# GEOMETRY–CONDITION FACTORIZED NEURAL OPERATORS FOR ELECTROMAGNETIC MODELING

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

## ABSTRACT

Electromagnetic forward modeling is fundamental to applications ranging from mineral exploration to cerebral electrical impedance tomography (EIT), yet repeated high-fidelity simulations remain computationally expensive. Starting from the Maxwell-governed forward operator, we analyze its local Fréchet sensitivity with respect to the three-dimensional electrical-property field. The resulting sensitivity exhibits a pronounced **geometry–condition** organization: the corresponding unfolding shows substantially stronger spectral concentration than alternative variable partitions, indicating that **observation geometry** and **excitation condition** admit distinct yet coupled representations. Guided by this structure, we propose the *Geometry–Condition Factorized Neural Operator* (GCFNO). A shared volumetric encoder extracts the latent electromagnetic state, a geometry factor models receiver interactions through coordinate-conditioned message passing, while a condition factor performs frequency-guided aggregation of depth-resolved features. A nonlinear readout then couples the two factors to predict sensor responses. On 3-D magnetotelluric modeling, GCFNO achieves the lowest three-seed mean total RMSE, while substantially improving held-out-frequency and receiver-geometry generalization. On head EIT, GCFNO achieves the lowest three-seed mean patient-macro error among evaluated surrogates. More broadly, these results suggest a path toward problem-specific neural operators whose architectures are derived from the governing physics, enabling more accurate and generalizable surrogates for complex physical systems. The methods and related analysis code of this manuscript are available at `https://anonymous.4open.science/r/iclr2027-code-artifact-788E/`.

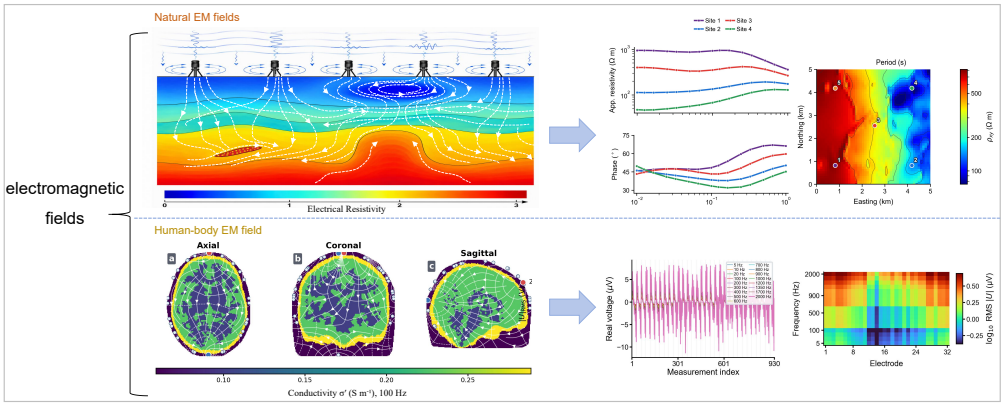


Figure 1: **Electromagnetic forward modeling in geophysics and biomedical imaging.** The upper panel maps subsurface resistivity to MT apparent-resistivity and phase responses; the lower panel maps head conductivity to multi-frequency EIT electrode voltages.

## 1 INTRODUCTION

Electromagnetic forward modeling underpins subsurface exploration and electrical impedance imaging (Chave & Jones, 2012; Adler & Boyle, 2017). Maxwell's equations and their appropriate quasi-static approximations relate heterogeneous electrical properties to electromagnetic responses



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.

under specified source and boundary conditions. Figure 1 illustrates two applications: magnetotellurics (MT) and head electrical impedance tomography (EIT). In MT, subsurface resistivity determines frequency-dependent responses at surface receivers (Qin & Grayver, 2026). In head EIT, tissue admittivity and electrode geometry determine boundary voltages under prescribed current injections (Candiani & Santacesaria, 2022; Abdelmoumene et al., 2026); multi-frequency measurements support applications such as stroke imaging (Goren et al., 2018; McDermott et al., 2020). Both applications require repeated forward evaluations for inversion, uncertainty quantification, and acquisition design, making high-fidelity three-dimensional simulation costly (Egbert & Kelbert, 2012; Kelbert et al., 2026). This motivates learned surrogates that account for material heterogeneity, observation geometry, and excitation conditions.

Neural operators provide a framework for learning such mappings between physical fields and responses (Kovachki et al., 2023). A central design question is how to organize geometry and excitation conditions within a shared representation. Receiver location determines where a field is observed, whereas frequency changes propagation and sensitivity to structures at different depths. Treating these variables as generic inputs does not explicitly encode their distinct physical roles. We therefore ask whether electromagnetic sensitivity supports a compact representation that shares geometric interactions across frequencies while retaining frequency-dependent depth information.

We address this question through a local sensitivity analysis and numerical diagnostics from the MT solver. The proposed geometry–condition partition concentrates sensitivity energy into a compact factorization, motivating the Geometry–Condition Factorized Neural Operator (GCFNO). Its geometry and condition pathways access a shared volumetric representation and interact through a nonlinear readout. This design connects a measurable property of the physical operator to an architectural inductive bias, which we evaluate on MT and head EIT.

Our contributions are threefold:

1. We establish a local geometry–condition approximation with an explicit spectral-tail and nonlinear-remainder bound. Reference-model analysis and solver sensitivity spectra support this partition. When the spectrum is concentrated, compact local approximations require fewer effective degrees of freedom, providing a theoretical motivation for data-efficient learning.
2. We introduce GCFNO, coupling geometric message passing with condition-dependent aggregation of volumetric features. Shared geometric interactions and frequency-dependent depth access support both dense-grid and coordinate-query prediction.
3. Experiments on three-dimensional MT and head EIT demonstrate improved predictive accuracy. Held-out-frequency, receiver-geometry, and reduced-data experiments on MT assess generalization and data efficiency.

## 2 Related Work

**Neural operators for physical simulation** DeepONet (Lu et al., 2021) pairs input-function encoding with coordinate-based evaluation, while Fourier neural operators (FNOs) (Li et al., 2021) use spectral convolutions on regular grids. Graph neural operators (GNOs) (Li et al., 2020) parameterize coordinate-dependent interaction kernels, and geometry-informed neural operators (GINO) (Li et al., 2023) connect irregular domains with regular latent grids. Other developments emphasize latent-space computation and query decoding (Alkin et al., 2024), multi-PDE pretraining (Herde et al., 2024; Holzschuh et al., 2025), and physical-parameter modulation (Jing et al., 2026). These approaches inform operator representation and conditioning. GCFNO focuses on choosing a geometry–condition partition from electromagnetic sensitivity structure and preserving the volumetric information needed by each condition.

**Learning-based electromagnetic forward modeling** Convolutional surrogates learn three-dimensional MT resistivity-to-response maps (Wang et al., 2024), with attention-gated extensions improving feature fusion (Zhong et al., 2026). EFNO (Peng et al., 2022) and EFKAN (Wang et al., 2026b) use coordinate- and frequency-conditioned readouts; EFDO (Liao et al., 2025) fuses resistivity and frequency encodings before spatial processing. In EIT, MCMC-Net (Majee et al., 2025) uses a learned forward surrogate for likelihood evaluations in Bayesian inversion. GCFNO builds on

explicit acquisition conditioning by sharing geometry-dependent interactions across frequencies and aggregating depth-resolved features according to the requested condition.

**Factorized representations for operator learning** F-FNO (Tran et al., 2023) factorizes spectral convolutions across coordinate directions, while SepONet (Yu et al., 2024) constructs tensor-product bases using coordinate-wise trunk networks. Tucker-FNO (Zhou et al., 2026) combines one-dimensional Fourier operators through a Tucker representation. These methods motivate efficient representations through coordinate factorization. GCFNO instead groups receiver–horizontal geometry and frequency–depth dependence according to local electromagnetic sensitivity, supported by comparisons against alternative variable partitions. This grouping motivates distinct geometric and condition pathways within a nonlinear operator; the local approximation does not impose a globally separable response map.

