

Physics-Informed Neural Networks for Porosity Prediction and Uncertainty Quantification in Niger Delta Wells with Missing Sonic Logs

^{*1}Babalola Oludare Olukayode, ²Ogunleye Odunayo Christiana, ¹Adebo Babatunde A.,
¹Adedokun Damilola Rukayat, ¹Oloruntola Racheal Foluke and ³Oduah Oluwaseun Adewunmi

¹Department of Physics, Lead City University, Ibadan, Oyo State, Nigeria.

²Department of Geology, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria.

³Federal College of Forestry, Ibadan, Oyo State, Nigeria.

*Corresponding authors' email: darekay35@gmail.com

ABSTRACT

Effective porosity is a key petrophysical parameter, yet a well without a measured sonic (DT) curve cannot support sonic porosity calculation or seismic-to-well ties. This study presents a physics-informed neural network (PINN) that jointly synthesizes a DT log and predicts porosity for three Niger Delta wells, none of which carries a measured sonic curve. A shared-trunk, dual-head architecture produces both outputs from one representation. Gardner's equation, the Wyllie time-average equation, and a density-porosity cross-check act as physics terms in the loss function. Aleatoric uncertainty comes from a heteroscedastic likelihood and epistemic uncertainty from a five-member deep ensemble with Monte Carlo dropout, which together give a predictive uncertainty band for every porosity value. These bands are Gaussian-approximation predictive intervals, not confidence intervals, and they have not been recalibrated. Under leave-one-well-out cross-validation against a no-physics ablation and a Random Forest baseline, the PINN outperforms the no-physics ablation on every held-out well ($R^2 = 0.938, 0.682$, and 0.806 on Jay_01, Jay_02, and Jay_03, versus $0.881, 0.592$, and 0.691) and is competitive with the Random Forest baseline, trailing narrowly on Jay_01 (0.938 vs 0.943), leading narrowly on Jay_02 (0.682 vs 0.678), and leading clearly on Jay_03 (0.806 vs 0.753). No measured sonic log exists for validation, so the accuracy of the synthetic DT curves remains unverified.

Received: 23 Sep. 2026

Accepted: 08 Oct. 2026

Published: 09 Oct. 2026

Keywords:

Physics-informed neural network, Porosity, Sonic logs, Aleatoric uncertainty, Epistemic uncertainty

INTRODUCTION

Background to the Study

Porosity is a critical reservoir property on which volumetric reserve estimation depends, and deriving accurate values from wireline logs has long been a central concern in petrophysics (Ahmed & Farman, 2023). Traditional methods rely on deterministic transforms, namely density porosity, neutron porosity, and sonic porosity from the Wyllie time-average equation, each calibrated against core samples (Wyllie et al., 1956). Machine learning has since entered petroleum geoscience: random forests, support vector regression, and gradient boosting applied to raw logs usually outperform single-equation transforms under cross-validation (Xu et al., 2022). These data-driven approaches sacrifice physical interpretability, and they extrapolate poorly when a well or facies departs from the training data (Isah et al., 2025). A second, practical problem is the reconstruction of missing curves. Technical or cost constraints have left many wells without a sonic log, which rules out sonic-porosity workflows and hampers seismic-to-well ties (Pham et al., 2020). This raises a natural question: can a model use the available log curves to generate a synthetic sonic response, check its own porosity estimate for physical consistency, and report the uncertainty of that estimate at the same time?

Review of Related Work

Physics-informed neural networks work by folding governing differential or algebraic equations into the training loss as soft constraints (Raissi et al., 2019), which lets a network learn

from less data while staying consistent with known physics. In petrophysics specifically, this idea has mostly shown up in permeability and porosity prediction through Kozeny-Carman-type relations, and in seismic or reservoir-simulation surrogate modelling. Sonic log synthesis has followed its own, separate track: empirical regressions such as Gardner's equation (Gardner et al., 1974) and Faust's equation (Faust, 1953), and more recently recurrent or convolutional networks trained across multiple wells (Pham et al., 2020). What remains largely absent from that second line of work is an explicit physics constraint that ties the synthesized curve back to whatever petrophysical product comes after it.

