Comparative Evaluation of Neural Surrogate Formulations for Coupled Thermo-Hydro-Mechanical Storage-Diffusion Processes in Geothermal Systems

Bromine J Poke[1], Rashel[2*]

[1]Independent Researcher

[2]Enterprises LLC, United States of America

Email: office@drrashelenterprises.com

Abstract

Neural surrogate models could support uncertainty analysis and optimization in geothermal systems, but comparisons among data-driven, physics-informed, and hybrid methods are often confounded by different data, model sizes, or test splits. This study provides a controlled comparison on a reduced, dimensionless three-dimensional thermo-hydro-mechanical problem. Pressure and temperature evolve through a coupled storage-diffusion system, and a scalar displacement potential supplies a reduced mechanical closure. Forty-eight parameter scenarios are generated per run on a $9 \times 9 \times 6$ grid at 10 times. The scenarios are partitioned into 32 training, 8 validation, and 8 testing cases. A supervised coordinate network, a physics-informed neural network, and a hybrid residual-correction network are trained with comparable multilayer perceptrons and evaluated across three random seeds. The hybrid model produces mean relative L2 errors of 0.0134 for pressure, 0.0121 for temperature, and 0.0134 for mechanical potential. It also yields the smallest normalized discrete residuals. The pure network obtains field-dependent relative errors of 0.2200 to 0.3022. Under the reported CPU budget, adding automatic-differentiation residual penalties does not improve generalization: the physics-informed model obtains errors of 0.2472 to 0.3246 and requires longer training. These results show that a useful low-fidelity physics prior can be more valuable than a residual penalty alone for this reduced benchmark.

Keywords: geothermal reservoir; thermo-hydro-mechanical coupling; neural surrogate; physics-informed neural network; hybrid modeling; reproducibility

DOI: 10.18091/ijsts.v10i2-07

1. Introduction

Thermo-hydro-mechanical (THM) interactions govern many subsurface energy processes. Fluid injection and production alter pressure, heat transport changes the thermal state, and both effects modify deformation and stress. High-fidelity THM solvers therefore combine multiple partial differential equations, material constitutive laws, spatial heterogeneity, and demanding nonlinear solution procedures. Such models are essential for physical interpretation, yet their cost can limit parameter studies, uncertainty analysis, inverse problems, and repeated optimization. Parallel finite-element developments have long addressed the computational burden of coupled porous-media simulation (Wang et al., 2009),

while recent geothermal studies have used machine-learning surrogates to accelerate simulation-assisted optimization (Jin et al., 2022).

Neural surrogates offer several routes to reduced evaluation cost. A purely data-driven model learns a mapping from coordinates and parameters to state variables. This approach is flexible, but accuracy depends on the coverage and quality of the training simulations. Physics-informed neural networks (PINNs) add differential-equation residuals and, in some formulations, boundary and initial conditions to the learning objective (Raissi et al., 2019). Physics-informed machine learning more broadly seeks to combine observations, simulations, and governing knowledge rather than treating them as separate sources of information (Karniadakis et al., 2021). Geoscience is a promising but difficult application area because multiple time scales, strongly coupled variables, non-smooth material fields, and heterogeneous boundary conditions can make optimization stiff (Degen et al., 2023). Advances in geothermal energy recovery depend on accurate characterization of thermal behavior across both subsurface reservoirs and associated heat-exchange systems. Investigations of reactive heat-exchanger configurations have emphasized the importance of thermal transport processes and heat-transfer performance in improving geothermal energy recovery (Williams, 2024). Within reservoir-scale numerical modeling, reliable prediction of pressure and temperature distributions provides complementary information for evaluating subsurface thermal behavior and developing computationally efficient simulation approaches. Recent deep-learning studies have also predicted geothermal temperature fields from geological and thermophysical attributes, demonstrating the potential of multi-attribute surrogates for hot-dry-rock characterization (Gao & Zhao, 2024).

PINNs have been applied to nonlinear diffusivity and Biot poroelastic equations (Kadeethum et al., 2020) and to coupled THM problems in porous media (Amini et al., 2022). Physics-constrained learning has also been used for three-dimensional flow in heterogeneous porous media (Yan et al., 2022). These studies demonstrate the feasibility of encoding physical structure in learning algorithms. They do not, however, imply that a residual penalty will always outperform supervised learning. A PINN objective balances field fit, several residuals, and boundary or initial conditions whose magnitudes and derivative orders can differ substantially. Its result depends on nondimensionalization, collocation sampling, weighting, optimizer schedule, and the relationship between continuous residuals and the discrete reference solver.

Hybrid residual learning provides a different route. Rather than asking a neural network to learn the complete state or satisfy all equations directly, a low-fidelity physical model supplies an approximate solution and a network learns its correction. This can reduce the complexity and dynamic range of the target. The approach is especially attractive when a simplified simulator is inexpensive and structurally informative, even if its parameters or couplings are biased.

The present study compares these three strategies under one controlled design. All learned models use compact coordinate multilayer perceptrons, the same scenario pool, and scenario-level train-validation-

test splits. The comparison includes a deliberately biased low-fidelity baseline, a supervised neural network, a PINN using automatic-differentiation residual penalties, and a hybrid residual-correction model. The reference problem is dimensionless and intentionally reduced. It contains coupled pressure and temperature diffusion plus a scalar mechanical potential, not the full vector elasticity, contact, fracture, or multiphase physics of a field-scale enhanced geothermal system.