# 3 METHOD

## 3.1 PROBLEM FORMULATION

Let $\mathbf{X}$ denote an electrical-property field discretized over a three-dimensional domain. For receiver coordinates $\mathcal{P}_q = \{\mathbf{p}_i\}_{i=1}^{N_q}$ and excitation condition $c$, the physical forward operator is

$$\mathcal{G}^\star : (\mathbf{X}, \mathcal{P}_q, c) \mapsto \mathbf{Y}(\mathcal{P}_q, c), \tag{1}$$

where $\mathbf{Y}(\mathcal{P}_q, c) \in \mathbb{R}^{N_q \times C_y}$ contains $C_y$ response components per receiver. Complex-valued outputs are represented by their real and imaginary components.

In MT, $\mathbf{X} \in \mathbb{R}^{C_x \times D \times H \times W}$ is a voxelized resistivity volume, $c = f$ is excitation frequency, and outputs contain apparent-resistivity and phase components. In EIT, $\mathbf{X}$ contains cellwise admittivity spectra on an unstructured head mesh. Here $c = (f, s)$ specifies a frequency and prescribed current-injection pattern $s$ from the fixed acquisition protocol; outputs are complex electrode voltages.

We learn a parametric approximation $\mathcal{G}_\theta$ to $\mathcal{G}^\star$. For requested conditions $\mathcal{C}_q = \{c_m\}_{m=1}^{M_q}$, predictions are stacked along the condition axis:

$$\widehat{\mathbf{Y}}_q = \mathrm{stack}_{m=1}^{M_q} \mathcal{G}_\theta(\mathbf{X}, \mathcal{P}_q, c_m) \in \mathbb{R}^{N_q \times M_q \times C_y}. \tag{2}$$

## 3.2 SENSITIVITY STRUCTURE AND DESIGN PRINCIPLES

### 3.2.1 LOCAL ELECTROMAGNETIC SENSITIVITY

The local sensitivity describes how a material perturbation changes the measured response. In MT, the electric field satisfies $\mathcal{L}_{\sigma,f} E = b_f$, where $\mathcal{L}_{\sigma,f} = \nabla \times \mu^{-1} \nabla \times + i\omega\sigma$ and $\omega = 2\pi f$, under the $e^{i\omega t}$ convention; $\mu$ is magnetic permeability and $\sigma$ is conductivity. Fix the sources, boundary data, and conductivity background $\sigma_0$, with corresponding field $E_0$. Assume that $\mathcal{L}_{\sigma_0,f}$ has a bounded inverse on homogeneous field variations and that the sensor response has a bounded derivative. Differentiating the field equation gives the first variation

$$\mathcal{L}_{\sigma_0,f}\, \delta E = -i\omega\, \delta\sigma\, E_0, \qquad \delta E = -i\omega \mathcal{L}_{\sigma_0,f}^{-1}(\delta\sigma\, E_0). \tag{3}$$

Thus the response variation is a linear functional of the material perturbation. Applying the response derivative and the chain rule for the chosen input coordinates $\mathbf{X}$ yields

$$[D_{\mathbf{X}} \mathcal{G}^\star(\mathbf{X}_0, \mathcal{P}_q, f)[\Delta\mathbf{X}]]_i = \int_\Omega K_{\mathbf{X}_0}(\mathbf{p}_i, \boldsymbol{\xi}, z, f)\, \Delta\mathbf{X}(\boldsymbol{\xi}, z)\, \mathrm{d}\boldsymbol{\xi}\, \mathrm{d}z, \tag{4}$$

where $\boldsymbol{\xi}$ and $z$ denote horizontal position and depth. The kernel measures how a perturbation at $(\boldsymbol{\xi}, z)$ affects receiver $\mathbf{p}_i$ at frequency $f$. Its adjoint representation combines the forward and adjoint fields with the input and response transformations. For EIT with fixed electrodes, currents, and contact impedances, the corresponding kernel combines forward and adjoint potential gradients.

The kernel's variables have distinct physical roles. In a homogeneous conducting half-space, the field amplitude decays as $e^{-z/\delta(f)}$, with $\delta(f) = \sqrt{2/(2\pi f \mu \sigma_0)}$. Frequency therefore sets a characteristic depth scale. In a horizontally translation-invariant reference problem with uniform excitation,

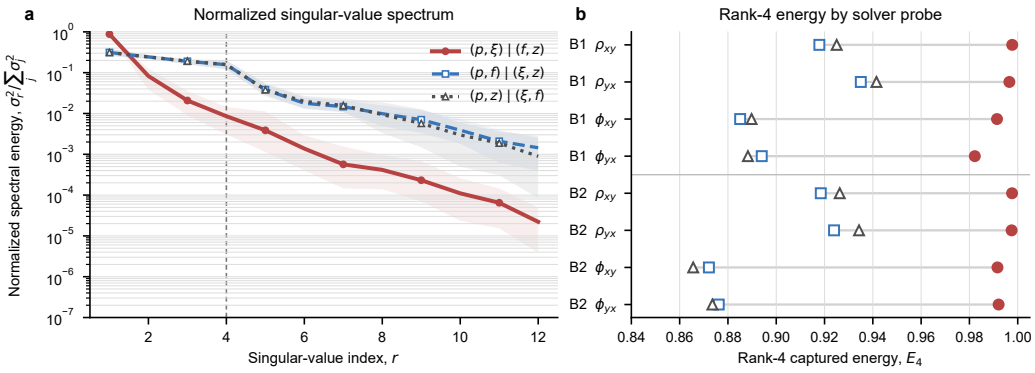


Figure 2: **Spectral concentration of solver sensitivity estimates.** (a) Normalized squared singular values for three variable groupings. Spectra are normalized separately for each response component and averaged within each background; curves show the mean and shading shows the range across 12 backgrounds. (b) Difference in rank-four captured energy between the proposed grouping and the dimension-matched control. Each point represents one background, averaged over four response components; positive values favor the proposed grouping.

translating both the receiver and material perturbation leaves the response variation unchanged. Consequently,

$$K_{\mathbf{X}_0}(\mathbf{p}, \boldsymbol{\xi}, z, f) = \kappa_{\mathbf{X}_0}(\mathbf{p} - \boldsymbol{\xi}, z, f). \tag{5}$$

This motivates grouping receiver–horizontal geometry $(\mathbf{p}, \boldsymbol{\xi})$ and frequency–depth dependence $(f, z)$. We next quantify the accuracy of a compact representation under this grouping.

### 3.2.2 A Local Geometry–Condition Approximation

Let $F : \mathbb{R}^{N_\xi \times N_z} \to \mathbb{R}^{N_p \times N_f}$ map the material field to one real response component. In MT, $\mathbf{X} = \log_{10} \rho$ and responses are log-apparent resistivities or phases normalized by $180^\circ$ on a locally continuous branch. At $\mathbf{X}_0$, group the sensitivity indices into geometry $(p, \xi)$ and frequency–depth $(f, z)$:

$$K_{(p,\xi),(f,z)} = \frac{\partial F_{p,f}}{\partial X_{\xi,z}}(\mathbf{X}_0), \qquad K = \sum_{r \geq 1} s_r u_r v_r^\top, \tag{6}$$

with $s_1 \geq s_2 \geq \cdots \geq 0$.