Uncertainty quantification in well-log machine learning has leaned mostly on ensemble variance or quantile regression (Singh Sandhu, 2026). The heteroscedastic-plus-epistemic decomposition that is now fairly standard in Bayesian deep learning (Kendall & Gal, 2017) shows up far less in petrophysics, which is a little surprising given how naturally the two ideas map onto the problem. Logging-tool measurement noise is aleatoric by nature, and the risk of extrapolating to an unfamiliar well or facies is epistemic.

Aim and Contributions of the Study

This study addresses that question with a physics-informed neural network that treats synthetic DT generation and porosity prediction as two coupled outputs of one shared representation, regularized by two classical relations, Gardner's equation (Gardner et al., 1974) and the Wyllie time-average equation (Wyllie et al., 1956), and cross-checked

against the density-porosity transform. On top of that, the model carries a full uncertainty decomposition: aleatoric uncertainty through a heteroscedastic likelihood, and epistemic uncertainty through a deep ensemble combined with Monte Carlo dropout (Kendall & Gal, 2017). Each porosity prediction therefore comes with an explicit predictive uncertainty band rather than a bare point estimate. This work makes three contributions. First, a PINN architecture that treats the synthetic sonic log as an intermediate, physically constrained representation rather than an end product in itself, coupled to the porosity estimate through a shared trunk and the physics terms in the loss function rather than by feeding the synthesized DT value directly into the porosity head. Second, a full aleatoric-plus-epistemic uncertainty scheme built directly into physics-constrained training. Third, a transparent evaluation against a no-physics ablation and a Random Forest baseline, all run under the identical leave-one-well-out protocol, that reports where the physics constraints help and where they do not, a comparison the applied PINN-petrophysics literature tends to underreport.

MATERIALS AND METHODS

Study Area and Data

The dataset draws on three wells, Jay_01, Jay_02, and Jay_03, from an anonymized field in the Niger Delta Basin, Nigeria, referred to throughout as the "Jay Field" to preserve operator confidentiality. This is a standard convention in Niger Delta well log studies. The basin's petroleum system is Tertiary in age and built around three formations: the Akata (shale), the Agbada (alternating sandstone and shale), and the

Benin (continental sandstone) (Diab et al., 2023) as shown in figure 1. Hydrocarbons here sit almost entirely in stacked deltaic sandstone bodies within the Agbada Formation: point-bar, barrier-bar, and channel or slope-fan sands (Olowoyo, 2010). It's that siliciclastic sandstone-shale setting, rather than a carbonate one, that explains why the physics terms in Section 2.2 use sandstone-specific matrix and fluid constants (Lawson-Jack et al., 2018).

Each well has gamma ray (GR, API units), neutron porosity (NPHI, v/v), deep resistivity (ohm·m), bulk density (RHOB, g/cc), and effective porosity (PHIE, v/v) curves sampled at 0.5 ft. None of the three carries a sonic (DT) curve at all, which is exactly the gap this pipeline is built to close. The interpreted target throughout is effective porosity (PHIE) as delivered in the source log set; no independent re-picking of porosity cut-offs was performed for this study. Table 1 lists the approximate measured-depth interval of each well, all within the Agbada Formation's stacked deltaic sandstone-shale sequence described above; no formal facies log was available to further subdivide these intervals, which we note as a limitation in Section 4.1. Quality control was limited to two steps: rows missing a value in any of the four predictor curves or in the porosity target were dropped, and resistivity was log10-transformed before modelling to correct its strongly right-skewed distribution, a routine step in petrophysical preprocessing. No additional despiking, environmental correction, or core-to-log calibration was applied, as these were not available for the anonymized dataset. Table 1 gives the row counts for each well before and after this screening; rows with missing values were removed rather than imputed.

Table 1: Approximate Depth Interval, Sampling and Row Counts of Each Well Before and After Quality Control (Curves Sampled at 0.5 Ft; Depths are Measured Depth)

Well	Approx. depth interval (ft)	Rows before QC	Rows after QC	Rows removed
Jay_01	6,750–7,450	1,461	1,380	81
Jay_02	6,200–6,950	1,461	1,461	0
Jay_03	6,650–7,300	1,351	1,268	83
Total	—	4,273	4,109	164

