# ECHO: AN EMBEDDED COUPLED HYBRID OPERATOR FOR THERMO-HYDRO-MECHANICAL MODELLING IN GEOTHERMAL SYSTEMS

**Anonymous authors**
Paper under double-blind review

## ABSTRACT

Coupled thermo-hydro-mechanical (THM) modelling is central to geothermal reservoir simulation, but resolving these interactions with high-fidelity numerical simulators is computationally expensive. Existing geomechanical surrogates reduce this cost but do not explicitly account for the distinct spatial and temporal behaviour of discrete fracture networks. We propose ECHO (Embedded Coupled Hybrid Operator), a hybrid surrogate that combines a Fourier neural operator (FNO) for reservoir-scale geomechanical prediction, a residual convolutional network for local near-well correction, and a graph-convolutional recurrent network (GCN–GRU) for fracture-specific dynamics. Evaluated on the Milford coupled reservoir–geomechanics dataset against a 3D U-Net, a full-grid GCN, DeepONet, a standard FNO, and U-FNO, ECHO obtains the lowest free-running fracture-response error in every evaluated operating and temporal regime, reducing elastic-dilation RMSE by 55.5–82.4% and fracture-normal-stress RMSE by 49.0–64.3% relative to the strongest competing architecture, while reaching $R^2 > 0.99$ on both targets. The evaluation considers held-out operating and material-parameter settings within a single fixed reservoir geometry and fracture network, and therefore measures interpolation within this geological configuration rather than generalisation to unseen geometries. Controlled ablations show that temporal recurrence is the most consistently beneficial architectural component, whereas graph connectivity and message passing contribute more conditionally across regimes. The complete surrogate runs in 6.6 ms per coupling transition, corresponding to a measured $4.1 \times 10^3$ replacement-runtime ratio relative to the reference finite-element geomechanics solve, and is integrated with a commercial reservoir simulator through a two-way coupling interface. Within the evaluated setting, these results support representation-specific neural surrogates as an effective approach for accelerating coupled physical simulation when a domain contains sub-regions with qualitatively different spatial structure.

## 1 INTRODUCTION

Enhanced geothermal systems (EGS) extract heat from hot, dry rock formations that lack the natural fluid pathways of conventional geothermal resources, by injecting fluid under pressure to create or reactivate fracture networks through which it can circulate (Robles et al., 2025; Aljubran & Horne, 2024). Because EGS do not depend on pre-existing hydrothermal conditions, they can in principle be deployed far more widely than conventional geothermal power, and recent field projects have demonstrated single wells sustaining on the order of 10 MW of continuous output (Singh et al., 2025). This makes EGS an attractive complement to weather-dependent renewables, which alone are estimated to leave a substantial cost gap in a fully decarbonised grid unless paired with dispatchable low-carbon generation (Tong et al., 2021; Sepulveda et al., 2018).

Realising this potential requires managing substantial operational risk. Injecting cold fluid simultaneously changes reservoir pressure and temperature, altering the stress state and deformation of the surrounding rock and, in turn, the aperture and permeability of the fracture network (Mahmoodpour et al., 2022a;b). These thermo-hydro-mechanical (THM) processes are tightly coupled, and mismanaging them can trigger induced seismicity with severe commercial consequences: the



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.

Mw 5.5 event at Pohang, Korea led to permanent facility closure and an estimated $300 million in combined damages and economic loss (Zhou et al., 2024). Assessing this risk across the many operating strategies and geological realisations relevant to a real deployment requires repeated coupled reservoir–geomechanics simulation, and the industrial workflow considered here couples SLB's INTERSECT reservoir simulator (SLB, 2026a) with the VISAGE finite-element geomechanics solver (SLB, 2026b) in a two-way loop (Hosseinzadeh et al., 2024). In benchmarking performed for this project, the geomechanical solve alone accounts for 50–80% of total coupled runtime, making it the dominant bottleneck for large-ensemble studies.

Deep-learning surrogates provide a route to amortising this cost: a network is trained on simulator-generated examples and then substituted for the expensive component during inference (Navrátil et al., 2019; Jin et al., 2020). Because the reservoir state is a spatial field rather than a fixed-length vector, this is naturally framed as operator learning, and Fourier neural operators (FNOs) in particular exploit the structured grid used by reservoir simulators through global spectral mixing (Li et al., 2021). A predecessor to this work used a hybrid FNO–XGBoost surrogate to approximate VISAGE from hydraulic forcing alone, reducing an hour-scale calculation to seconds. However, it omitted temperature as an input, used no explicit representation of the discrete fracture network, and produced one-directional predictions that could not feed back into the reservoir simulator, three limitations that make it unsuitable for THM-coupled EGS modelling on its own.

This motivates our central question: *how can reservoir-scale and fracture-specific representations be combined within a single surrogate to predict THM geomechanical response accurately and efficiently enough for repeated coupled EGS simulation?* We argue that the reservoir matrix and the discrete fracture network have fundamentally different spatial structures: the former is a dense, regular Cartesian domain well suited to global spectral operators, while the latter is a sparse, irregular set of cells for which graph-based message passing provides a more natural inductive bias. This motivates testing whether a representation-specific surrogate that treats the structured reservoir and fracture-associated cells separately can improve prediction relative to monolithic formulations operating over the complete domain.

