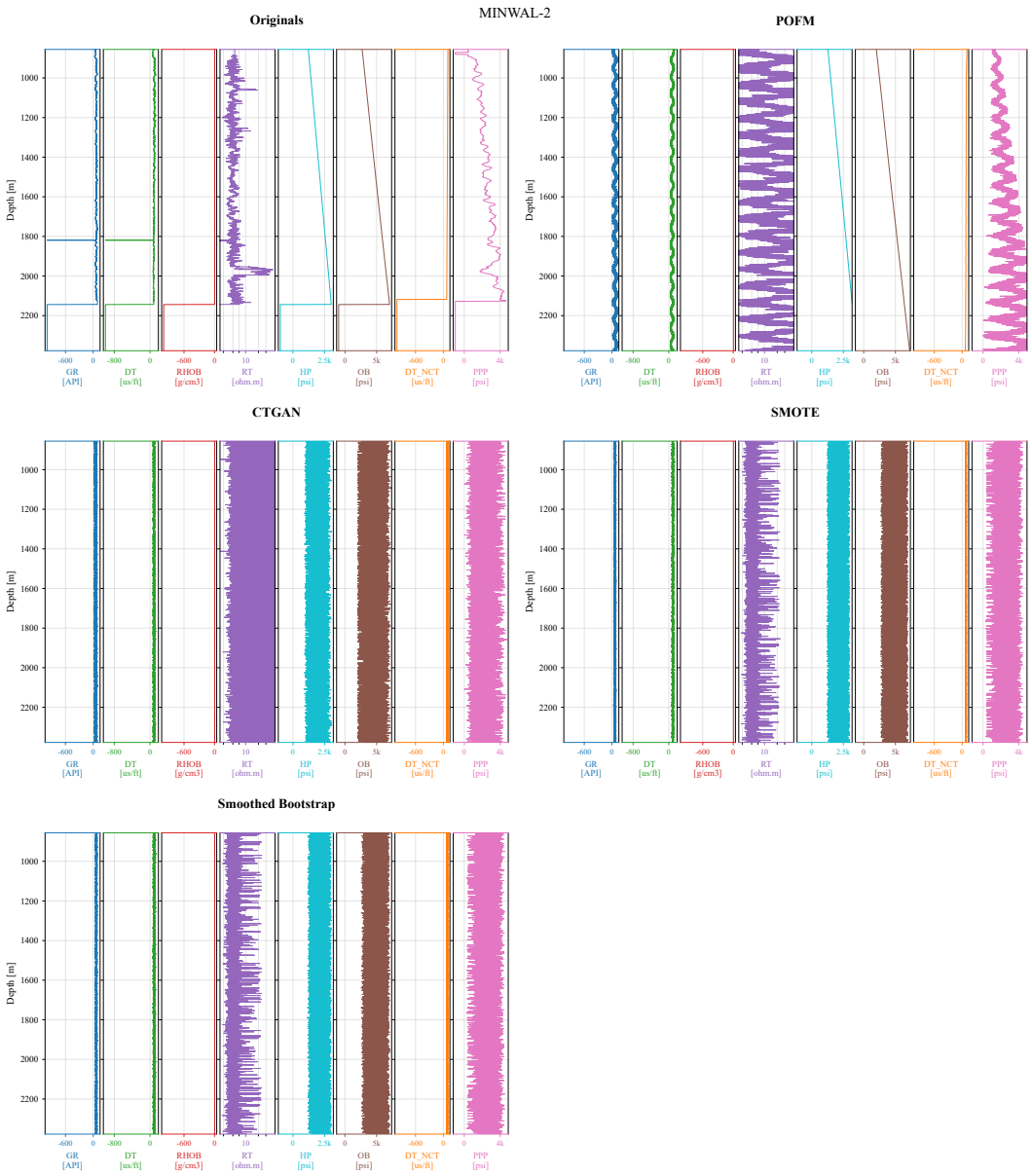

Figure 7: MINWAL-2: real, POFM, CTGAN, SMOTE, and Smoothed Bootstrap logs.

For MINWAL-2 (Figure 7), CTGAN's logs are dense vertical bands with no depth-dependent structure, reflecting the GAN's tendency to approximate the empirical marginal distribution; richer conditioning would not remove this averaging, as it stems from mode-seeking rather than missing conditioning inputs. POFM instead produces depth-varying logs across features. CTGAN stays close to the training distribution while POFM is plausible but more dispersed, so CTGAN's relative advantage over POFM arises from a closer distribution approximation, while POFM's broader sampling loses predictive alignment with the held-out well. Many CTGAN per-variable JSD values fall below the real-to-real baseline—impossible for a genuinely novel generator, but expected for one copying training wells.