The contributions are fourfold. First, the paper specifies an internally consistent reduced three-dimensional THM benchmark with sampled coupling and source parameters. Second, it prevents point-level leakage by holding out complete parameter scenarios. Third, it reports field errors, discrete-equation residual checks, wall-clock training times, model sizes, and variability across three seeds. Fourth, it provides executable generation, training, evaluation, testing, and figure scripts. The objective is not to prove that one learning paradigm is universally superior. It is to show, transparently, what each paradigm achieves under a modest and reproducible budget.

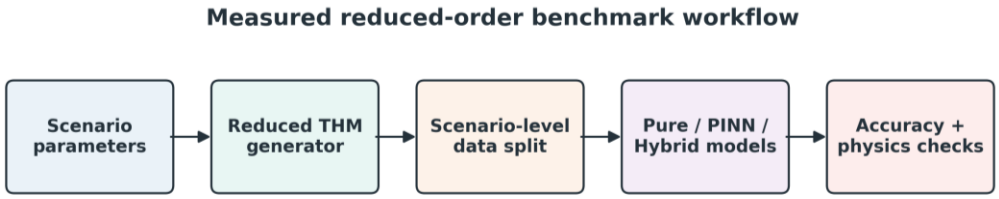

Figure 1. Reproducible benchmark workflow. Scenario parameters drive the reduced THM generator and the deliberately biased low-fidelity baseline. Data are split by complete scenarios before fitting. Pure NN, PINN, and hybrid models are evaluated with field errors and discrete residual checks.

2. Reduced THM problem

2.1 Domain, fields, and scaling

The dimensionless spatial domain is the unit cube, $\Omega = (0, 1)^{3,}$ and dimensionless time satisfies $\tau \in [0, 0.225]$. The state variables are pressure perturbation $p(x, \tau)$, temperature perturbation $\theta(x, \tau)$, and scalar mechanical potential $\psi(x, \tau)$. The variables are dimensionless. They can be interpreted as physical perturbations divided by user-selected characteristic scales p_c, T_c, and ψ_c, with $x = x/L$ *and* $\tau = t/t_c$. The numerical study operates directly in dimensionless form so that the reported ranges and errors are not tied to a particular reservoir.

The mechanical potential is a reduced closure. It represents a scalar field driven by pressure and temperature and is useful for comparing coupled surrogate strategies. It is not the displacement vector u, does not reconstruct a full strain or stress tensor, and cannot represent shear, anisotropic elasticity, faults, or contact. Biot's poroelastic framework motivates pressure-mechanics coupling in the broader problem class (Biot, 1941), but the present benchmark should not be read as a complete discretization of three-dimensional Biot thermoelasticity.

2.2 Coupled equations

The pressure-temperature subsystem is

$$a_{11}\,\partial p/\partial \tau \;+\; a_{12}\,\partial \theta/\partial \tau \;-\; D_p\,\nabla^2 p \;=\; q_p \quad (1)$$

$$a_{21}\,\partial p/\partial \tau \;+\; a_{22}\,\partial \theta/\partial \tau \;-\; D_t\,\nabla^2 \theta \;=\; q_t \quad (2)$$

The reduced mechanical closure is

$$\nabla^2 \psi \;+\; c_{mp}\,p \;-\; c_{mt}\,\theta \;=\; 0 \quad (3)$$

The storage-coupling coefficients are

$$a_{11} \;=\; 1 \;+\; c_{pm}c_{mp}, \quad a_{12} \;=\; -(c_{pt} \;+\; c_{pm}c_{mt}) \quad (4)$$

$$a_{21} \;=\; -(c_{tp} \;-\; c_{tm}c_{mp}), \quad a_{22} \;=\; 1 \;-\; c_{tm}c_{mt} \quad (5)$$

The feedback factors $c_{pm} = 0.070$ and $c_{tm} = 0.040$ are fixed. D_p and D_t are dimensionless hydraulic and thermal diffusivities. The sampled coefficients c_{pt} and c_{tp} couple the pressure and temperature storage equations, while c_{mp} and c_{mt} couple pressure and temperature to the mechanical potential. The determinant $a_{11}a_{22} - a_{12}a_{21}$ remains positive throughout the sampled range.

The source functions are Gaussian perturbations with width $\sigma = 0.11$. A pressure source is centered at $x_i = (0.25, 0.50, 0.50)$, and a pressure sink with 70% of the source magnitude is centered at $x_o = (0.75, 0.50, 0.50)$. Cooling is collocated with the pressure source:

$$q_p \;=\; A_p[exp(-\|x \;-\; x_i\|^2/(2\sigma^2)) \;-\; 0.70\,exp(-\|x \;-\; x_o\|^2/(2\sigma^2))] \quad (6)$$

$$q_t \;=\; -A_t\,exp(-\|x \;-\; x_i\|^2/(2\sigma^2)) \quad (7)$$

Homogeneous Dirichlet conditions $p = \theta = \psi = 0$ are imposed on the external boundary. The initial perturbations are zero. These conditions make the experiment self-contained and produce smooth fields that isolate coupling and surrogate behavior without introducing fracture discontinuities or complicated boundary controls.