**Contributions.** We introduce ECHO, a hybrid surrogate combining (i) a thermo-hydro-mechanical FNO (THM-FNO) for reservoir-wide prediction, (ii) a residual 3D CNN that corrects spatially concentrated error in a near-well region, and (iii) a fracture-specific graph-convolutional recurrent network (GCN–GRU) that models the discrete fracture cells autoregressively. We evaluate this design against five neural baselines under a common data protocol, isolate the contribution of each architectural choice through controlled ablations, and integrate the complete surrogate with a commercial reservoir simulator through a gRPC-based two-way coupling interface, verified at the software level. Within the evaluated EGS dataset, ECHO achieves the lowest fracture-prediction error of all six compared architectures, with a measured replacement-runtime ratio of approximately $4.1 \times 10^3$ relative to the reference VISAGE geomechanics solve.

## 2 Related Work

**THM simulation and sequential coupling.** EGS behaviour is governed by the interacting evolution of pore pressure, temperature, and mechanical deformation (Mahmoodpour et al., 2022b), most commonly solved through sequential coupling between a reservoir-flow simulator and a finite-element geomechanics solver (Hosseinzadeh et al., 2024). This decomposition is what motivates a surrogate targeted specifically at the geomechanical step, rather than the full coupled system.

**Neural operators for reservoir and PDE surrogates.** FNOs learn mappings between function spaces by combining pointwise and spectral-domain transformations (Li et al., 2021), and have been applied to subsurface flow and carbon-storage problems where the domain is a structured grid (Jin et al., 2020). DeepONet instead factorises the operator into separate branch and trunk networks (Lu et al., 2021). Extensions such as U-FNO combine spectral and convolutional paths for multiphase flow (Wen et al., 2022), and Geo-FNO learns a deformation onto a latent regular grid to support irregular geometries (Li et al., 2023). ECHO differs from these single-architecture surrogates by explicitly separating reservoir-scale and fracture-specific prediction rather than asking one global operator to represent both.

**Graph-based and recurrent surrogates.** Graph neural networks provide an inductive bias suited to irregular or non-Cartesian domains (Kipf & Welling, 2017; Bronstein et al., 2021), and adaptive-connectivity variants such as DeltaGNN show that message-passing topology can itself be learned rather than fixed a priori (Mancini & Rekik, 2025). For temporally evolving graph-structured systems, continuous-time formulations such as Temporal Graph ODEs target irregularly sampled dynamics (Gravina et al., 2024), which is directly relevant to coupled simulation schedules that mix monthly and yearly timesteps. We use a discrete GCN–GRU rather than a continuous-time model, and evaluate whether this simpler formulation is already sufficient to capture fracture-specific dynamics before considering that added complexity.

**Data-efficient extensions.** Because full-physics training data is expensive to generate, active learning has been used to prioritise the most informative additional simulations, reporting substantial error reductions over random acquisition in neural PDE solvers (Musekamp et al., 2025), and predictive-uncertainty estimates have been used to decide when a surrogate prediction should be replaced with a full-physics solve (Winovich et al., 2026). We return to both directions in Section 6 as means of addressing ECHO's principal limitation: a fixed geometry and a narrow training ensemble.

# 3 METHOD

**Problem formulation.** The coupled simulator advances through states $t_0 < t_1 < \cdots < t_T$. At each coupling transition, INTERSECT supplies reservoir pressure and temperature, and the surrogate must predict the resulting mechanical response in place of VISAGE. Rather than predicting absolute mechanical states, all ECHO components predict the *increment* over the transition,

$$\Delta a_t = a_{t+1} - a_t, \qquad \hat{a}_{t+1} = \hat{a}_t + \Delta\hat{a}_t, \tag{1}$$

which concentrates learning on the transition-induced response rather than the much larger pre-existing state, and matches the sequential structure of the simulator itself. During free-running deployment, pressure and temperature forcing continue to be supplied externally by the simulator, while the mechanical state is reconstructed autoregressively from the surrogate's own predictions. The two ECHO branches share this formulation but operate over different spatial supports: the structured branch covers the full Cartesian grid, while the fracture branch is restricted to the cells associated with the discrete fracture network (DFN).

## 3.1 STRUCTURED-GRID THM-FNO AND NEAR-WELL CORRECTION

At each transition, the THM-FNO receives the pressure and temperature increments, the current effective-stress and plastic-strain tensors, and the grid coordinates, and predicts the change in total strain, effective stress, and plastic strain (18 channels total) using four spectral layers of width 16 with $16 \times 16 \times 8$ retained Fourier modes, following the formulation of Li et al. (2021). We retain the predecessor surrogate's XGBoost-based plastic-activation gating formulation rather than its trained classifier parameters. For plastic-strain component $c$, a training cell is labelled active when the magnitude of the denormalised VISAGE increment exceeds $\tau_p = 10^{-8}$. A binary XGBoost classifier is trained for each component containing both active and inactive examples, and the corresponding FNO plastic-strain prediction is set to zero where the classifier predicts inactivity. The retained XGBoost gate does not receive temperature explicitly as an input.