**Proposition 1** (Local approximation by geometry–condition factors)**.** *Suppose $F$ is $C^2$ near $\{\mathbf{X}_0 + t\Delta\mathbf{X} : 0 \leq t \leq 1\}$, with second derivative bounded by $L$ in bilinear Frobenius operator norm. For $1 \leq R \leq \min(N_p N_\xi, N_f N_z)$, reshape $u_r, v_r$ into $A_r \in \mathbb{R}^{N_p \times N_\xi}$ and $B_r \in \mathbb{R}^{N_f \times N_z}$. Then*

$$J_R[\Delta\mathbf{X}] = \sum_{r=1}^{R} s_r A_r \Delta\mathbf{X} B_r^\top \tag{7}$$

*satisfies*

$$\|F(\mathbf{X}_0 + \Delta\mathbf{X}) - F(\mathbf{X}_0) - J_R[\Delta\mathbf{X}]\|_F \leq \varepsilon_R \|\Delta\mathbf{X}\|_F + \frac{L}{2}\|\Delta\mathbf{X}\|_F^2, \tag{8}$$

*where $\varepsilon_R^2 = \sum_{r>R} s_r^2$.*

The full proof is provided in Appendix B. Here $A_r$ maps horizontal perturbations to receivers, while $B_r$ aggregates depth information for each frequency. A small spectral tail therefore supports a compact local operator with shared geometric interactions and frequency-dependent depth access. The separation rank $R$ characterizes the compactness of this local sensitivity representation.

### 3.2.3 Implications for Operator Learning and Generalization

**Alignment with solver sensitivities.** For nonzero $K$, define the rank-$R$ captured energy as $E_R = \sum_{r=1}^{R} s_r^2 / \|K\|_F^2$. The spectral tail then satisfies $\varepsilon_R^2 = (1 - E_R)\|K\|_F^2$. We evaluate

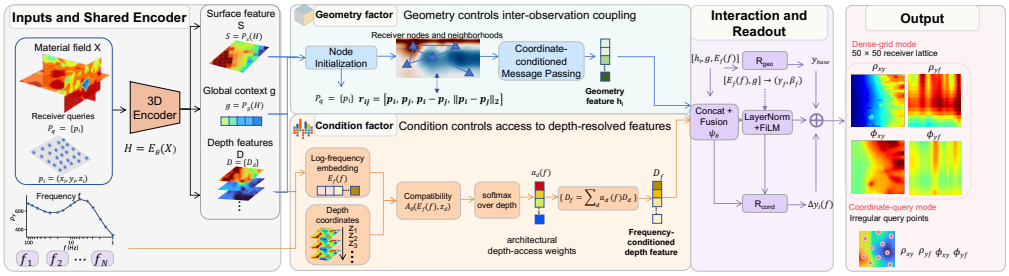


Figure 3: **Geometry–Condition Factorized Neural Operator (GCFNO).** A shared 3-D encoder produces surface-aligned, global, and depth-resolved volumetric features. The geometry factor models inter-observation coupling through coordinate-conditioned message passing over the queried receiver set, while the condition factor performs frequency-dependent aggregation of depth-resolved features. Their representations are coupled through a nonlinear fusion and residual readout to predict electromagnetic responses. The same factorization is instantiated for both dense-grid and coordinate-query outputs.

the proposed grouping on 12 MT backgrounds selected to span resistivity contrasts and anomaly depths. Sensitivities are estimated at five frequencies from 1 to 100 Hz and 64 spatially distributed receivers using centered block finite differences. Figure 2 compares the resulting spectra and rank-four captured energy against alternative variable groupings. MT finite-difference sensitivities retain $98.22\%$–$99.77\%$ energy at rank four under the proposed grouping. The dimension-matched control retains $86.56\%$–$94.15\%$. Thus the proposed grouping achieves a smaller measured approximation error at equal separation rank.

**Learning a shared operator.** Sharing geometric interactions lets supervision at different frequencies jointly constrain the same spatial representation. Condition-dependent depth aggregation retains the frequency-specific response structure. For a rank-$R$ local sensitivity matrix of size $a \times b$, with $R < \min(a, b)$, factorization reduces the effective degrees of freedom from $ab$ to $R(a + b - R)$. Under the Gaussian observation model, this reduction lowers the leading-order estimation error in the small-noise limit (Appendix C).

**Generalization across queries.** Coordinate-conditioned interactions reuse learned spatial relationships at new receiver locations. Frequency-conditioned aggregation reuses depth-resolved features with weights adapted to the requested frequency. Parameter sharing thus connects predictions across locations and frequencies, allowing supervision at observed queries to inform predictions at new queries. This transfer is evaluated on held-out frequencies, unseen receiver layouts, and varying query-set sizes in Section 4.3.

### 3.3 Geometry–Condition Factorized Architecture

**From local sensitivity to architecture.** Equation (7) motivates shared spatial processing across frequencies and frequency-dependent aggregation over depth. GCFNO adopts this organization in a learned volumetric representation: the encoder embeds material information into spatial features, receiver sampling and graph interactions form a shared geometry representation, and the condition pathway aggregates depth features before sampling them at each receiver. A nonlinear readout couples the resulting representations:

$$\mathcal{G}_\theta(\mathbf{X}, \mathcal{P}_q, c) = D_\theta(T_{\text{geo}}(\mathbf{X}, \mathcal{P}_q), T_{\text{cond}}(\mathbf{X}, \mathcal{P}_q, c)) \,. \tag{9}$$

The encoder, receiver sampling, and graph interactions jointly carry material information to receiver features. The condition pathway explicitly separates geometry and frequency–depth dependence in its latent feature access, as shown below. Figure 3 illustrates the MT instantiation; Appendix E.1 details both MT and EIT.

**Shared Volumetric Encoder.** Given a volumetric physical-property field $\mathbf{X}$, a shared 3-D encoder first produces a latent volume $\mathbf{H} = E_\theta(\mathbf{X})$. From this latent volume we derive three representations:

$$\mathbf{S} = P_s(\mathbf{H}), \qquad \mathbf{g} = P_g(\mathbf{H}), \qquad \mathbf{D} = P_d(\mathbf{H}). \tag{10}$$

The projections $P_s$, $P_g$, and $P_d$ produce complementary views: $\mathbf{S}$ initializes receiver features, $\mathbf{g}$ summarizes the sample's global state, and $\mathbf{D} = \{\mathbf{D}_d\}_{d=1}^{N_d}$ retains depth-organized information for condition-dependent aggregation. Each $\mathbf{D}_d \in \mathbb{R}^{C\times H'\times W'}$ preserves a horizontal feature map with $C$ channels.

**Geometry Factor.** The geometry factor treats queried sensors as graph nodes. Each node $i$ is associated with a coordinate $\mathbf{p}_i \in \mathcal{P}_q$ and a latent feature initialized by sampling $\mathbf{S}$ at $\mathbf{p}_i$ and incorporating the global context. Message passing refines these material-dependent receiver features through nearby observation nodes, with the resulting geometry representation shared across frequencies:

$$\mathbf{h}_i^{\ell+1} = \Phi_\ell\left(\mathbf{h}_i^\ell, \mathrm{Agg}_{j\in\mathcal{N}(i)}\,\kappa_\ell(\mathbf{r}_{ij}, \mathbf{h}_j^\ell)\right), \qquad \mathbf{r}_{ij} = [\mathbf{p}_i, \mathbf{p}_j, \mathbf{p}_i - \mathbf{p}_j, \|\mathbf{p}_i - \mathbf{p}_j\|_2]. \tag{11}$$

Here $\mathcal{N}(i)$ is the neighborhood of node $i$, Agg aggregates messages, and $\Phi_\ell$ updates the node feature. Absolute coordinates, relative displacement, and pairwise distance condition the learned kernel $\kappa_\ell$, allowing the geometry pathway to represent spatial relationships on both regular lattices and arbitrary receiver sets.