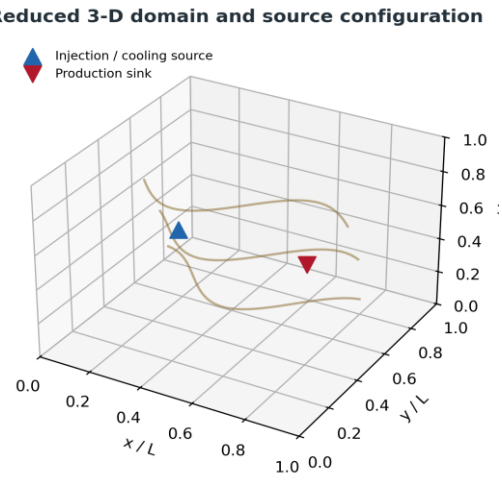

Figure 2. Reduced three-dimensional domain. Blue and red symbols identify the pressure/cooling source and pressure sink, respectively. Brown curves are conceptual flow-path guides and are not discrete fractures used by the numerical generator.

2.3 Parameter space

Eight scenario parameters are sampled with stratified uniform draws. Table 1 gives the ranges. The values define a stable reduced problem and are not calibrated geothermal properties.

Table 1. Dimensionless scenario parameters.

Parameter	Meaning	Range
D_p	Pressure diffusivity	0.015 to 0.035
D_t	Thermal diffusivity	0.006 to 0.015
c_{pt}	Thermal-to-hydraulic storage coupling	0.040 to 0.120
c_{tp}	Hydraulic-to-thermal storage coupling	0.020 to 0.070
A_p	Pressure-source amplitude	0.800 to 1.400
A_t	Cooling-source amplitude	0.400 to 0.800
c_{mp}	Pressure-mechanical coupling	0.500 to 1.000
c_{mt}	Thermal-mechanical coupling	0.100 to 0.300

2.4 Reference and low-fidelity solvers

The reference generator uses a regular $9 \times 9 \times 6$ $grid$ and 10 stored times. The dimensionless step is $\Delta\tau = 0.025$. Centered second-order finite differences approximate the Laplacians, and an explicit Euler update advances the coupled pressure-temperature system after inversion of its two by two storage matrix. At each time, a three-dimensional discrete sine transform solves the homogeneous-Dirichlet Poisson equation for ψ. The modest grid is adequate for a computationally inspectable learning benchmark, but it is not a mesh-convergence claim.

The low-fidelity solver uses the same grid and numerical operators, with intentional structural and parametric bias: D_p is multiplied by $0.80, D_t \times 1.20$, both pressure-temperature storage couplings are removed, A_p is multiplied by $0.95, A_t \times 1.05, c_{mp} \times 0.75$, and $c_{mt} \times 0.50$. Thus, the baseline retains the source geometry, boundary behavior, and diffusive structure while producing systematic field errors. The hybrid network receives the baseline fields and learns their correction.

3. Surrogate models and experimental design

3.1 Common inputs and architecture

Each record contains 12 inputs: x, y, z, τ, and the eight scenario parameters. The outputs are p, θ, and ψ. The pure and PINN models use a $12 - 64 - 64 - 64 - 3$ multilayer perceptron with sigmoid linear unit activations. Inputs and outputs are standardized using training-split statistics. Both models contain 9,347 trainable parameters.

The hybrid model appends the three low-fidelity fields to the 12 physical inputs. Its $15 - 64 - 64 - 64 - 3$ network predicts the residual between the reference and low-fidelity solutions. The physical output equals the baseline plus the learned residual. This model contains 9,539 trainable parameters. All standardization statistics are calculated from training scenarios only.

3.2 Objectives

For the pure network, the objective is the normalized mean-squared field error

$$\mathcal{L}data = (1/N_b)\, \Sigma_i \, \|S_y{}^{-1}(\hat{y}_i - y_i)\|_2{}^2 \quad (8)$$

where y = (p, θ, ψ), S_γ is the diagonal vector of training-target standard deviations, and N_β is the batch size. The hybrid model uses the same form after replacing the target with the reference-minus-baseline correction.

The PINN uses

$$\mathcal{L}PINN = \mathcal{L}data + \lambda_{phys}(\langle r_p{}^2 \rangle + \langle r_t{}^2 \rangle + \langle r_m{}^2 \rangle) \quad (9)$$

with $\lambda_{phys} = 0.01$. The residuals are obtained with automatic differentiation. The pressure and temperature residuals are normalized by A_p and A_t. The mechanical residual is normalized by the standard deviation of c_{mp} p - c_{mt} θ at the collocation batch, with a lower bound of 10^{-4}. Sixty-four interior collocation points are sampled per optimizer step from the training scenarios. Boundary and initial states also appear in the supervised training records; no separate boundary penalty is added.

3.3 Data generation and splits

Each seed produces 48 independently stratified parameter scenarios and 233,280 point records. Complete scenarios are allocated to 32 training, 8 validation, and 8 test scenarios. A held-out scenario contributes

no spatial or temporal point to training. This grouping is important because a random point-level split would expose the same parameter scenario and nearby coordinates on both sides of the split and could overstate generalization.