Analysis of the standalone THM-FNO shows that its error is highly non-uniform: a near-well region containing a small fraction of reservoir cells accounts for over 90% of the cumulative squared stress error. We therefore introduce a residual 3D CNN (three $3 \times 3 \times 3$ convolutional layers, GELU activations, zero-initialised final layer) that takes the FNO input together with the FNO's own prediction and outputs a signed correction, applied only within a fixed Chebyshev-radius-3 neighbourhood of the well trajectories (Figure 1):

$$\Delta\hat{Y}_t^{\text{corr}} = \Delta\hat{Y}_t^{\text{FNO}} + \alpha\, M_{\text{well}} \odot \Delta\hat{Y}_t^{R}, \tag{2}$$

where $M_{\text{well}}$ is a fixed binary mask determined solely by well geometry (never by test-set error) and $\alpha$ is a learned global scale initialised to one so that the combined model initially reproduces the pre-trained FNO.

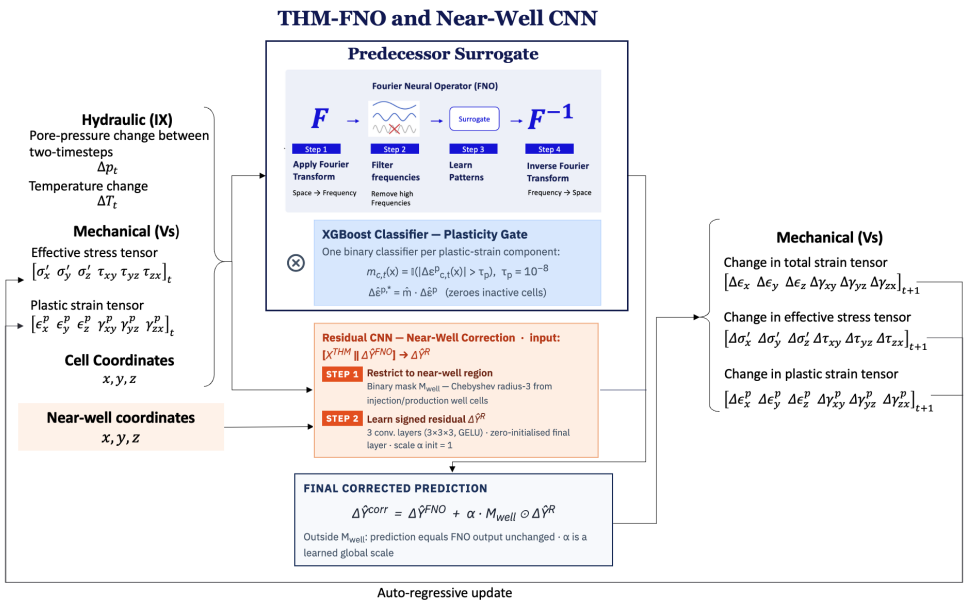



Figure 1: Structured-grid pipeline. Reservoir pressure/temperature increments and the current mechanical state are mapped by the FNO to reservoir-wide increments, gated by an XGBoost plasticity classifier, and refined within a near-well region by a residual CNN. Outside this region the corrected prediction is identical to the raw FNO output.

### 3.2 Fracture-aware GCN–GRU

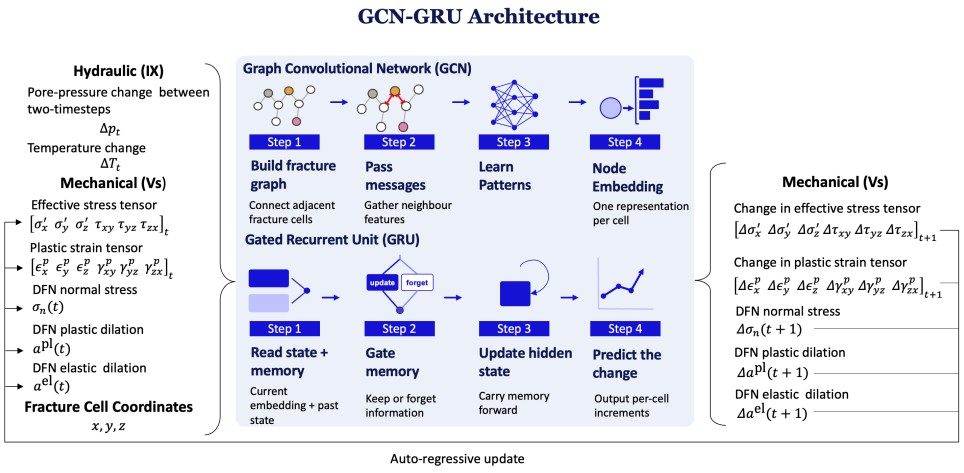



Figure 2: Fracture surrogate. A fixed adjacency graph over DFN-associated cells is processed by three graph-convolution layers, then propagated through a node-wise GRU that carries state across coupling transitions before a linear readout predicts the fracture-specific mechanical increment.