**Condition Factor.** The condition factor determines how a requested frequency accesses the depth-resolved latent volume. Let $E_f(f)$ denote the embedding of numerical log-frequency, $\mathbf{z}_d$ the encoded depth coordinate, and $\mathcal{S}_{\mathbf{p}_i}$ bilinear sampling at receiver $\mathbf{p}_i$. GCFNO aggregates depth features while preserving the horizontal grid, then samples the resulting map:

$$a_d(f) = \mathrm{softmax}_d\, A_\theta(E_f(f), \mathbf{z}_d), \qquad \mathbf{d}_{i,f} = \mathcal{S}_{\mathbf{p}_i}\left[\sum_{d=1}^{N_d} a_d(f)\mathbf{D}_d\right] \in \mathbb{R}^C. \tag{12}$$

At a fixed frequency, the weights are shared across samples and horizontal positions; sample-specific information enters through $\mathbf{D}_d$. Let $w_i(\eta)$ be the sampling weight at horizontal grid location $\eta$. For each channel $k$, the latent access operator satisfies

$$d_{i,f,k} = \sum_{\eta,d} w_i(\eta) a_d(f) D_{d,k}(\eta), \qquad \frac{\partial d_{i,f,k}}{\partial D_{d,k}(\eta)} = w_i(\eta) a_d(f). \tag{13}$$

Thus its kernel separates receiver–horizontal geometry $(i, \eta)$ from frequency–depth dependence $(f, d)$, matching the variable grouping in Eq. (6). This identity describes the latent feature access; the encoder and readout learn the nonlinear dependence on the material field.

**Interaction and Readout.** The geometry and condition pathways are coupled at the response readout. Let $\mathbf{h}_i = T_{\mathrm{geo}}(\mathbf{X}, \mathcal{P}_q)_i$ denote the shared geometry feature, $\mathbf{d}_{i,f} = T_{\mathrm{cond}}(\mathbf{X}, \mathcal{P}_q, f)_i$ the condition feature sampled at the same receiver, and $\mathbf{g}$ the global context. The fused prediction is

$$\widehat{\mathbf{y}}_i(f) = R_\theta([\mathbf{h}_i, \mathbf{d}_{i,f}, \mathbf{g}, E_f(f)])\,, \tag{14}$$

where $R_\theta$ combines nonlinear fusion, FiLM (Perez et al., 2018) modulation, layer normalization (Ba et al., 2016), and a residual readout. The detailed formulation is given in Appendix E.1. This readout couples receiver-specific geometry with condition-dependent volumetric information and supports prediction on both fixed receiver lattices and arbitrary coordinate queries through $\mathcal{P}_q$.

**Domain-specific instantiations.** The MT model uses receiver-graph interactions and vertical depth features. The EIT model applies geometric message passing to mesh cells and organizes condition features into boundary-distance bins; its electrode decoder is conditioned on frequency and current-injection pattern. Both implementations follow the same geometry–condition design (Appendix E.1).

## 4 Experiments

We evaluate GCFNO on two electromagnetic forward problems: three-dimensional magnetotellurics (MT) and head electrical impedance tomography (EIT). The MT benchmark contains heterogeneous subsurface resistivity models with localized anomalies relevant to geophysical and mineral exploration, while the EIT benchmark uses subject-specific head anatomies to model multi-frequency boundary responses under multiple current-injection patterns. These tasks span geological and biomedical

Table 1: **Predictive accuracy across two electromagnetic forward problems.** Results are reported as mean ± standard deviation over three training seeds on the held-out test sets.

| Model | 3-D MT | | | Head EIT | |
|---|---|---|---|---|---|
| | $\mathrm{RMSE}_{\mathrm{tot}} \downarrow$ | $\mathrm{RMSE}_{\log \rho_a} \downarrow$ | $\mathrm{RMSE}_{\phi} \downarrow$ | RMSE (mV)↓ | Rel. $L_2$ (%)↓ |
| 3D U-Net (Çiçek et al., 2016) | $0.04742 \pm 0.00463$ | $0.06546 \pm 0.00639$ | $(2.562 \pm 0.258)^\circ$ | $0.3899 \pm 0.0184$ | $2.470 \pm 0.110$ |
| FNO (Li et al., 2021) | $0.03178 \pm 0.01593$ | $0.04401 \pm 0.02203$ | $(1.637 \pm 0.840)^\circ$ | $0.3882 \pm 0.0181$ | $2.454 \pm 0.100$ |
| DeepONet (Lu et al., 2021) | $0.06860 \pm 0.00184$ | $0.09554 \pm 0.00260$ | $(3.005 \pm 0.045)^\circ$ | $0.6601 \pm 0.0159$ | $4.039 \pm 0.095$ |
| Perceiver IO (Jaegle et al., 2022) | $0.06043 \pm 0.00392$ | $0.08448 \pm 0.00562$ | $(2.234 \pm 0.119)^\circ$ | $0.4798 \pm 0.0958$ | $3.046 \pm 0.610$ |
| DeepSets (Zaheer et al., 2017) | $0.05732 \pm 0.00012$ | $0.07970 \pm 0.00017$ | $(2.636 \pm 0.003)^\circ$ | $0.4071 \pm 0.0194$ | $2.580 \pm 0.110$ |
| EFKAN (Wang et al., 2026a) | $0.04809 \pm 0.00230$ | $0.06712 \pm 0.00324$ | $(1.949 \pm 0.062)^\circ$ | $9.3627 \pm 0.0003$ | $60.484 \pm 0.002$ |
| GNO (Li et al., 2020) | $0.01907 \pm 0.00031$ | $0.02644 \pm 0.00044$ | $(0.939 \pm 0.016)^\circ$ | $0.3922 \pm 0.0287$ | $2.490 \pm 0.170$ |
| GCFNO | $\mathbf{0.01762 \pm 0.00041}$ | $\mathbf{0.02444 \pm 0.00058}$ | $(\mathbf{0.868 \pm 0.013})^\circ$ | $\mathbf{0.3735 \pm 0.0342}$ | $\mathbf{2.391 \pm 0.230}$ |

electromagnetic systems with distinct discretizations, observation geometries, acquisition protocols, and response variables. MT further enables controlled variation of frequency and receiver geometry for evaluating the two factors in GCFNO separately.

### 4.1 Experimental Setup

**MT.** The MT benchmark contains 900 synthetic $50 \times 50 \times 50$ resistivity volumes, split into 630/135/135 training, validation, and test samples. Responses are evaluated on a $50 \times 50$ surface receiver grid at 15 frequencies between 1 and 100 Hz and contain $\{\rho_{xy}, \rho_{yx}, \varphi_{xy}, \varphi_{yx}\}$. We report Total RMSE, $\log_{10} \rho_a$ RMSE, and phase RMSE. The baseline set includes 3D U-Net (Çiçek et al., 2016), FNO (Li et al., 2021), DeepONet (Lu et al., 2021), Perceiver IO (Jaegle et al., 2022), DeepSets (Zaheer et al., 2017), EFKAN (Wang et al., 2026a), and GNO (Li et al., 2020).

**EIT.** The head-EIT benchmark follows the synthetic construction released with BREIT (Abdelmoumene et al., 2026) and contains 193 subjects represented by approximately 76k finite-element cells, with 17 frequencies, 31 current-injection patterns, and 33 electrodes. The split contains 135/29/29 training, validation, and test subjects. We compare GCFNO with 3D U-Net, FNO, DeepONet, Perceiver IO, DeepSets, EFKAN, and GNO using patient-macro RMSE and relative $L_2$ error. Full data-generation, training, and evaluation protocols are provided in Appendix D and Appendix E.

### 4.2 Predictive Accuracy Across Electromagnetic Systems

GCFNO achieves the lowest three-seed mean prediction error among all evaluated baselines on both electromagnetic benchmarks. On fixed-grid MT, GCFNO obtains a three-seed mean total RMSE of $0.01762 \pm 0.00041$, compared with $0.01907 \pm 0.00031$ for GNO, corresponding to a $7.60\%$ reduction in the mean, with consistent improvements in both $\log_{10} \rho_a$ and phase. The frequency-resolved results in Fig. 4 further show lower error at all 15 supervised frequencies.

Head EIT provides a complementary evaluation under a substantially different sensing configuration, with subject-specific unstructured head geometries, multiple excitation frequencies, and multiple current-injection patterns. GCFNO attains the lowest three-seed mean patient-macro error, with an RMSE of $0.3735 \pm 0.0342$ mV and a relative $L_2$ error of $2.391 \pm 0.230\%$, compared with $0.3882 \pm 0.0181$ mV and $2.454 \pm 0.100\%$ for FNO. Figure 5 further shows that GCFNO closely reproduces the boundary-voltage responses at representative low and high frequencies and preserves the response structure across all 17 excitation frequencies. Together, the MT and EIT results show consistent predictive gains across electromagnetic systems with distinct discretizations, observation geometries, and acquisition protocols.