Each seed changes the sampled parameter set, scenario assignment, minibatch sequence, and neural initialization. Reported means and sample standard deviations summarize these three runs. They indicate sensitivity to this limited resampling procedure, not full epistemic or aleatoric uncertainty.

3.4 Optimization and software

AdamW with weight decay 10^{-6} is used for all learned models, and gradient norms are clipped at 10. Table 2 lists the measured configuration. The PINN receives a larger training budget because each step requires first and second derivatives. The code uses Python, PyTorch, NumPy, SciPy, and scikit-learn.

Table 2. Model and training configuration.

Model	Epochs	Steps per epoch	Batch size	Learning rate	Physics weight
Pure NN	55	18	1,024	0.001	0
PINN	80	12	768	0.0008	0.01
Hybrid	45	15	1,024	0.001	0

3.5 Evaluation

Field accuracy is measured with root-mean-square error, mean absolute error, coefficient of determination R^2, and relative L2 error. The principal scale-independent comparison is

$$Erel(y) = \|\hat{y} - y\|_2 / \|y\|_2 \quad (10)$$

Discrete residuals are recomputed after prediction by reconstructing each held-out scenario on the regular grid. Pressure and temperature time derivatives and Laplacians use the same finite-difference operators as the reference generator. The mechanical residual uses the same discrete Poisson operator. These checks are distinct from the PINN's automatic-differentiation training residual and therefore test consistency with the reference discretization. Wall-clock training and batched inference times are measured with Python's high-resolution performance counter. Timing applies only to the reported runtime and problem size.

4. Results

4.1 Field accuracy

Table 3 reports the three-seed results. The low-fidelity baseline is already accurate for pressure, with R^2 = 0.9989 and relative L2 error 0.0337, but its temperature and mechanical-potential errors are 0.1022 and

0.2624. This behavior is consistent with a baseline that retains correct spatial source structure while intentionally removing cross-coupling and weakening the mechanical coefficients.

The pure network obtains R^2 values of 0.9177 for pressure, 0.8947 for temperature, and 0.9488 for potential. Its relative L2 errors range from 0.2200 to 0.3022. Variability is largest for pressure and temperature because each run uses a new parameter sample and only eight held-out scenarios.

Under the stated optimization budget, the PINN does not improve these metrics. Its R^2 values are 0.9017, 0.8795, and 0.9365, and its relative errors range from 0.2472 to 0.3246. These results are not evidence that PINNs are generally inferior. They show that this residual formulation, weighting, collocation budget, network size, and optimizer schedule did not convert the added differential constraints into better held-out field prediction.

The hybrid model produces the smallest field errors for all three outputs. Mean relative L2 error is 0.0134 for pressure, 0.0121 for temperature, and 0.0134 for potential. Mean R^2 exceeds 0.9998 for every field. Relative to the pure model, the mean error decreases by approximately 95% for pressure, 96% for temperature, and 94% for potential. The small seed-to-seed standard deviations indicate that learning the structured baseline correction is substantially easier in this parameter range than learning the complete solution from coordinates and parameters.

Table 3. Field metrics across three seeds, mean ± sample standard deviation.

Model	p R^2	p rel. L2	θ R^2	θ rel. L2	ψ R^2	ψ rel. L2
Low-fidelity	0.9989 ± 0.0002	0.0337 ± 0.0026	0.9890 ± 0.0033	0.1022 ± 0.0157	0.9295 ± 0.0023	0.2624 ± 0.0040
Pure NN	0.9177 ± 0.0435	0.2797 ± 0.0767	0.8947 ± 0.0823	0.3022 ± 0.1263	0.9488 ± 0.0235	0.2200 ± 0.0491
PINN	0.9017 ± 0.0505	0.3067 ± 0.0781	0.8795 ± 0.0974	0.3246 ± 0.1301	0.9365 ± 0.0197	0.2472 ± 0.0376
Hybrid	0.9998 ± 0.0000	0.0134 ± 0.0004	0.9998 ± 0.0001	0.0121 ± 0.0048	0.9998 ± 0.0000	0.0134 ± 0.0014

Figure 3 shows the relative-error comparison, and Figure 4 displays reference, hybrid, and error slices for one held-out primary-seed scenario. The error panels use field-specific scales because ψ is smaller in magnitude than p or θ. The visual agreement should be interpreted together with the aggregate table rather than as an independent validation sample.