The 235 fracture-associated cells of the evaluated dataset form a sparse, irregular subset of the reservoir grid; within the receptive field of a $3 \times 3 \times 3$ convolution, these cells have on average only 6.5 same-domain neighbours, compared with 21.9 for near-well cells, motivating a graph- structured rather than purely convolutional treatment. Each graph node corresponds directly to one DFN-associated cell of the structured reservoir grid, and a fixed node ordering is retained throughout feature extraction, graph construction, inference, and state reconstruction. Consequently, no spatial interpolation between separate meshes is required: pressure, temperature, and structured-branch

mechanical quantities are gathered at the corresponding fracture-cell indices and reordered into the graph-node convention. We connect these fracture-associated cells using a fixed Cartesian-offset stencil (edges added where two cells' grid indices differ by a permitted offset), producing an adjacency matrix $A$ used throughout each temporal sequence. For the reported ECHO results, we use an unweighted six-face stencil connecting $\pm i$, $\pm j$, and $\pm k$ neighbours; alternative Cartesian stencils are evaluated in the supplementary ablation. At each transition, the 20-dimensional node feature (pressure/temperature increments, stress, strain, and fracture-specific state, plus grid coordinates) is processed by three symmetric-normalised graph-convolution layers of width 64,

$$H_t^{(\ell+1)} = \phi\Big(\tilde{D}^{-1/2}\tilde{A}\tilde{D}^{-1/2}H_t^{(\ell)}W^{(\ell)} + b^{(\ell)}\Big), \tag{3}$$

producing per-node embeddings $z_{u,t}$. To retain information across coupling transitions – which the graph convolution alone cannot, since it is applied independently at each step – a node-wise GRU maintains a recurrent hidden state $h_{u,t}$ from $z_{u,t}$ and $h_{u,t-1}$. For standalone fracture-surrogate evaluation, a linear readout predicts a 15-channel mechanical increment comprising six effective-stress increments, six total-strain increments, and the three fracture-specific increments (elastic dilation, plastic dilation, and fracture normal stress).

Because elapsed physical time $\Delta t$ is not supplied to the GRU, monthly and yearly transitions are treated as separate cadence-homogeneous prediction regimes and are modelled by separately trained instances of the same architecture rather than by a single time-aware model. The recurrent hidden state is initialised to zero at the beginning of each independent reconstruction trajectory and is propagated only within that trajectory; it is therefore not transferred from the monthly model to the yearly model or between reservoir realisations. Elastic dilation, the more dynamically variable of the two principal targets, is represented in the final fracture models using a log-ratio target. For fracture cell $u$ at transition $t$, we define

$$r_{u,t} = \log\big(\delta^e_{u,t+1} + \epsilon\big) - \log\big(\delta^e_{u,t} + \epsilon\big),$$

where $\epsilon > 0$ is a small constant ensuring that the logarithm is well defined over the evaluated trajectories. The next elastic-dilation state is reconstructed multiplicatively as

$$\hat{\delta}^e_{u,t+1} = \Big(\hat{\delta}^e_{u,t} + \epsilon\Big)\exp(\hat{r}_{u,t}) - \epsilon.$$

We use this formulation as a target parameterisation rather than as a general physical constraint; its suitability is specific to the elastic-dilation range represented in the dataset.

The standalone and coupled fracture surrogates use the same graph-recurrent architecture but are trained as separate models with different prediction interfaces. During standalone fracture-surrogate evaluation, the model predicts the full 15-channel mechanical increment described above. Within the complete ECHO system, effective stress and total strain at the fracture-associated cells are instead supplied by the structured branch, so predicting these quantities again would duplicate the role of the THM-FNO. The coupled fracture model is therefore trained with a three-channel readout that predicts only elastic dilation, plastic dilation, and fracture normal stress, which are the fracture-specific quantities required for state reconstruction and the deterministic property update returned to INTERSECT. The standalone and coupled models share the selected architecture and training configuration, but not their learned weights or output layers. Fracture normal stress remains a distinct fracture-specific simulator quantity rather than a duplicate of the reservoir-scale effective-stress tensor; the latter is supplied to the graph model as part of its local mechanical state, while fracture normal stress is predicted as a separate target.

## 4 Experimental Setup

**Dataset.** We used a $180 \times 60 \times 25$ coupled INTERSECT–VISAGE simulation ensemble (≈292,500 finite elements) containing one injector, one producer, and a discrete fracture network of 235 cells (0.087% of the grid) across four fracture stages. Sixteen realisations are split evenly between rate-controlled and pressure-controlled (BHP) operation, with Latin hypercube sampling over the operating condition, thermal expansion coefficient, fracture normal stiffness, and dilation angle. These realisation-level parameters are varied during full-physics data generation but are not supplied explicitly as surrogate input features; instead, ECHO conditions on the resulting reservoir and mechanical state available at each coupling transition. Each realisation contains 22 coupled states sampled

monthly during the early portion of the trajectory and annually thereafter. The monthly and yearly transitions are partitioned into separate cadence-homogeneous sequences before recurrent training and evaluation because the GRU does not receive elapsed physical time as an input. The monthly regime contains 11 transitions and the yearly regime contains 8 transitions; each regime starts with an independently initialised recurrent hidden state. Cases are partitioned 6/1/1 (train/validation/held-out test) at the realisation level within each control family to avoid leaking trajectory-specific information, giving four evaluation settings (BHP-monthly, BHP-yearly, rate-monthly, rate-yearly).