### 4.3 Generalization across Conditions and Geometry

**Condition generalization.** We evaluate generalization to unseen acquisition conditions by withholding selected frequency groups from supervision and testing only on the excluded frequencies. As shown in Fig. 6(a), a standard GNO with frequency-specific output channels (denoted GNO-fixed) yields a macro-averaged held-out-frequency RMSE of $2.419$. Explicit frequency conditioning (denoted GNO freq-only) reduces the error to $1.264$, while the structured condition pathway in GCFNO further lowers it to $0.546$. GCFNO improves three of the four completion settings, although

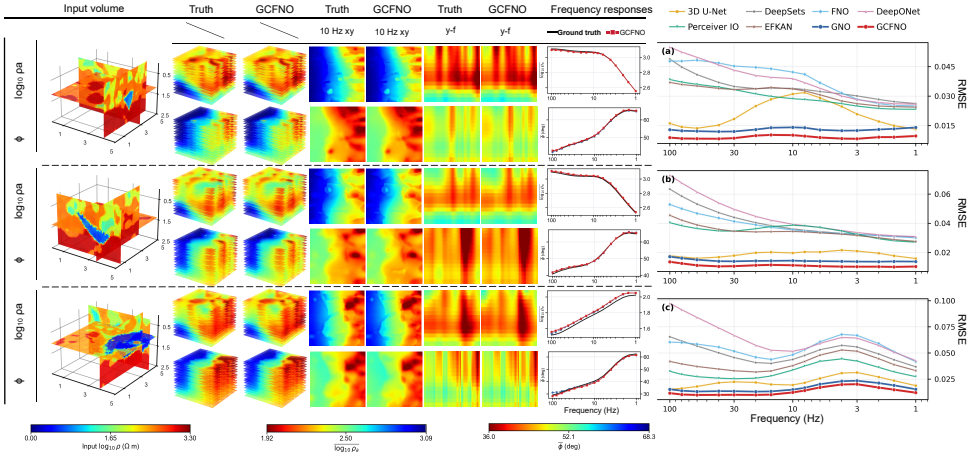


Figure 4: **Qualitative and frequency-resolved MT results.** Three held-out resistivity models are shown with ground-truth and GCFNO predictions in volumetric, spatial-slice, $y$–frequency, and frequency-response views. The corresponding panels on the right report per-frequency RMSE for all evaluated models. GCFNO accurately captures both spatial response structure and frequency-dependent behavior across the representative test cases.

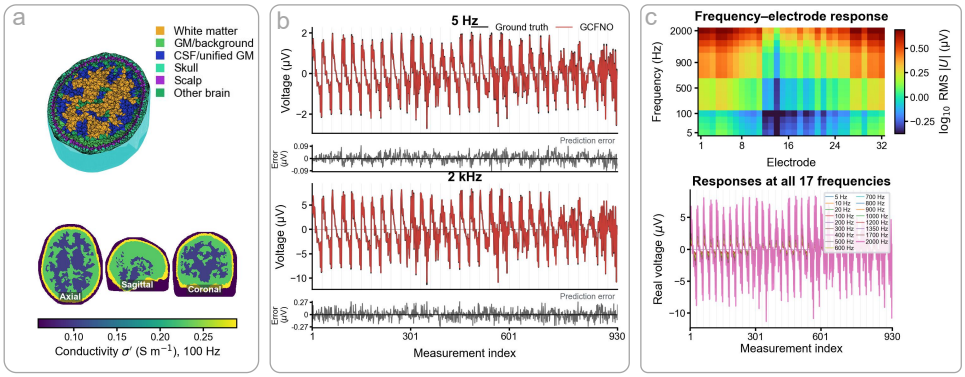


Figure 5: **Representative head-EIT forward prediction.** (a) Subject-specific head conductivity model and anatomical slices. (b) Ground-truth and GCFNO boundary voltages at 5 Hz and 2 kHz, together with the corresponding prediction errors. (c) Frequency–electrode response map and voltage traces across all 17 excitation frequencies.

low-to-high extrapolation remains the most difficult, with an RMSE of $0.563$ compared with $0.478$ for direct frequency conditioning.

**Geometry generalization.** Figure 6(b) evaluates geometry generalization by training query models on a regular $25 \times 25$ receiver grid and testing them on four coordinate-disjoint 625-station layouts selected from the solver-generated $50 \times 50$ station field. GCFNO improves over GNO on all four unseen layouts. The geometry pathway also remains stable under substantial changes in query-set size: when the number of requested receivers varies from 144 to 10,000, GCFNO maintains an RMSE of $0.0243$–$0.0253$ and reduces the average error by $9.18\%$ relative to GNO.

### 4.4 Data Efficiency

We evaluate GCFNO under reduced training data by subsampling the MT training set to $5\%$, $10\%$, $20\%$, $50\%$, and $100\%$, while keeping the validation and test sets fixed. To summarize performance across data regimes, we report the normalized area under the error curve on the logarithmic training-

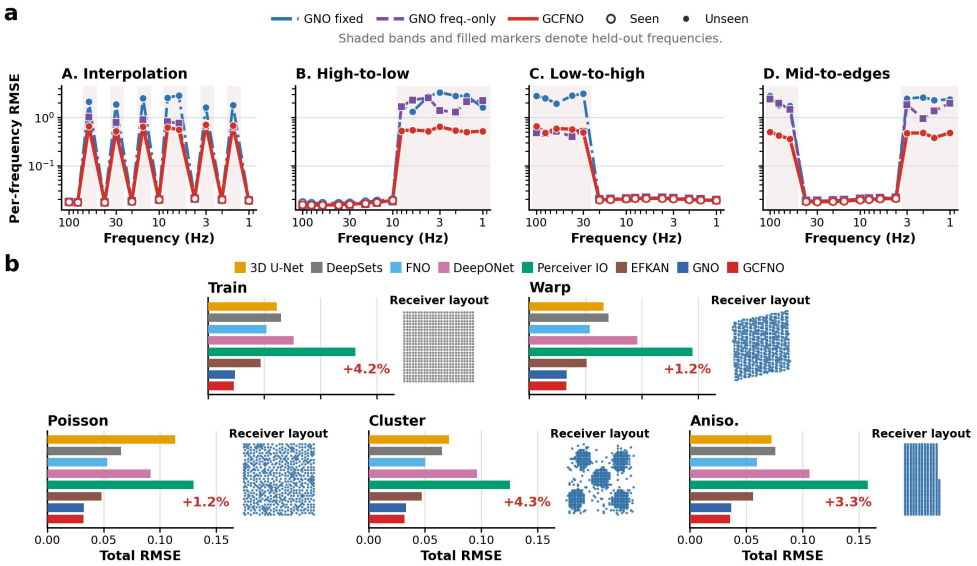


Figure 6: **Generalization across conditions and receiver geometries.** (a) Per-frequency RMSE for four held-out-frequency settings. (b) Total RMSE on the training layout and four coordinate-disjoint receiver layouts, with annotations showing GCFNO versus GNO.

Table 2: **Data efficiency on MT.** Total RMSE at selected training-data fractions and normalized area under the error curve on the logarithmic data-fraction scale. Lower is better.

| Model | 5% | 20% | 100% | AUC↓ |
|---|---|---|---|---|
| 3D U-Net (Çiçek et al., 2016) | 0.07415 | 0.05376 | 0.04219 | 0.05446 |
| FNO (Li et al., 2021) | 0.08214 | 0.05669 | 0.04421 | 0.05945 |
| DeepONet (Lu et al., 2021) | 0.08427 | 0.07915 | 0.07247 | 0.07863 |
| GNO (Li et al., 2020) | 0.04711 | 0.03130 | 0.01953 | 0.03184 |
| GCFNO | **0.04419** | **0.02897** | **0.01853** | **0.02952** |

data scale,

$$\mathrm{AUC}_{\log}(E) = \frac{1}{\log p_{\max} - \log p_{\min}} \int_{\log p_{\min}}^{\log p_{\max}} E(e^u)\,\mathrm{d}u, \tag{15}$$

where $p$ denotes the training-data ratio and $E(p)$ the corresponding test error. Lower $\mathrm{AUC}_{\log}$ indicates better performance across the full range of training-set sizes.