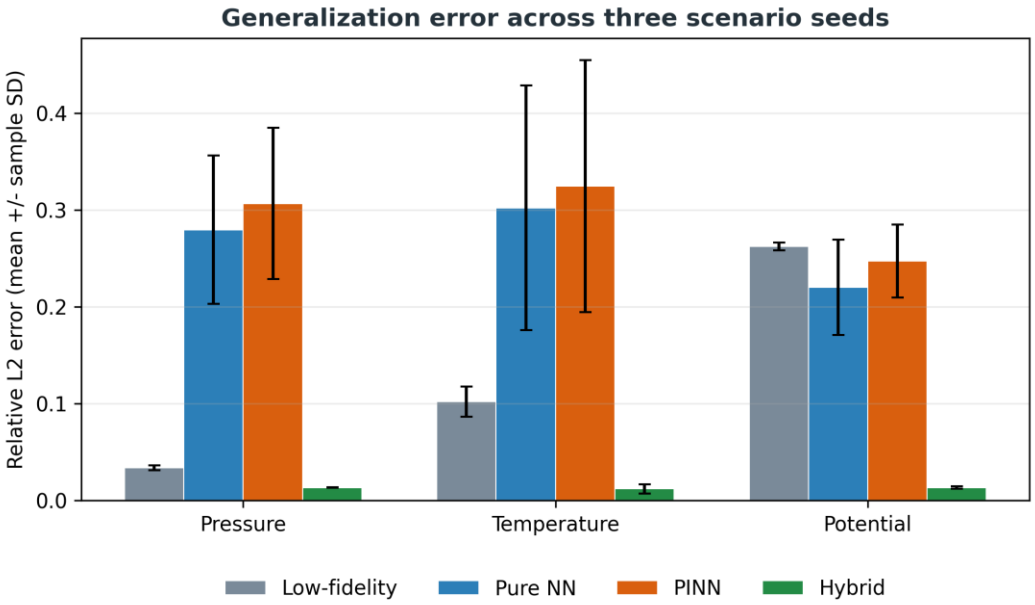

Figure 3. Relative L2 generalization error across three random seeds. Bars show the mean across seeds, and error bars show the sample standard deviation. Each seed uses a separately sampled 48-scenario dataset and scenario-level data splitting.

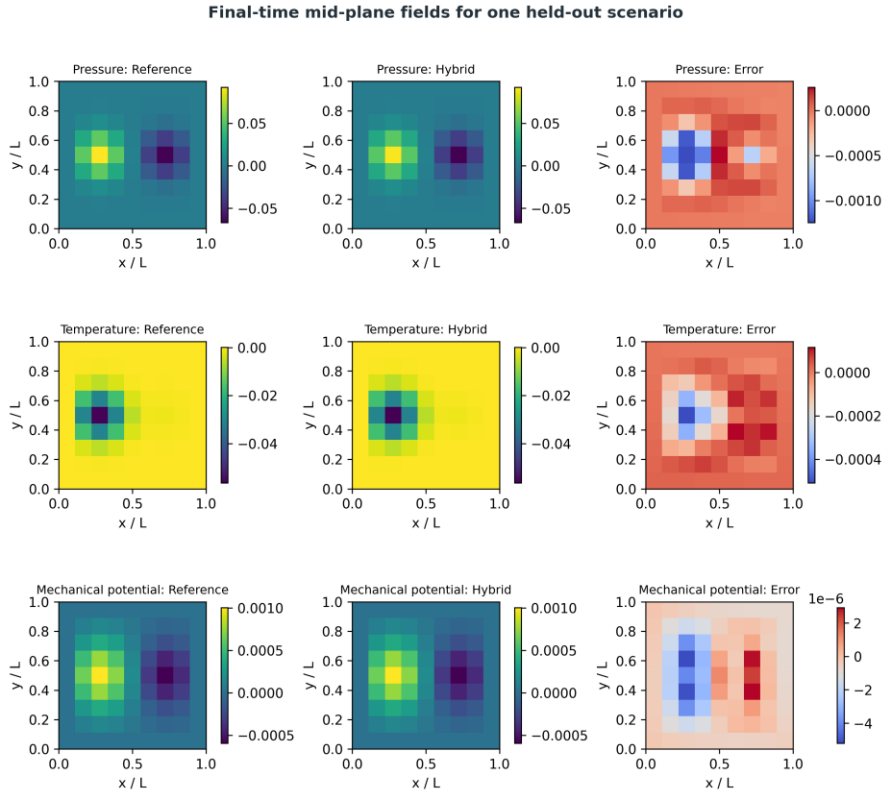

Figure 4. Reference and hybrid-predicted mid-plane fields at the final time for one held-out scenario from the primary seed. The error column uses an independent color scale for each field.

4.2 Discrete residual consistency

Table 4 gives normalized discrete residual root-mean-square values. The hybrid model has the lowest values for all equations: 0.00283 for mass, 0.00191 for energy, and 0.08470 for the mechanical closure. It also improves on the low-fidelity baseline, whose corresponding values are 0.00591, 0.01033, and 0.41991.

The pure and PINN networks yield larger discrete residuals. The PINN obtains 0.05106 for mass, 0.04539 for energy, and 0.68717 for mechanics. Its training objective uses continuous automatic-differentiation residuals at scattered collocation points, while evaluation uses finite differences on predicted grid fields. The two residual measures approach the same differential equations but are not numerically identical. A low continuous collocation loss does not guarantee low discrete residuals after gridding, particularly with limited collocation coverage and uneven output scales.

Table 4. Normalized discrete residual RMS on held-out scenarios, mean ± sample standard deviation.