**Metrics.** We report increment-level RMSE and $R^2$ (pooled over cells and transitions) for elastic dilation and fracture normal stress – the two quantities that determine the fracture-property update used for two-way coupling – together with state MAE after autoregressive reconstruction. Negative reconstructed elastic-dilation states are additionally inspected as a diagnostic of rollout behaviour, but are not used for cross-architecture ranking because reconstruction parameterisations differ between model formulations. Each configuration is trained with five random seeds, and results are reported as mean $\pm$ sample standard deviation across seeds. Because each control family contains only one held-out physical realisation, these error bars quantify optimisation variability conditional on the same test trajectory rather than variability across independently sampled operating or material-parameter realisations.

**Baselines.** Fracture-level results are compared against persistence, a training-mean baseline, and ridge regression to isolate gains attributable to state-dependent and non-linear learning. The complete ECHO system is also benchmarked against a 3D U-Net (Çiçek et al., 2016), a full-grid GCN (Kipf & Welling, 2017), DeepONet (Lu et al., 2021), a standard FNO (Li et al., 2021), and U-FNO (Wen et al., 2022), all trained on the same six full-physics realisations per control family and evaluated under the same free-running protocol. These complete-model baselines differ substantially in architecture, spatial representation, and parameter count, so this benchmark compares the performance of the implemented surrogate formulations rather than providing a parameter-matched isolation of architecture. Controlled fracture-domain comparisons and parameter-matched ablations are used separately to examine the contributions of representation, recurrence, and message passing.

## 5 RESULTS

**Effect of temperature and local correction.** Supplying the true temperature increment to the FNO reduces mean stress-increment RMSE by 4.24% and strain-increment RMSE by 4.95% relative to a hydro-mechanical baseline that receives no thermal signal, consistent with the physical role of injection-driven cooling. This does not, however, remove the dominant spatial error pattern: across five seeds, cells within a fixed near-well mask account for $92.08 \pm 0.17\%$ of the cumulative squared stress error. Adding the residual CNN under joint optimisation reduces mean stress-increment RMSE by a further 47.88% (to $0.01595 \pm 0.00036$ bar) and strain-increment RMSE by 47.94%, with joint optimisation of the FNO and residual parameters together outperforming a frozen-backbone variant by 7–9%.

**Capacity-controlled fracture representation comparison.** To distinguish the effect of fracture-specific representation from simply adding modelling capacity, we compare the GCN–GRU with a fracture-masked residual CNN operating on the same 235 fracture-associated cells under a capacity-controlled setup. Although the near-well correction already reaches 202 of these 235 cells, the Cartesian CNN control leaves a persistent, spatially concentrated error pattern, whereas the GCN–GRU reduces error more broadly across the fracture domain. This controlled comparison provides evidence that the fracture-specific representation contributes beyond additional local Cartesian capacity: on the held-out realisations, the GCN–GRU reaches elastic-dilation $R^2$ of 0.9922–0.9950 and fracture-normal-stress $R^2$ of 0.9970–0.9979 across the four regimes.

**Ablations.** Controlled, parameter-matched ablations (35,599 parameters in every variant) isolate the contribution of recurrence and message passing. Removing the recurrent pathway increases elastic-dilation RMSE by 57.4–100.9% and normal-stress RMSE by 41.4–130.1% across all four regimes – the single most consistent effect observed in this work. Removing graph message passing instead produces a regime-dependent effect: it degrades both targets under BHP control but *improves* elastic-dilation and normal-stress RMSE by 12.5% and 5.6% respectively under monthly rate con-

Table 1: Held-out free-running fracture-response error, ECHO vs. the strongest competing architecture in each regime (from among 3D U-Net, full-grid GCN, DeepONet, FNO, U-FNO). Ranges are computed across the four evaluation regimes (BHP/rate × monthly/yearly).

| Target | Increment RMSE reduction | Final-step RMSE reduction | Reconstructed-state RMSE reduction |
|---|---|---|---|
| Elastic dilation | 55.5–82.4% | 37.2–68.7% | 62.2–87.7% |
| Fracture normal stress | 49.0–64.3% | 46.8–57.6% | 58.5–73.4% |

trol, indicating that the fixed six-face fracture connectivity is not uniformly informative. A separate connectivity sweep (2-, 4-, 6-, 8-, and 18-direction stencils) finds that the densest (18-direction) graph consistently increases error by 8–54% relative to the six-face graph across all regimes, while sparser topologies help in some regime–target combinations but not others; six-face connectivity is retained as the most consistent compromise. An autoregressive-exposure sweep over the training rollout window $W$ shows a similar regime split: BHP models prefer a short rollout ($W{=}5$), while rate-controlled models prefer full-sequence exposure, with the choice treated as a regime-dependent training decision rather than a universal setting.