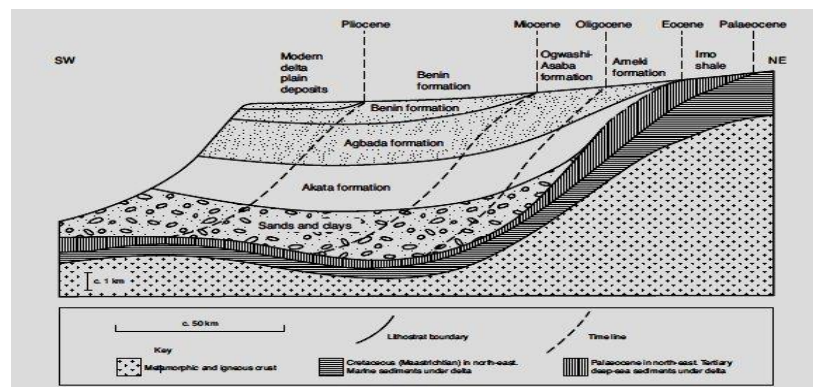


Figure 1: Lithostratigraphic Description of the Niger Delta Basin Showing the Stratigraphic Equivalences between the Outcropping and the Subsurface Niger Delta (after Wright et al., 1985)

Physics Formulation

Three classical petrophysical relations keep the network's outputs anchored to known physics. Gardner's equation (Gardner et al., 1974) relates bulk density to compressional velocity:

$$\rho = aV^b \quad (1)$$

Where $a = 0.23$ and $b = 0.25$ for the clean-sandstone case, with V in ft/s and ρ in g/cc, the customary English-unit form

of Gardner's coefficient (the commonly quoted $a = 0.31$ applies when V is in m/s; the implementation and all reported results use $a = 0.23$). Inverting Eq. 1 gives $V(\rho) = (\rho/a)^{1/b}$ in ft/s, from which the expected transit time (μ s/ft) follows for the measured bulk density log:

$$\Delta t_{\text{Gardner}} = \frac{10^6}{V(\rho)} \quad (2)$$

The Wyllie time-average equation (Wyllie et al., 1956) ties the network's synthesized DT back to its own porosity output:

$$\Delta t = \phi \Delta t_f + (1 - \phi) \Delta t_{ma} \quad (3)$$

In Eq. 3, Δt is the transit time synthesized by the network ($\mu\text{s}/\text{ft}$), ϕ is the porosity output of the network (v/v), and Δt_f and Δt_{ma} are the fluid and matrix transit times ($\mu\text{s}/\text{ft}$). Matrix and fluid constants come from standard clean-sandstone defaults rather than a fit to this particular dataset ($\Delta t_{ma} = 55.5 \mu\text{s}/\text{ft}$, $\Delta t_f = 189 \mu\text{s}/\text{ft}$; $\rho_{\text{matrix}} = 2.65 \text{ g/cc}$, $\rho_{\text{fluid}} = 1.00 \text{ g/cc}$; Table 2), chosen to match the Agbada Formation's deltaic sandstone lithology described in Section 2.1. No shale correction was applied to the Wyllie or density-porosity terms, which is a real simplification given how much the

Agbada Formation alternates between sand and shale. We return to this in Section 4.1. With $a = 0.23$ (Eq. 1), the Gardner-implied transit time for a bulk density of 2.30 g/cc is $100 \mu\text{s}/\text{ft}$, consistent with the synthesized DT curves reported in Figures 2–4, which range from approximately 60 to $130 \mu\text{s}/\text{ft}$ across the Agbada Formation sand-shale sequence. The density-porosity transform serves as an independent cross-check on the porosity head:

$$\phi_{\text{density}} = \frac{\rho_{\text{matrix}} - \rho}{\rho_{\text{matrix}} - \rho_{\text{fluid}}} \quad (4)$$

Where ρ is the measured bulk density (g/cc) and ϕ_{density} is the resulting density porosity (v/v). The transform is reported separately in Section 3 as a physics-only baseline curve.