As shown in Table 2, GCFNO consistently outperforms GNO across the reported data fractions. With only $5\%$ of the training data, GCFNO reaches a total RMSE of $0.044185$, comparable to full-data FNO at $0.044212$. GCFNO also achieves the lowest aggregate $\mathrm{AUC}_{\log}$, indicating a consistently better data–accuracy tradeoff.

## 5 Discussion and Conclusion

GCFNO organizes electromagnetic operator learning around shared geometric interactions and condition-dependent access to volumetric features. Local approximation and estimation results, together with solver sensitivity spectra, explain how this structure can preserve response information with fewer effective degrees of freedom. Experiments show improved predictive accuracy on MT and head EIT, with additional gains in frequency and geometry generalization and data efficiency on MT. These findings support sensitivity-guided factorization as an inductive bias for learning reusable electromagnetic response mappings from limited simulations. More broadly, this sensitivity-guided approach can inform the design of problem-specific neural operators for related physical systems.

## AI USE STATEMENT

Generative AI tools assisted with manuscript organization, language editing, LaTeX preparation. The authors take responsibility for the scientific content and verification of the final manuscript.

## REFERENCES


Djahid Abdelmoumene, Ishak Ayad, Maï K. Nguyen, and Christian Daveau. Breit: A framework for brain stroke reconstruction using multi-frequency 3d eit, 2026. URL https://arxiv.org/abs/2606.28787.

Andy Adler and Alistair Boyle. Electrical impedance tomography: Tissue properties to image measures. *IEEE Transactions on Biomedical Engineering*, 64(11):2494–2504, 2017. doi: 10.1109/TBME.2017.2728323.

Benedikt Alkin, Andreas Fürst, Simon Schmid, Lukas Gruber, Markus Holzleitner, and Johannes Brandstetter. Universal physics transformers: A framework for efficiently scaling neural operators. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 25152–25194. Curran Associates, Inc., 2024. doi: 10.52202/079017-0793. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/2cd36d327f33d47b372d4711edd08de0-Paper-Conference.pdf.

Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E. Hinton. Layer normalization. arXiv:1607.06450, 2016. URL https://arxiv.org/abs/1607.06450.

Valentina Candiani and Matteo Santacesaria. Neural networks for classification of strokes in electrical impedance tomography on a 3d head model. *Mathematics in Engineering*, 4(4):1–22, 2022. ISSN 2640-3501. doi: 10.3934/mine.2022029. URL https://www.aimspress.com/article/doi/10.3934/mine.2022029.

Emmanuel J. Candès and Yaniv Plan. Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Transactions on Information Theory*, 57 (4):2342–2359, 2011. doi: 10.1109/TIT.2011.2111771.

Alan D. Chave and Alan G. Jones (eds.). *The Magnetotelluric Method: Theory and Practice*. Cambridge University Press, Cambridge, 2012.

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 2016*, volume 9901 of *Lecture Notes in Computer Science*, pp. 424–432. Springer, 2016. doi: 10.1007/978-3-319-46723-8_49.

Marion Darbas, Jérémy Heleine, Renier Mendoza, and Arrianne Crystal Velasco. Sensitivity analysis of the complete electrode model for electrical impedance tomography. *AIMS Mathematics*, 6 (7):7333–7366, 2021. ISSN 2473-6988. doi: 10.3934/math.2021431. URL https://www.aimspress.com/article/doi/10.3934/math.2021431.

Gary D. Egbert and Anna Kelbert. Computational recipes for electromagnetic inverse problems. *Geophysical Journal International*, 189(1):251–267, 04 2012. ISSN 0956-540X. doi: 10.1111/j.1365-246X.2011.05347.x. URL https://doi.org/10.1111/j.1365-246X.2011.05347.x.

Nir Goren, James Avery, Thomas Dowrick, Eleanor Mackle, Anna Witkowska-Wrobel, David Werring, and David Holder. Multi-frequency electrical impedance tomography and neuroimaging data in stroke patients. *Scientific Data*, 5(1):180112, 2018. ISSN 2052-4463. doi: 10.1038/sdata.2018.112. URL https://doi.org/10.1038/sdata.2018.112.

Maximilian Herde, Bogdan Raonić, Tobias Rohner, Roger Käppeli, Roberto Molinaro, Emmanuel de Bézenac, and Siddhartha Mishra. Poseidon: Efficient foundation models for pdes, 2024. URL https://arxiv.org/abs/2405.19101.

Benjamin Holzschuh, Qiang Liu, Georg Kohl, and Nils Thuerey. Pde-transformer: Efficient and versatile transformers for physics simulations, 2025. URL https://arxiv.org/abs/2505.24717.

Peter J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35 (1):73–101, 1964. doi: 10.1214/aoms/1177703732.

Andrew Jaegle, Sebastian Borgeaud, Jean-Baptiste Alayrac, Carl Doersch, Catalin Ionescu, David Ding, Skanda Koppula, Daniel Zoran, Andrew Brock, Evan Shelhamer, Olivier Hénaff, Matthew M. Botvinick, Andrew Zisserman, Oriol Vinyals, and João Carreira. Perceiver io: A general architecture for structured inputs & outputs, 2022. URL https://arxiv.org/abs/2107.14795.

Cheng Jing, Uvini Balasuriya Mudiyanselage, Abhishek Verma, Kallol Bera, Shahid Rauf, and Kookjin Lee. Extending fourier neural operators for modeling parameterized and coupled pdes, 2026. URL https://arxiv.org/abs/2607.23466.

Anna Kelbert, Paul A. Bedrosian, Adam Schultz, Gary D. Egbert, Louise Pellerin, Jeffrey J. Love, Andrew Frassetto, and Benjamin S. Murphy. The united states magnetotelluric array and the national impedance map. *Reviews of Geophysics*, 64(2):e2024RG000850, 2026. doi: https://doi.org/10.1029/2024RG000850. URL https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2024RG000850. e2024RG000850 2024RG000850.

Vladimir Koltchinskii, Karim Lounici, and Alexandre B. Tsybakov. Nuclear norm penalization and optimal rates for noisy low rank matrix completion. *The Annals of Statistics*, 39(5):2302–2329, 2011. ISSN 00905364, 21688966. URL http://www.jstor.org/stable/41713579.

Nikola Kovachki, Zongyi Li, Burigede Liu, Kamyar Azizzadenesheli, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces with applications to pdes. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL http://jmlr.org/papers/v24/21-1524.html.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Graph kernel network for partial differential equations, 2020. URL https://arxiv.org/abs/2003.03485.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations, 2021. URL https://arxiv.org/abs/2010.08895.

Zongyi Li, Nikola Borislavov Kovachki, Chris Choy, Boyi Li, Jean Kossaifi, Shourya Prakash Otta, Mohammad Amin Nabian, Maximilian Stadler, Christian Hundt, Kamyar Azizzadenesheli, and Anima Anandkumar. Geometry-informed neural operator for large-scale 3d pdes, 2023. URL https://arxiv.org/abs/2309.00583.

Weiyang Liao, Ronghua Peng, Xiangyun Hu, Yue Zhang, Wenlong Zhou, Xiaonian Fu, and Haikun Lin. Fast forward modeling of magnetotelluric data in complex continuous media using an extended fourier deeponet architecture. *Geophysics*, 90(2):F11–F25, 02 2025. ISSN 0016-8033. doi: 10.1190/geo2023-0613.1. URL https://doi.org/10.1190/geo2023-0613.1.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, March 2021. ISSN 2522-5839. doi: 10.1038/s42256-021-00302-5. URL http://dx.doi.org/10.1038/s42256-021-00302-5.