**Benchmark against alternative architectures.** Figure 3 compares all six surrogate families on held-out free-running prediction. ECHO obtains the lowest increment RMSE in every regime–target combination, with the margin summarised in Table 1; increment $R^2$ for ECHO ranges from 0.9925–0.9966 (elastic dilation) and 0.9975–0.9978 (normal stress), while DeepONet produces negative $R^2$ in most comparisons and the full-grid GCN produces negative elastic-dilation $R^2$ in all four regimes. This ordering is preserved after autoregressive reconstruction of the fracture-state trajectory (Table 1, right column), indicating that ECHO's advantage is not confined to isolated transition-level predictions. The final ECHO fracture models use the multiplicative log-ratio reconstruction described in Section 3.2. Because this target parameterisation differs from a direct additive reconstruction and can itself affect the admissible range of reconstructed states, we do not use the incidence of negative reconstructed elastic-dilation states as evidence of architectural superiority over the compared baselines.

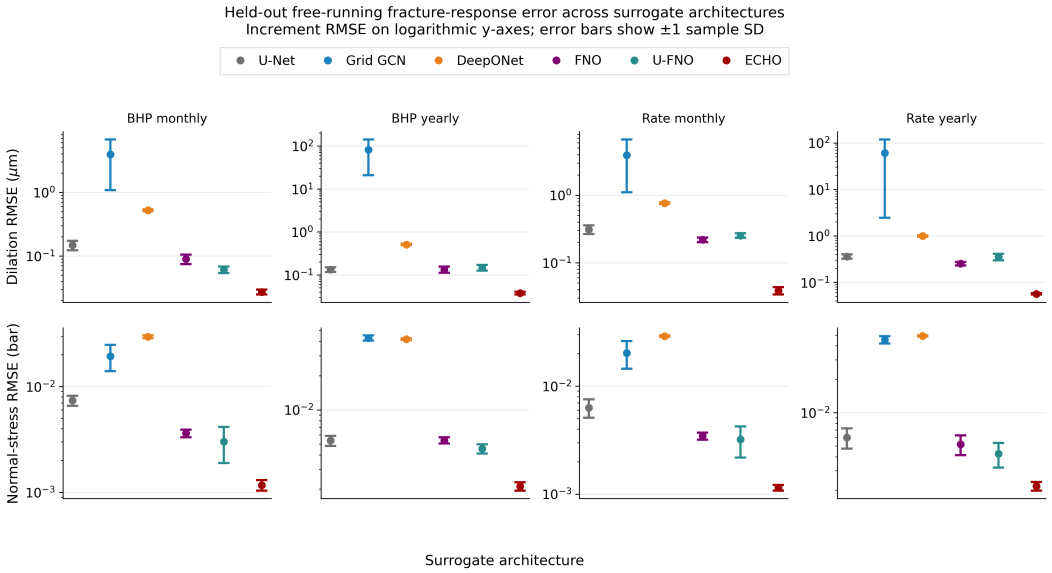



Figure 3: Held-out free-running increment RMSE for elastic dilation (top) and fracture normal stress (bottom) across the four evaluation regimes, for all six compared architectures. Axes are logarithmic; error bars show $\pm 1$ sample standard deviation across five seeds.

Table 2: Inference cost relative to the measured VISAGE geomechanics-solve time. Neural-model timings use batch size one on a single NVIDIA A100 80GB and report the mean over three sweeps of 20 timed 11-transition rollouts. The VISAGE hardware configuration is not available in the present benchmark record; runtime ratios therefore compare the measured replacement costs and should not be interpreted as hardware-matched solver speed-ups.

| Architecture | Parameters | Latency (ms/trans.) | Peak mem. (MiB) | Runtime ratio vs. VISAGE |
|---|---|---|---|---|
| 3D U-Net | 22,430,351 | $12.14 \pm 0.04$ | 864 | 2,239× |
| Full-grid GCN | 2,767 | $9.00 \pm 0.01$ | 366 | 3,025× |
| DeepONet | 269,359 | $6.54 \pm 0.00$ | 336 | 4,164× |
| Vanilla FNO | 5,258,031 | $5.73 \pm 0.00$ | 341 | 4,753× |
| U-FNO | 4,866,519 | $25.03 \pm 0.01$ | 505 | 1,087× |
| **ECHO (both branches)** | 1,423,576 | 6.63 | 233 | **4,107**× |
| VISAGE solve | — | 27,217 | — | 1× |

**Computational cost.** Table 2 reports neural inference cost measured on a single NVIDIA A100 80GB GPU (batch size one, checkpoint loading excluded). The complete ECHO path (reservoir and fracture branches together) requires 6.63 ms per coupling transition, compared with a measured 27.217 s average VISAGE geomechanics solve in the reference coupled workflow, corresponding to a $4.1 \times 10^3$ replacement-runtime ratio. Because the hardware configuration associated with the reference VISAGE timing is not available in the present benchmark record, this quantity is not a hardware-matched solver speed-up. Retaining the measured non-mechanical share of the coupled step gives an estimated end-to-end Amdahl factor of approximately $2.3\times$ for the reference workflow; this estimate inherits the same hardware-comparison limitation.