Model	Mass	Energy	Mechanical closure
Low-fidelity	0.00591 ± 0.00030	0.01033 ± 0.00174	0.41991 ± 0.00895
Pure NN	0.04582 ± 0.00716	0.04067 ± 0.00704	0.66246 ± 0.22689
PINN	0.05106 ± 0.00777	0.04539 ± 0.00786	0.68717 ± 0.16692
Hybrid	0.00283 ± 0.00017	0.00191 ± 0.00044	0.08470 ± 0.02690

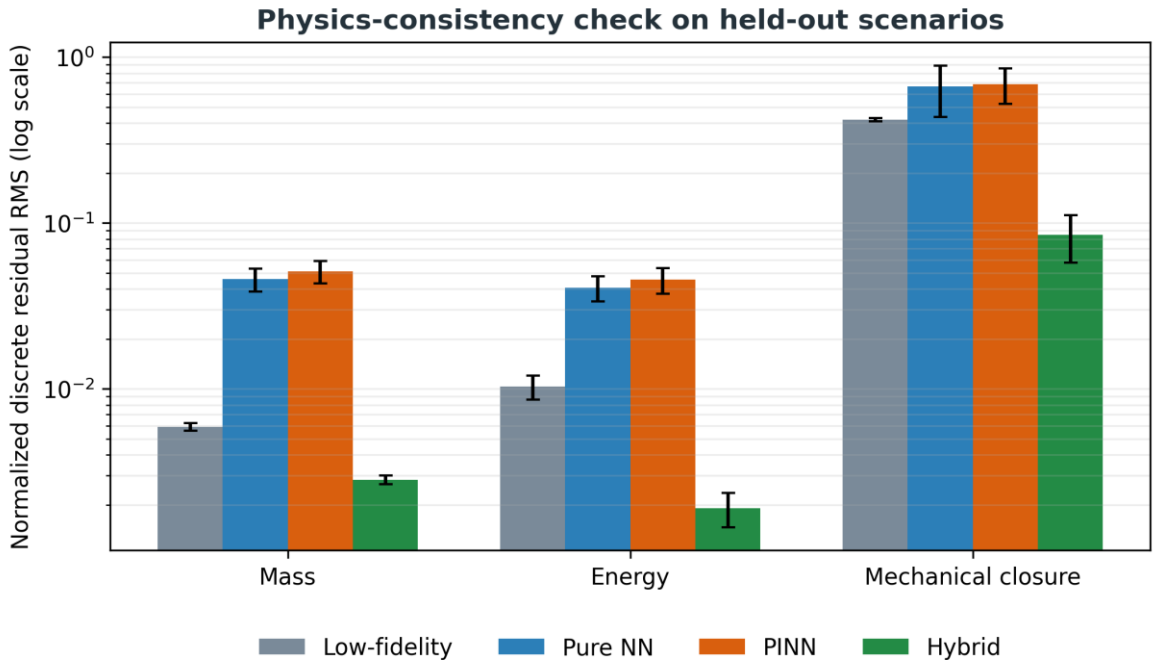

Figure 5. Normalized discrete residual root-mean-square values on held-out scenarios. Residuals are recomputed from predicted gridded fields with the same finite-difference operators used in the generator. Lower values indicate better consistency with the reduced equations.

4.3 Computational observations

Mean CPU training times are 4.72 ± 1.13 s for the pure model, 36.08 ± 3.57 s for the PINN, and 2.44 ± 0.26 s for the hybrid model. The PINN cost reflects additional epochs and first- and second-order automatic differentiation. Mean batched inference times are 0.414, 0.208, and 0.282 microseconds per point, respectively. Differences at this scale are sensitive to batch scheduling and runtime noise.

Figure 6 therefore plots accuracy against measured training cost, not operational geothermal acceleration. Figure 7 shows primary-seed training objectives. Absolute values differ because the PINN objective contains residual terms and the hybrid target is a smaller correction field.

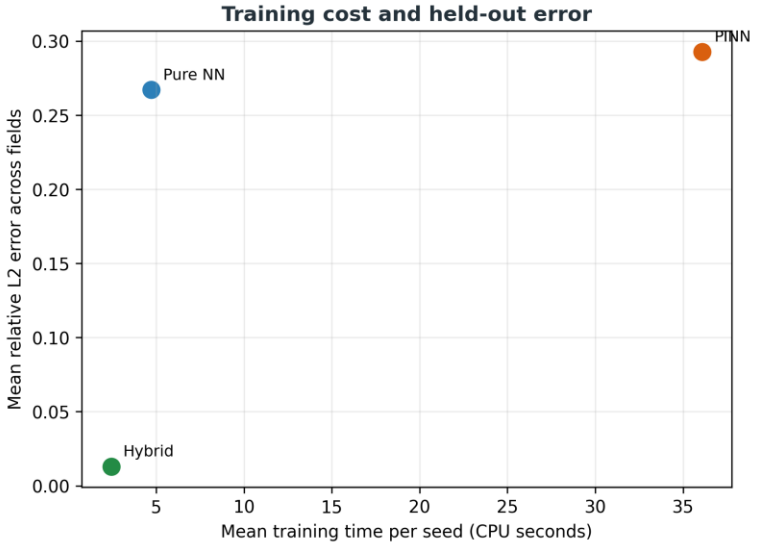

Figure 6. Mean CPU training time per seed versus the mean relative L2 error across pressure, temperature, and mechanical potential. Timing describes this small runtime and is not a field-scale speedup claim.