Sudeb Majee, Anuj Abhishek, Thilo Strauss, and Taufiquar Khan. Mcmc-net: accelerating markov chain monte carlo with neural networks for inverse problems. *Inverse Problems*, 41(9):095013, Sept 2025. ISSN 1361-6420. doi: 10.1088/1361-6420/ae05c2. URL http://dx.doi.org/10.1088/1361-6420/ae05c2.

Barry McDermott, Adnan Elahi, Adam Santorelli, Martin O'Halloran, James Avery, and Emily Porter. Multi-frequency symmetry difference electrical impedance tomography with machine learning for human stroke diagnosis. *Physiological Measurement*, 41(7):075010, aug 2020. doi: 10.1088/1361-6579/ab9e54. URL https://doi.org/10.1088/1361-6579/ab9e54.

Zhong Peng, Bo Yang, Yixian Xu, Feng Wang, Lian Liu, and Yi Zhang. Rapid surrogate modeling of electromagnetic data in frequency domain using neural operator. *IEEE Transactions on Geoscience and Remote Sensing*, 60:1–12, 2022. doi: 10.1109/TGRS.2022.3222507.

Ethan Perez, Florian Strub, Harm de Vries, Vincent Dumoulin, and Aaron Courville. FiLM: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, pp. 3942–3951, 2018. doi: 10.1609/aaai.v32i1.11671.

Ce Qin and Alexander Grayver. Efficient and scalable finite-element magnetotelluric modelling on high-order meshes. *Geophysical Journal International*, 244(2):ggaf471, 02 2026. ISSN 1365-246X. doi: 10.1093/gji/ggaf471. URL https://doi.org/10.1093/gji/ggaf471.

Alasdair Tran, Alexander Mathews, Lexing Xie, and Cheng Soon Ong. Factorized fourier neural operators, 2023. URL https://arxiv.org/abs/2111.13802.

P.J. Vauhkonen, M. Vauhkonen, T. Savolainen, and J.P. Kaipio. Three-dimensional electrical impedance tomography based on the complete electrode model. *IEEE Transactions on Biomedical Engineering*, 46(9):1150–1160, 1999. doi: 10.1109/10.784147.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018. ISBN 978-1-108-41519-4. doi: 10.1017/9781108231596.

Feng Wang, Hong Qiu, Yingying Huang, Xiaozhe Gu, Renfang Wang, and Bo Yang. Efkan: A kan-integrated neural operator for efficient magnetotelluric forward modeling. *Computers & Geosciences*, 207:106052, 2026a. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2025.106052. URL https://www.sciencedirect.com/science/article/pii/S009830042500202X.

Feng Wang, Hong Qiu, Yingying Huang, Xiaozhe Gu, Renfang Wang, and Bo Yang. Efkan: A kan-integrated neural operator for efficient magnetotelluric forward modeling. *Computers & Geosciences*, 207:106052, 2026b. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2025.106052. URL https://www.sciencedirect.com/science/article/pii/S009830042500202X.

Xuben Wang, Peifan Jiang, Fei Deng, Shuang Wang, Rui Yang, and Chongxin Yuan. Three-dimensional magnetotelluric forward modeling through deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024. doi: 10.1109/TGRS.2024.3401587.

Xinling Yu, Sean Hooten, Ziyue Liu, Yequan Zhao, Marco Fiorentino, Thomas Van Vaerenbergh, and Zheng Zhang. Separable operator networks, 2024. URL https://arxiv.org/abs/2407.11253.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabás Póczos, Ruslan Salakhutdinov, and Alexander J. Smola. Deep sets. In *Advances in Neural Information Processing Systems*, volume 30, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/f22e4747da1aa27e363d86d40ff442fe-Abstract.html.

Xin Zhong, Weiwei Ling, Kejia Pan, Chaofei Liu, Pinxia Wu, Jiajing Zhang, Zhiliang Zhan, and Wenbo Xiao. A neural network architecture based on attention gate mechanism for 3d magnetotelluric forward modeling. *Computers & Geosciences*, 207:106079, 2026. ISSN 0098-3004. doi: https://doi.org/10.1016/j.cageo.2025.106079. URL https://www.sciencedirect.com/science/article/pii/S0098300425002298.

Guancheng Zhou, Zelin Zeng, Yisi Luo, Qi Xie, and Deyu Meng. Tucker-FNO: Tensor tucker-fourier neural operator and its universal approximation theory. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=UJvkXnuozY.

## SUPPLEMENTAL MATERIAL

Appendix A derives the electromagnetic sensitivities. Appendix B proves the approximation results, and Appendix C analyzes structured local estimation. Appendices D–E document the datasets, metrics, implementations, and training protocols. Appendices F–H provide full generalization results, data-efficiency and robustness analyses, and EIT diagnostics.

## A PHYSICAL MODELS AND SENSITIVITY DERIVATIONS

### A.1 MAGNETOTELLURIC FIELDS AND RESPONSE SENSITIVITIES

We use the $e^{i\omega t}$ convention, with $\omega = 2\pi f$, and neglect displacement current in the MT model. Let $V$ be the energy space for homogeneous field variations. The source and boundary data are fixed with respect to conductivity, and the total electric field satisfies

$$a_\sigma(E, v) := \int_\Omega \mu^{-1}(\nabla \times E) \cdot \overline{\nabla \times v} + i\omega\sigma E \cdot \overline{v}\, dx = b_\omega(v), \qquad v \in V. \tag{16}$$

For prescribed tangential boundary data, $E$ belongs to a fixed affine translate of $V = H_0(\mathrm{curl}; \Omega)$. Assume that the induced operator $\mathcal{A}_\sigma : V \to V^*$ has a bounded inverse and that $h \in L^\infty(\Omega)$ defines a bounded multiplication operator $M_h$. These assumptions hold, for example, on a bounded Lipschitz domain with this choice of $V$, real coefficients $0 < \mu_{\min} \le \mu \le \mu_{\max}$ and $0 < \sigma_{\min} \le \sigma \le \sigma_{\max}$, and $\omega > 0$: multiplication of $a_\sigma(v, v)$ by $e^{-i\pi/4}$ gives a coercive real part in the $H(\mathrm{curl})$ norm.

**Lemma 2** (Local Maxwell sensitivity)**.** *Under these assumptions, the field is Fréchet differentiable with respect to conductivity. For a bounded complex-linear observation $\ell_p$, define the adjoint field $E^*_{p,\omega} \in V$ by $a_\sigma(v, E^*_{p,\omega}) = \ell_p(v)$ for all $v \in V$. Then*

$$DE_\sigma[h] = -i\omega\mathcal{A}_\sigma^{-1} M_h E, \qquad D\ell_p(E_\sigma)[h] = -i\omega \int_\Omega hE \cdot \overline{E^*_{p,\omega}}\, dx. \tag{17}$$

*Proof.* Subtracting the weak equations for $\sigma + h$ and $\sigma$ yields $E_{\sigma+h} - E_\sigma = -i\omega\mathcal{A}_\sigma^{-1} M_h E_{\sigma+h}$. Set $T_h = i\omega\mathcal{A}_\sigma^{-1} M_h$, regarded as an operator on the total-field energy space. If $q_h = \|T_h\| < 1$, the Neumann expansion gives

$$E_{\sigma+h} = (I + T_h)^{-1} E_\sigma, \qquad \|E_{\sigma+h} - E_\sigma + T_h E_\sigma\| \le \frac{q_h^2}{1 - q_h} \|E_\sigma\|. \tag{18}$$

Since $q_h = O(\|h\|_{L^\infty})$, the linear term is the Fréchet derivative. Testing its weak equation with $E^*_{p,\omega}$ proves the observation formula. □

Equation (17) gives the forward–adjoint sensitivity representation used in electromagnetic inversion (Egbert & Kelbert, 2012). Observations can be bounded sensor averages, finite-dimensional solver outputs, or point evaluations on a solution space with the required regularity.