## 6 Discussion and Limitations

Across the evaluated complete formulations and controlled fracture-domain ablations, reservoir-wide and fracture-specific mechanical behaviour are well represented by different spatial and temporal components within the same surrogate. Temporal recurrence is the most consistently beneficial component we identify; graph connectivity and message passing help in most, but not all, of the evaluated regimes, suggesting that the fixed Cartesian-offset fracture graph does not uniformly capture the interactions most relevant to prediction.

These results are also bounded in scope. All realisations share one reservoir geometry and one 235-cell fracture network, so the held-out evaluation measures interpolation across unseen operating and material-parameter settings within a fixed geological configuration rather than generalisation to new geometries; addressing this would likely require pairing geometry-general operator learning (e.g. Li et al. 2023) with active data acquisition to offset the cost of generating additional full-physics labels (Musekamp et al., 2025). Moreover, each control family contains only one held-out physical realisation, so the reported test performance should be interpreted as out-of-sample accuracy on these reserved trajectories rather than as an estimate of variability across the full sampled parameter space. The sampled constitutive parameters, including thermal expansion coefficient, fracture normal stiffness, and dilation angle, are varied during full-physics data generation but are not supplied explicitly as surrogate input features. Their influence is therefore represented indirectly through the evolving reservoir and mechanical state available at each coupling transition. The present results consequently demonstrate empirical interpolation within the sampled ensemble rather than a parameter-explicit constitutive mapping; explicit conditioning on realisation-level material parameters is a natural direction for future work. The evaluated ensemble also remains in the elastic mechanical regime, so ECHO has not been tested where fracture deformation is irreversible. In addition, the retained XGBoost plastic-activation gate does not receive temperature explicitly as an input; its adequacy under thermally induced plastic activation therefore remains unevaluated. The fixed six-face fracture graph could plausibly be improved by weighting or learning edges from fracture-specific quantities such as orientation or transmissibility (Mancini & Rekik, 2025), or by replacing the cadence-specific monthly/yearly models with a single time-conditioned formulation that receives $\Delta t$ explicitly (Gravina et al., 2024). Finally, the present study evaluates two aspects of sequential deployment separately. First, the free-running evaluation measures accumulation of surrogate error as the predicted mechanical state is reconstructed autoregressively

across the held-out sequence, while reservoir pressure and temperature remain externally supplied. Second, software-level integration tests verify that surrogate-predicted fracture-property updates are propagated through the two-way coupling interface and consumed at the subsequent simulator step. These results establish autoregressive surrogate accuracy and functional two-way coupling independently, but they do not constitute a quantitative validation of long-horizon closed-loop agreement between an INTERSECT–ECHO trajectory and the corresponding INTERSECT–VISAGE solution. Because prediction errors in fracture state may alter permeability and thereby modify subsequent reservoir states, such closed-loop physical validation remains an important direction for future work; predictive-uncertainty-gated fallback to the full-physics solver (Winovich et al., 2026) provides one possible mechanism for bounding this risk during deployment.

## 7 Conclusion

We presented ECHO, a hybrid neural surrogate that pairs a thermo-hydro-mechanical FNO and local residual correction for reservoir-scale prediction with a graph-convolutional recurrent network for fracture-specific dynamics, and integrated it with a commercial reservoir simulator through a two-way coupling interface. Across a coupled EGS benchmark, ECHO achieves the lowest held-out fracture-prediction error of six compared architectures in every evaluated regime, with the advantage persisting through autoregressive rollout. The complete surrogate has a measured replacement-runtime ratio of approximately $4.1 \times 10^3$ relative to the reference finite-element geomechanics solve. Within the evaluated setting, these results support representation-specific neural surrogates as an effective approach for accelerating coupled physical simulation when a domain contains sub-regions with qualitatively different spatial structure.

### Reproducibility Statement

The THM-FNO, residual CNN, and GCN–GRU architectures are specified in Section 3, and the training objectives, rollout evaluation, and model-selection protocol are described in Sections 3–4. The supplementary material provides complete architecture and feature definitions, fracture-graph construction and temporal-rollout details, the data-partition and evaluation protocol, controlled ablations, complete benchmark results, and simulator-interface verification. All reported multi-seed results use five fixed training seeds, with the resulting variability interpreted as sensitivity to stochastic training initialisation as described in Section 4. The underlying simulation ensemble, trained model weights, ECHO source code (including model definitions, training and evaluation pipelines, and simulator-integration code), and commercial simulator infrastructure are proprietary and cannot be released; the methodological details provided in the paper and supplementary material are intended to support independent reimplementation on an equivalent coupled reservoir–geomechanics dataset.

### Ethics Statement

This work uses synthetic reservoir-simulation data and does not involve human participants, personal data, or dual-use research of concern.