Table 2: Physics Constants used in Eqs. 1–4 (Clean-Sandstone Literature Defaults; No Shale Correction Applied)

Symbol	Quantity	Value	Units
a	Gardner coefficient (English-unit form)	0.23	g/cc , with V in ft/s
b	Gardner exponent	0.25	dimensionless
Δt_{ma}	Matrix transit time	55.5	$\mu\text{s}/\text{ft}$
Δt_f	Fluid transit time	189	$\mu\text{s}/\text{ft}$
ρ_{matrix}	Matrix density	2.65	g/cc
ρ_{fluid}	Fluid density	1.00	g/cc

Network Architecture

Architecturally, this is a shared-trunk, dual-head multilayer perceptron. The four normalized predictor curves, gamma ray, neutron, log-resistivity, and bulk density, go in as inputs. The trunk runs three fully connected layers (128, 128, 64 units, ReLU activations), followed by dropout at a rate of 0.2, which stays active at inference time so it can support Monte Carlo sampling. From that shared 64-unit representation, two heads branch off: a DT head (32-unit hidden layer, one linear output for the synthetic transit time) and a porosity head (32-unit hidden layer, two linear outputs feeding the mean and log-variance of a heteroscedastic Gaussian likelihood).

The model was implemented directly in NumPy, with forward and backward passes derived by hand rather than relying on a deep learning framework, and trained with Adam (learning rate 3×10^{-3}) over 600 full-batch epochs.

Loss Function

The training objective combines a supervised heteroscedastic negative log-likelihood on porosity with three physics residuals plus a boundary penalty:

$$L = L_{\text{NLL}}(\phi_{\text{true}}, \phi_{\text{mean}}, \phi_{\text{logvar}}) + \lambda_1 \cdot L_{\text{Gardner}} + \lambda_2 \cdot L_{\text{Wyllie}} + \lambda_3 \cdot L_{\text{density}} + \lambda_4 \cdot L_{\text{boundary}} \quad (5)$$

In Eq. 5, L_{NLL} is the heteroscedastic Gaussian negative log-likelihood of the porosity target; L_{Gardner} penalizes the departure of the synthesized DT from the Gardner-implied transit time (Eq. 2); L_{Wyllie} penalizes the departure of the synthesized DT from the value implied by the porosity output through Eq. 3; L_{density} penalizes the departure of the porosity output from the density porosity (Eq. 4); and L_{boundary} penalizes a predictive mean outside $[0, 1]$. In the main experiments, $\lambda_1 = \lambda_2 = \lambda_3 = 0.3$ and $\lambda_4 = 0.05$, flat defaults held fixed across folds rather than tuned per well. The Gardner and Wyllie residuals are computed in a normalized DT space, z-scored using training-fold statistics of the Gardner-implied transit time, so their gradient magnitude sits on roughly the same scale as the porosity terms, which live natively in $[0, 1]$.

Uncertainty Quantification

At inference, predictive uncertainty is split into its aleatoric and epistemic parts. Each of five independently initialized ensemble members draws 20 stochastic forward passes with dropout switched on (Monte Carlo dropout), giving 100

samples of $(\phi_{\text{mean}}, \phi_{\text{logvar}})$ at every depth point. Aleatoric variance is the sample mean of $\exp(\phi_{\text{logvar}})$; epistemic variance is the sample variance of ϕ_{mean} across those 100 draws; total predictive variance is simply the sum of the two, and the porosity estimate reported is the sample mean of ϕ_{mean} . The reported 95% band is the predictive mean $\pm 1.96 \times$ the square root of this total variance, the usual Gaussian approximation; because the mean and variance are unbounded, the resulting band can extend below 0 or above 1 in physically inadmissible porosity territory even though the boundary penalty (λ_4 in Eq. 5) discourages the predictive mean itself from leaving $[0, 1]$ — the penalty constrains the mean, not the width of the Gaussian tails, which is why Figures 2–4 show bands dipping slightly negative. Table 4 reports, for each held-out well, the empirical coverage of the nominal 95% band, that is, the fraction of reference porosity values that fall inside it. These intervals remain the model's internal, decomposed uncertainty estimate: they are Gaussian-approximation predictive intervals rather than confidence intervals, and no recalibration was applied. We flag this distinction here, in the Abstract, and in Section 4.1, and note that a bounded predictive distribution (e.g. a logit-normal or truncated Gaussian likelihood) is a natural fix we have not yet implemented.