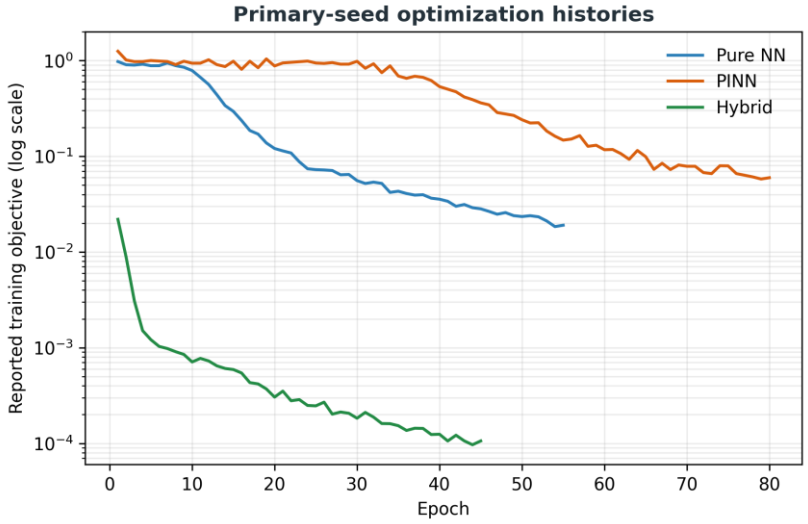

Figure 7. Primary-seed training objectives. Curves are not directly comparable in absolute magnitude because the PINN includes a residual penalty and the hybrid predicts corrections with a much smaller target scale.

5. Discussion

5.1 Why the hybrid model succeeds here

The hybrid result is consistent with target simplification. The low-fidelity solver provides the correct domain, boundary conditions, source locations, time direction, and broad diffusion patterns. Its errors arise from controlled parameter and coupling biases. The residual network therefore learns a smooth correction over an eight-dimensional parameter space rather than reconstructing all solution structure from coordinates. The reduction in target variance is encoded in the model's standardization and is visible in the shorter training time.

This finding supports a practical strategy for geothermal surrogate development: retain an inexpensive physical calculation when it is available and train the neural component on what that calculation misses. Similar logic underlies simulation-assisted machine learning for reservoir optimization (Jin et al., 2022) and physics-constrained surrogates for porous-media flow (Yan et al., 2022). The value of the strategy still depends on the baseline. If low-fidelity physics has the wrong regime, source topology, boundary response, or constitutive behavior, residual learning may inherit a misleading representation or require a correction as difficult as the original solution.

5.2 Interpreting the PINN result

The PINN's weaker result is informative because physics penalties are sometimes treated as a guaranteed improvement. Several mechanisms can explain the outcome. First, the objective combines normalized field errors with three residuals that have different derivative order and scale. A single fixed weight may not balance these contributions through training. Second, second derivatives of a compact coordinate network can be costly and noisy to optimize. Third, only 64 collocation points are used per step, which is sparse relative to the spatial, temporal, and parameter domain. Fourth, the reference data come from a finite-difference solver while the PINN uses continuous automatic differentiation. Discretization error and boundary implementation can create a mismatch between the field that best fits the data and the field that minimizes the continuous residual.

These observations align with broader cautions in physics-informed machine learning: encoding equations does not remove optimization, sampling, and model-form challenges (Karniadakis et al., 2021; Degen et al., 2023). More advanced PINN experiments could use adaptive residual weights, curriculum sampling, Fourier or sinusoidal features, domain decomposition, hard boundary constraints, separate field networks, quasi-Newton refinement, or a discrete residual matching the reference stencil. Such changes are valuable future work, but adding them selectively after observing test results would weaken the controlled nature of the present benchmark unless the complete tuning protocol were repeated for all models.

5.3 Limits of inference

The main limitation is physical scope. The scalar Poisson closure cannot represent the vector displacement and stress tensors required for induced-seismicity or fracture-stability analysis. The model omits advective heat transport, temperature- and stress-dependent properties, multiphase flow, geochemical reactions, wells, explicit fractures, anisotropy, and material heterogeneity. The smooth Gaussian sources and homogeneous outer boundaries are numerical conveniences.

The grid is coarse and no mesh or time-step convergence. The sampled ranges are dimensionless and were selected for stable, smooth solutions. There is no calibration to a named geothermal field and no comparison with experimental or private data. The three seeds quantify only a narrow combination of sampling, splitting, and initialization variability. They are not confidence intervals for real reservoir performance.

The comparison is also budget-specific. The PINN receives more steps than the other networks, but no exhaustive hyperparameter search is performed. The pure and hybrid networks use dense supervised records from every training scenario. A different regime, such as sparse observations with abundant collocation points, could change the relative value of a physics penalty. Finally, the excellent hybrid score partially reflects that the low-fidelity solver shares the reference grid and numerical operators. A stronger test would pair different meshes, discretizations, or physical approximations.