### AI use statement

Several generative AI tools were used in the research and preparation of this submission. Perplexity was used for literature search, citation discovery, and locating relevant passages in source papers. ChatGPT and Gemini were used for writing assistance, including drafting and condensing portions of the manuscript, restructuring text, improving grammar and clarity, and providing feedback on organisation. Cursor and Claude Code were used to support implementation and code development. All AI-assisted text and code were reviewed by the authors before inclusion. Generative AI tools were not used to generate the experimental results reported in this paper. The authors reviewed and verified the final manuscript, analyses, claims, and reported results and take responsibility for the final content of the work, including text and code produced with the assistance of generative AI.

## REFERENCES


M. J. Aljubran and R. N. Horne. Power supply characterization of baseload and flexible enhanced geothermal systems. *Scientific Reports*, 14:17619, 2024.

M. M. Bronstein, J. Bruna, T. Cohen, and P. Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 424–432, 2016.

A. Gravina, D. Zambon, D. Bacciu, and C. Alippi. Temporal graph ODEs for irregularly-sampled time series. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 4025–4034, 2024.

B. Hosseinzadeh, F. Amour, M. R. Hajiabadi, A. Abdollahi, and H. M. Nick. Is two-way coupled geomechanical modeling essential for CO2 storage in deformable chalk reservoirs? In *Proceedings of the 17th International Conference on Greenhouse Gas Control Technologies (GHGT-17)*, 2024.

Z. L. Jin, Y. Liu, and L. J. Durlofsky. Deep-learning-based surrogate model for reservoir simulation with time-varying well controls. *Journal of Petroleum Science and Engineering*, 192:107273, 2020.

T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations (ICLR)*, 2017.

Z. Li, N. Kovachki, K. Azizzadenesheli, B. Liu, K. Bhattacharya, A. M. Stuart, and A. Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations (ICLR)*, 2021.

Z. Li, D. Z. Huang, B. Liu, and A. Anandkumar. Fourier neural operator with learned deformations for PDEs on general geometries. *Journal of Machine Learning Research*, 24(388):1–26, 2023.

L. Lu, P. Jin, G. Pang, Z. Zhang, and G. E. Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3): 218–229, 2021.

S. Mahmoodpour, M. Singh, K. Bär, and I. Sass. Thermo-hydro-mechanical modeling of an enhanced geothermal system in a fractured reservoir using carbon dioxide as heat transmission fluid – a sensitivity investigation. *Energy*, 254:124266, 2022a.

S. Mahmoodpour, M. Singh, A. Turan, K. Bär, and I. Sass. Simulations and global sensitivity analysis of the thermo-hydraulic-mechanical processes in a fractured geothermal reservoir. *Energy*, 247:123511, 2022b.

K. Mancini and I. Rekik. DeltaGNN: Graph neural network with information flow control. *arXiv preprint arXiv:2501.06002*, 2025.

D. Musekamp, M. Kalimuthu, D. Holzmüller, M. Takamoto, and M. Niepert. Active learning for neural PDE solvers. In *International Conference on Learning Representations (ICLR)*, 2025.

J. Navrátil, A. King, J. Rios, G. Kollias, R. Torrado, and A. Codas. Accelerating physics-based simulations using end-to-end neural network proxies: An application in oil reservoir modeling. *Frontiers in Big Data*, 2:33, 2019.

A. Robles, J. Perez, C. Patiño Suarez, M. Gambús-Ordaz, and J. Santos. Hot dry rock (hdr) production systems for geothermal energy development: A focus on egs and ags systems. In *Third EAGE Workshop on Geothermal Energy*, 2025.

N. A. Sepulveda, J. D. Jenkins, F. J. de Sisternes, and R. K. Lester. The role of firm low-carbon electricity resources in deep decarbonization of power generation. *Joule*, 2(11):2403–2420, 2018.

A. Singh, G. Galban, M. McClure, K. Briggs, and J. Norbeck. Designing the record-breaking enhanced geothermal system at project cape. In *Unconventional Resources Technology Conference (URTeC)*, 2025.

SLB. Intersect reservoir simulator. `https://www.slb.com/`, 2026a.

SLB. Visage finite-element geomechanics simulator. `https://www.slb.com/`, 2026b.

D. Tong, D. J. Farnham, L. Duan, Q. Zhang, N. S. Lewis, K. Caldeira, et al. Geophysical constraints on the reliability of solar and wind power worldwide. *Nature Communications*, 12:6146, 2021.

G. Wen, Z. Li, K. Azizzadenesheli, A. Anandkumar, and S. M. Benson. U-FNO—an enhanced fourier neural operator-based deep-learning model for multiphase flow. *Advances in Water Resources*, 163:104180, 2022.

N. Winovich, M. Daneker, L. Lu, and G. Lin. Active operator learning with predictive uncertainty quantification for partial differential equations. *Journal of Computational Physics*, 555:114791, 2026.

W. Zhou, F. Lanza, I. Grigoratos, R. Schultz, J. Cousse, E. Trutnevyte, et al. Managing induced seismicity risks from enhanced geothermal systems: A good practice guideline. *Reviews of Geophysics*, 62(4):e2024RG000849, 2024.