Experimental Setup

Leave-one-well-out (LOWO) cross-validation was used across the three wells: each fold trains on two wells and holds the third out entirely for testing, rotating through all three combinations. Feature normalization statistics and the Gardner-DT normalization statistics come from the training fold only, so no test-set information leaks in. Two comparison models ran under the same LOWO protocol: a no-physics ablation, where λ_1 through λ_4 are all zeroed out and the network reduces to a purely data-driven heteroscedastic-ensemble MLP with an unconstrained auxiliary DT output, and a Random Forest regressor (300 trees, max depth 12) trained directly on the four predictor curves, a classical machine-learning baseline in line with prior work on this field. Network sizes, dropout rate, learning rate, epoch count, ensemble size, and the λ weights in Eq. 5 were fixed from preliminary experiments on the two training wells of each fold only, never on the held-out well, and then held constant

across all folds and both the physics-informed and no-physics models; they were not re-tuned per fold. Each fold-model combination was trained with a single random seed rather than averaged over repeated seeds, so the RMSE and R^2 values in Table 3 do not carry a seed-to-seed variance estimate; repeated-seed training is identified in Section 4.1 as a needed extension before the fold-by-fold differences reported here can be treated as statistically robust rather than directional. Note on reproducibility: seeds are now assigned deterministically per fold and ensemble member, so the

values in Tables 3 and 4 are exactly reproducible from the released code; an earlier internal version of this pipeline drew seeds non-deterministically, which produced different per-well numbers across separate runs before this was identified and corrected.

RESULTS AND DISCUSSION

Table 3 reports porosity RMSE and R^2 for all three models across the three LOWO folds.

Table 3: LOWO Cross-Validation Results: Physics-Informed PINN, No-Physics Ablation, and Random Forest Baseline

Held-out well	PINN RMSE	PINN R^2	No-physics RMSE	No-physics R^2	RF RMSE	RF R^2
Jay_01	0.0277	0.938	0.0384	0.881	0.0265	0.943
Jay_02	0.0516	0.682	0.0583	0.592	0.0519	0.678
Jay_03	0.0519	0.806	0.0655	0.691	0.0586	0.753

The clearest signal in Table 3 is the comparison against the no-physics ablation: the physics-informed model has lower RMSE and higher R^2 than its own physics-stripped twin on all three held-out wells (Jay_01: R^2 0.938 vs 0.881; Jay_02: 0.682 vs 0.592; Jay_03: 0.806 vs 0.691), the same architecture and training procedure in both cases, differing only in whether Eqs. 2–4 contribute to the loss. That consistency is notable given that only a single random seed was used per fold-model combination (Section 2.6); it is not proof that the physics terms will help on a fourth well, but it is a more encouraging signal than a result that flipped sign between folds would have been. Against the Random Forest baseline the picture is closer: Random Forest is marginally ahead on Jay_01 (R^2 0.943 vs 0.938, RMSE 0.0265 vs 0.0277), the physics-informed model is marginally ahead on Jay_02 (0.682 vs 0.678), and clearly ahead on Jay_03 (0.806 vs 0.753, RMSE 0.0519 vs 0.0586). We read this as the physics-informed model performing on par with, and on one well meaningfully better than, a strong classical baseline, while additionally producing a synthesized DT curve and a

decomposed uncertainty estimate that Random Forest does not provide. We stop short of calling this a general result: three folds and a single seed per fold is a narrow evidentiary base, and Section 4.1 identifies repeated-seed training as the next step needed before these differences can be treated as more than directional.

Figures 2 through 4 show, for each held-out well, the true versus predicted porosity log alongside its 95% total-uncertainty band, the density-porosity physics-only baseline for reference, and the corresponding synthesized DT curve.

The uncertainty bands in Figures 2–4 are broadly similar in width across the three wells, as Table 4 confirms: mean total half-width ranges only from 0.072 to 0.079 v/v. What varies far more is how well that nominal 95% band is calibrated. Empirical coverage runs from 67.4% on Jay_03, well under the nominal rate, to 98.0% on Jay_01, somewhat over it, with Jay_02 close to nominal at 85.1%. Table 4 gives, for each held-out well, the mean total-uncertainty half-width in porosity units, the aleatoric and epistemic variance components, and this empirical coverage figure.