5.4 Implications for a field-scale study

A publication aimed at field-scale geothermal decisions should extend the benchmark in stages. The first extension is a full vector-elastic thermo-poroelastic formulation with physically dimensional variables and documented constitutive parameters. The second is heterogeneous permeability, porosity, thermal conductivity, and elastic moduli, followed by explicit or embedded fracture representations. The third is solver-to-surrogate validation on mesh-independent observables, such as production temperature, pressure drawdown, induced displacement, and energy recovery. The fourth is an external dataset or blind benchmark with uncertainty quantification. Only after these steps would claims about operational speedup or geothermal forecasting be justified. The broader application of geothermal surrogate models also requires careful consideration of how predicted subsurface thermal conditions relate to energy recovery and heat-transfer performance. Research on reactive heat exchangers has highlighted the relevance of thermal transport and heat-exchange processes to geothermal energy extraction (Williams, 2024). Although the present benchmark does not simulate surface heat-exchange equipment or integrated geothermal power systems, extending surrogate predictions to physically calibrated reservoir models could provide useful thermal inputs for subsequent evaluations of geothermal energy recovery.

6. Reproducibility and data availability

The accompanying archive contains the Python source, parameterized generator, model classes, PINN residual implementation, training and aggregation scripts, configuration files, tests, environment specifications, a compact example dataset, saved primary checkpoints, measured JSON and CSV outputs, a prediction sample, and editable SVG and EPS figures. Full three-seed datasets are omitted from the archive because they can be regenerated deterministically from the recorded seeds and configuration. The repository README provides the exact command sequence.

All reported numerical results have status measured in the JSON output. The environment record includes Python and PyTorch versions, platform, device, and CPU count. Unit tests check output shapes, finite values, deterministic structure, reduced mechanical closure, low-fidelity mismatch, and neural forward behavior. No private, proprietary, or human-participant data are used.

7. Conclusions

This study presents a reproducible comparison of supervised, physics-informed, and hybrid neural surrogates on a reduced three-dimensional THM geothermal benchmark. Holding out complete parameter scenarios produces a more defensible generalization test than random point splitting. Across three seeds, the hybrid residual-correction model achieves relative L2 errors near 0.013 for pressure, temperature, and mechanical potential and yields the smallest discrete residuals. The supervised model is less accurate, and the PINN does not improve on it under the reported weighting, collocation, and CPU budget.

The result supports a narrow conclusion: when an inexpensive physical baseline captures the correct qualitative solution structure, learning its correction can be more effective than applying a residual penalty to a coordinate network. The result does not rank methods universally and does not validate a field-scale enhanced geothermal system. Its value is the transparent benchmark, including an unfavorable PINN result, explicit limitations, and executable artifacts that allow the comparison to be rerun or extended.

References

1. Amini, D., Haghighat, E., & Juanes, R. (2022). Physics-informed neural network solution of thermo-hydro mechanical processes in porous media. *Journal of Engineering Mechanics, 148*(11), 04022070. https://doi.org/10.1061/(ASCE)EM.1943-7889.0002156
2. Biot, M. A. (1941). General theory of three-dimensional consolidation. *Journal of Applied Physics, 12*(2), 155-164. https://doi.org/10.1063/1.1712886
3. Degen, D., Caviedes-Voullième, D., Buiter, S. J. H., Hendricks Franssen, H.-J., Vereecken, H., González-Nicolás, A., & Wellmann, F. (2023). Perspectives of physics-based machine learning strategies for geoscientific applications governed by partial differential equations. *Geoscientific Model Development, 16*(24), 7375-7409. https://doi.org/10.5194/gmd-16-7375-2023

4. Jin, W., Atkinson, T. A., Doughty, C., Neupane, G., Spycher, N., McLing, T. L., Dobson, P. F., Smith, R., & Podgorney, R. (2022). Machine-learning-assisted high-temperature reservoir thermal energy storage optimization. *Renewable Energy, 197*, 384-397. https://doi.org/10.1016/j.renene.2022.07.118
5. Kadeethum, T., Jørgensen, T. M., & Nick, H. M. (2020). Physics-informed neural networks for solving nonlinear diffusivity and Biot's equations. *PLOS ONE, 15*(5), e0232683. https://doi.org/10.1371/journal.pone.0232683
6. Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics, 3*(6), 422-440. https://doi.org/10.1038/s42254-021-00314-5
7. Gao, W., & Zhao, J. (2024). Prediction of geothermal temperature field by multi-attribute neural network. *Geothermal Energy, 12*, Article 22. https://doi.org/10.1186/s40517-024-00300-x
8. 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
9. Wang, W., Kosakowski, G., & Kolditz, O. (2009). A parallel finite element scheme for thermo-hydro-mechanical coupled problems in porous media. *Computers & Geosciences, 35*(8), 1631-1641. https://doi.org/10.1016/j.cageo.2008.07.007
10. Williams, M. O. (2024). Development of reactive heat exchangers for enhanced geothermal energy recovery. *Power System Protection and Control, 52*(2), 18-37. https://doi.org/10.46121/pspc.52.2.3
11. Yan, B., Harp, D. R., Chen, B., & Pawar, R. (2022). A physics-constrained deep learning model for simulating multiphase flow in 3D heterogeneous porous media. *Fuel, 313*, 122693. https://doi.org/10.1016/j.fuel.2021.122693