Table 4: Total Predictive-Uncertainty Half-Width, Aleatoric and Epistemic Variance Components, And Empirical Coverage of the Nominal 95% Band for Each Held-Out Well

Held-out well	Mean total half-width (v/v)	Mean aleatoric variance (v/v) ²	Mean epistemic variance (v/v) ²	Empirical coverage of 95% band (%)
Jay_01	0.078	0.00122	0.00042	98.0
Jay_02	0.079	0.00136	0.00036	85.1
Jay_03	0.072	0.00127	0.00026	67.4

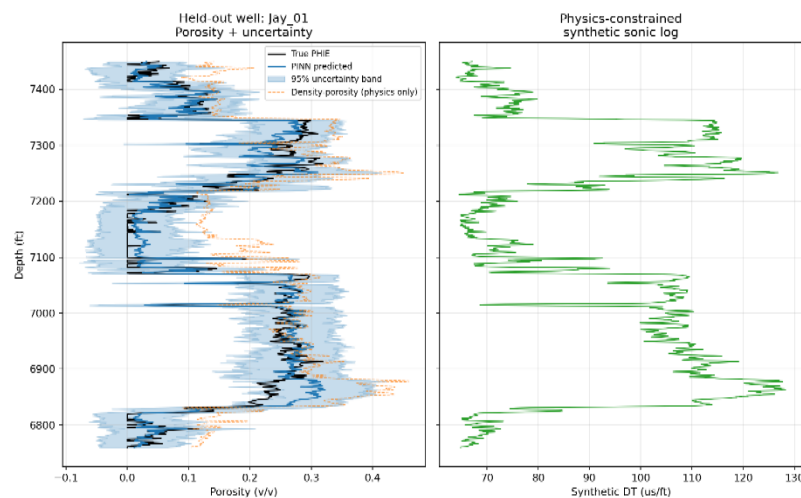


Figure 2: Jay_01 held out: predicted porosity with 95% uncertainty band (left) and synthesized sonic log (right)

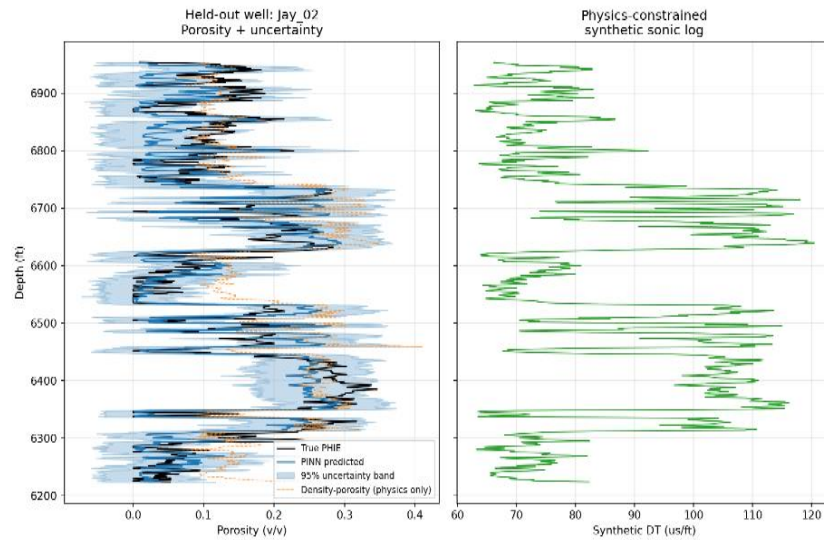


Figure 3: Jay_02 held out: predicted porosity with 95% uncertainty band (left) and synthesized sonic log (right)

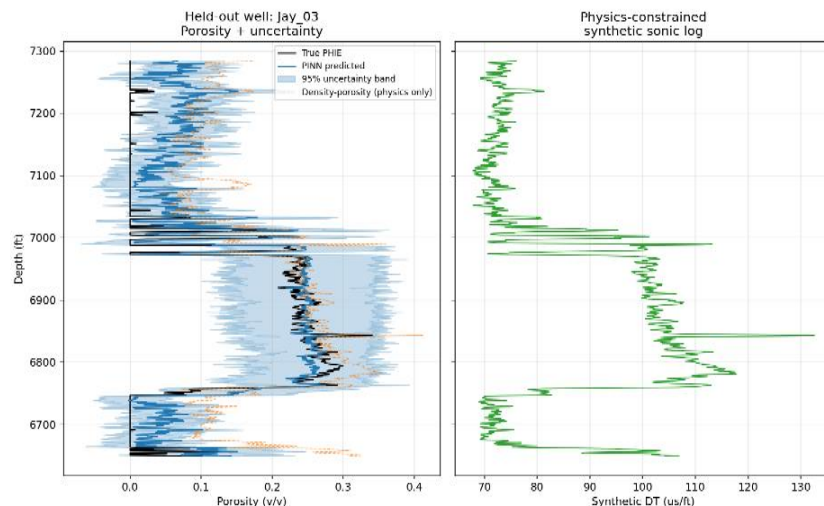


Figure 4: Jay_03 held out: predicted porosity with 95% uncertainty band (left) and synthesized sonic log (right).

Discussion

The central finding here is that physics-informed constraints built on flat, literature-default sandstone constants improved held-out porosity accuracy relative to the no-physics ablation on all three wells, and matched or exceeded a Random Forest baseline on two of three. That is a more encouraging result than an earlier, non-deterministic version of this experiment suggested during internal testing (see the Note on reproducibility in Section 2.6), and we are cautious about reading too much into it for the reason given there: a single training seed per fold-model combination means these differences are directional rather than statistically established. Two things are worth separating. First, the improvement over the no-physics ablation is architecture-controlled, the same network, the same training procedure, differing only in whether the physics residuals contribute to the loss, which makes it the cleaner of the two comparisons in Table 3. Second, the comparison against Random Forest is weaker evidence either way, since Random Forest is a structurally different model family; a narrow win or loss there is less diagnostic of what the physics terms specifically contributed.

We flag repeated-seed training, named as a limitation in Section 4.1, as the next step that would let us say with more confidence whether the per-well pattern in Table 3 reflects the physics terms themselves or the particular initialization each fold happened to draw.

Setting the accuracy comparison aside, the physics-informed model offers two things the Random Forest baseline cannot provide: a synthesized sonic curve that stays internally consistent with the measured density log, useful wherever no sonic tool was run, say for seismic time-depth ties or sonic-derived elastic properties, and a decomposed aleatoric and epistemic uncertainty band on every porosity estimate rather than a bare point prediction. The no-physics ablation also outputs an auxiliary DT curve and an uncertainty band, but nothing ties that curve to the density log or to the porosity output. These claims concern internal consistency only. None of the three wells has a measured sonic log, so the accuracy of the synthesized DT cannot be tested directly, and the curves should not be treated as substitutes for logged data. Together with the accuracy pattern documented above, these

two capabilities look like the main contribution of this framework for fields working with incomplete log suites. The 95% bands in Figures 2–4 occasionally dip below zero porosity, which is not physically admissible. As described in Section 2.5, the boundary term (λ) in the loss constrains the predictive mean toward [0,1]; it does not truncate the Gaussian tails used to build the 95% interval, so a mean close to zero with non-trivial predictive variance still produces a band that crosses zero. Clipping the band at zero after the fact would misrepresent its coverage, since the underlying Gaussian assumption would no longer match the truncated interval actually reported; any clipping would need to be paired with a recalibration step we have not carried out. A cleaner fix, left for future work, is to replace the heteroscedastic Gaussian likelihood with a distribution that is bounded on [0,1] by construction, such as a logit-normal or beta likelihood, so the predictive interval respects the physical bound without post hoc truncation.

Limitations of the Study

Three wells from one anonymized field and a single depositional system, the Agbada Formation's deltaic sandstone-shale sequence, is a narrow base, and it limits how far both the accuracy comparison and the LOWO protocol itself can generalize. Three folds give a directional signal, not a statistically powered claim, and the fact that the comparison against Random Forest is not uniformly one-sided (Section 3.1) makes clear that a larger well set is needed before drawing any general conclusion about physics-informed accuracy gains in this setting. Each result also comes from a single random seed, so the differences between models carry no estimate of training variability; repeated-seed training is needed before those differences can be treated as statistically robust. The physics constants (Gardner, Wyllie, density-porosity) came from standard clean-sandstone literature defaults rather than a calibration against core or formation-specific data, and no shale-volume correction went into the Wyllie or density-porosity terms despite the Agbada Formation's shaly-sand character. A Vshale-corrected version of these equations is a natural next step, and probably an important one. The synthetic DT output also has no ground-truth sonic log to validate against directly; its quality is inferred only through internal consistency with Gardner's equation and the density log, never confirmed against a held-out true sonic curve. The 95% uncertainty bands are Gaussian-approximation predictive intervals that have not been recalibrated, and their lower limits can fall below zero porosity; a bounded likelihood is a natural remedy. Finally, the physics loss weights (λ) stayed fixed at flat defaults rather than being tuned or annealed during training. A weight-annealing schedule, or adaptive weighting based on gradient norms, is a reasonable next step.

CONCLUSION

This study built a physics-informed neural network that synthesizes a sonic log, predicts porosity, and quantifies aleatoric-plus-epistemic uncertainty, all from a reduced well log suite that lacks a sonic curve, and evaluated it with leave-one-well-out cross-validation on three wells from an anonymized Niger Delta field in the Agbada Formation. Against a no-physics ablation under the same protocol, the physics-informed model improved porosity accuracy on every held-out well; against a Random Forest baseline it was closely competitive, trailing narrowly on one well and leading on the other two. That pattern, reported transparently alongside its single-seed, three-well limitations (Section 4.1), is evidence that flat, uncalibrated physics constants can be a

useful regularizer in this shaly-sand setting without being presented as a guaranteed or fully generalizable one. As it stands, the framework's practical value lies in its consistent edge over an architecturally identical data-driven model, its synthetic-log generation, and its explicit uncertainty decomposition, none of which the Random Forest baseline provides; the synthetic logs, however, remain unvalidated against measured sonic data. Shale-corrected physics terms, formation-calibrated constants, additional wells, repeated random seeds, and direct validation against measured sonic logs are the most promising directions for confirming whether this advantage holds up beyond the present three-well, single-seed evaluation.

REFERENCES

- Ahmed, R., & Farman, G. M. (2023). How to estimate the major petrophysical properties: A review. *Iraqi Journal of Oil and Gas Research (IJOGR)*, 3(1), 43–58. <https://doi.org/10.55699/ijogr.2023.0301.1037>
- Diab, A. I., Sanuade, O., & Radwan, A. E. (2023). An integrated source rock potential, sequence stratigraphy, and petroleum geology of (Agbada-Akata) sediment succession, Niger delta: Application of well logs aided by 3D seismic and basin modeling. *Journal of Petroleum Exploration and Production Technology*, 13(1), 237–257. <https://doi.org/10.1007/s13202-022-01548-4>
- Faust, L. Y. (1953). A velocity function including lithologic variation. *Geophysics*, 18(2), 271–288. <https://doi.org/10.1190/1.1437869>
- Gardner, G. H. F., Gardner, L. W., & Gregory, A. R. (1974). Formation velocity and density: The diagnostic basics for stratigraphic traps. *Geophysics*, 39(6), 770–780. <https://doi.org/10.1190/1.1440465>
- Isah, A., Tariq, Z., Mustafa, A., Mahmoud, M., & Okoroafor, E. R. (2025). A review of data-driven machine learning applications in reservoir petrophysics. *Arabian Journal for Science and Engineering*, 50(24), 20343–20377. <https://doi.org/10.1007/s13369-025-10329-0>
- Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 5574–5584). Curran Associates. <https://arxiv.org/abs/1703.04977>
- Lawson-Jack, O., Uko, E. D., Tamunobereton-Ari, I., & Alabraba, M. A. (2018). Geomechanical characterization of a reservoir in part of Niger Delta, Nigeria. *Asian Journal of Applied Science and Technology*, 3(1), 10–30.
- Olowoyo, K. O. (2010). *Structural and seismic facies interpretation of Fabi field, onshore Niger Delta, Nigeria*. Universal-Publishers.
- Pham, N., Wu, X., & Zabihi Naeini, E. (2020). Missing well log prediction using convolutional long short-term memory network. *Geophysics*, 85(4), WA159–WA171. <https://doi.org/10.1190/geo2019-0282.1>
- Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of*

- Computational Physics*, 378, 686–707. <https://doi.org/10.1016/j.jcp.2018.10.045>
- Singh Sandhu, S. K. (2026). Uncertainty quantification of well log predictions. *Interpretation*, 14(1), B31–B38.
- Wyllie, M. R. J., Gregory, A. R., & Gardner, L. W. (1956). Elastic wave velocities in heterogeneous and porous media. *Geophysics*, 21(1), 41–70. <https://doi.org/10.1190/1.1438217>
- Xu, C., Fu, L., Lin, T., Li, W., & Ma, S. (2022). Machine learning in petrophysics: Advantages and limitations. *Artificial Intelligence in Geosciences*, 3, 157–161. <https://doi.org/10.1016/j.aiig.2022.11.004>



©2026 This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license viewed via <https://creativecommons.org/licenses/by/4.0/> which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is cited appropriately.