**Input and output transformations.** For the MT input $m = \log_{10} \rho$, conductivity is $\sigma = 10^{-m}$, so $\delta\sigma = -(\ln 10)\sigma\, \delta m$. The corresponding input kernel is therefore $K_m = -(\ln 10)\sigma K_\sigma$. Let $E_t, H_t \in \mathbb{C}^{2\times 2}$ collect the tangential fields from two independent incident polarizations. If $H_t$ is invertible, the impedance tensor and its variation are

$$Z = E_t H_t^{-1}, \qquad \delta Z = (\delta E_t - Z\delta H_t) H_t^{-1}. \tag{19}$$

For a nonzero component $Z_{ab}$, set $\rho_{a,ab} = |Z_{ab}|^2/(\mu_0\omega)$ and choose a locally continuous phase. The response coordinates used in the sensitivity analysis satisfy

$$\delta \log_{10} \rho_{a,ab} = \frac{2}{\ln 10} \mathrm{Re}\, \frac{\delta Z_{ab}}{Z_{ab}}, \qquad \delta(\varphi_{ab,\deg}/180) = \frac{1}{\pi} \mathrm{Im}\, \frac{\delta Z_{ab}}{Z_{ab}}. \tag{20}$$

Thus the measured Jacobian includes both the material transformation and the response transformation.

**Reference depth scale.** In a homogeneous conductive half-space under plane-wave excitation, the tangential field satisfies $\partial_z^2 E = i\omega\mu\sigma_0 E$. Its decaying solution is

$$E(z,\omega) = E(0,\omega)e^{-(1+i)z/\delta(\omega)}, \qquad \delta(\omega) = \sqrt{\frac{2}{\omega\mu\sigma_0}}. \tag{21}$$

The reference propagation length therefore scales as $f^{-1/2}$, providing a frequency-dependent scale for organizing depth-resolved features.

### A.2 EIT WITH THE COMPLETE ELECTRODE MODEL

Let $\Omega$ be a connected bounded Lipschitz domain, with disjoint electrodes $e_1, \dots, e_{N_e}$ of positive surface measure. At frequency $f$, let $\gamma = \sigma + i\omega\varepsilon \in L^\infty(\Omega)$ satisfy $\operatorname{Re}\gamma \geq \gamma_* > 0$. Contact impedances $\zeta_l$ are fixed with respect to material perturbations and satisfy $\operatorname{Re}(\zeta_l^{-1}) > 0$. For an applied current vector $I_s \in \mathbb{C}^{N_e}$ with $\sum_l (I_s)_l = 0$, the complete electrode model (CEM) determines the interior potential $u_s$ and electrode voltages $U_s$ (Vauhkonen et al., 1999). On the gauge-fixed space $\mathcal{V} = \{(v, V) \in H^1(\Omega) \times \mathbb{C}^{N_e} : \sum_l V_l = 0\}$, its weak formulation is

$$\begin{aligned} a_\gamma((u_s, U_s), (v, V)) &:= \int_\Omega \gamma \nabla u_s \cdot \overline{\nabla v}\, dx + \sum_{l=1}^{N_e} \zeta_l^{-1} \int_{e_l} (u_s - (U_s)_l)\overline{(v - V_l)}\, dS \\ &= \sum_{l=1}^{N_e} (I_s)_l \overline{V_l}, \qquad (v, V) \in \mathcal{V}. \end{aligned} \tag{22}$$

The stated positivity conditions give coercivity on $\mathcal{V}$. For a complex admittivity perturbation $h \in L^\infty(\Omega)$, differentiation at fixed electrodes, currents, and contact impedances gives

$$a_\gamma((\delta u_s, \delta U_s), (v, V)) = -\int_\Omega h \nabla u_s \cdot \overline{\nabla v}\, dx. \tag{23}$$

The bounded-inverse argument in Lemma 2, with the gradient multiplier replacing $i\omega M_h$, establishes Fréchet differentiability (Darbas et al., 2021). For a voltage measurement $y_{s,q} = q^* U_s$, where $\sum_l q_l = 0$, define the adjoint pair by $a_\gamma((v, V), (v_q^*, V_q^*)) = q^* V$ for all $(v, V) \in \mathcal{V}$. Testing Eq. (23) with this pair yields

$$D_\gamma y_{s,q}[h] = -\int_\Omega h \nabla u_s \cdot \overline{\nabla v_q^*}\, dx. \tag{24}$$

Electrode geometry enters through the domain and boundary form, while frequency and injection pattern enter through $\gamma(f)$, the contact impedances, and $I_s$.

## B APPROXIMATION RESULTS AND PROOFS

### B.1 LOCAL FACTORIZATION OF THE DISCRETE RESPONSE

*Proof of Proposition 1.* Let $J = DF(\mathbf{X}_0)$ and $K_{(p,\xi),(f,z)} = J_{pf,\xi z}$. Write the truncated SVD as $K_R = \sum_{r=1}^R s_r u_r v_r^\top$, and reshape $u_r, v_r$ into $A_r \in \mathbb{R}^{N_p \times N_\xi}$ and $B_r \in \mathbb{R}^{N_f \times N_z}$. Contracting the input indices gives

$$(J_R[\Delta\mathbf{X}])_{pf} = \sum_{\xi,z} \sum_{r=1}^R s_r (A_r)_{p\xi} (B_r)_{fz} \Delta X_{\xi z} = \left[\sum_{r=1}^R s_r A_r \Delta\mathbf{X} B_r^\top\right]_{pf}, \tag{25}$$

which proves Eq. (7). The unfolding preserves the Frobenius norm. Applying Cauchy–Schwarz over $(\xi, z)$ for each $(p, f)$ and summing gives

$$\|(J - J_R)[\Delta\mathbf{X}]\|_F \leq \|K - K_R\|_F \|\Delta\mathbf{X}\|_F = \left(\sum_{r>R} s_r^2\right)^{1/2} \|\Delta\mathbf{X}\|_F. \tag{26}$$

The segment $\mathbf{X}_0 + t\Delta\mathbf{X}$, $0 \leq t \leq 1$, lies in the neighborhood where the second derivative bound holds. Taylor's integral remainder is

$$F(\mathbf{X}_0 + \Delta\mathbf{X}) - F(\mathbf{X}_0) - J[\Delta\mathbf{X}] = \int_0^1 (1-t) D^2 F(\mathbf{X}_0 + t\Delta\mathbf{X})[\Delta\mathbf{X}, \Delta\mathbf{X}]\, \mathrm{d}t, \tag{27}$$

whose norm is at most $L\|\Delta\mathbf{X}\|_F^2/2$. Adding this bound to Eq. (26) proves Eq. (8). □

**Solver sensitivity diagnostic.** We select 12 held-out MT backgrounds spanning resistivity contrasts and anomaly depths, with sample selection fixed before spectral analysis. Each background is evaluated at frequencies $\{100, 30, 10, 3, 1\}$ Hz and 64 spatially distributed receivers. The material field is partitioned into four horizontal blocks and four depth slabs. For each block direction $V_{\xi,z}$, we estimate the signed sensitivity by centered finite differences,

$$K^{(h)}_{p,f,\xi,z} = \frac{F_{p,f}(X_0 + hV_{\xi,z}) - F_{p,f}(X_0 - hV_{\xi,z})}{2h}, \qquad h = 0.03, \tag{28}$$

where $X_0 = \log_{10}\rho$ and $V_{\xi,z}$ equals one within the selected block and zero elsewhere. The four response components use the transformations specified in Appendix A.

The proposed $(p,\xi) \mid (f,z)$ unfolding and the dimension-matched $(p,z) \mid (\xi,f)$ control are both $256 \times 20$. The $(p,f) \mid (\xi,z)$ control is $320 \times 16$. We compute the singular values separately for each background and response component, without centering. For nonzero $K$, define

$$e_r = \frac{s_r^2}{\sum_j s_j^2}, \qquad E_R = \sum_{r=1}^{R} e_r, \qquad R_{95} = \min\{R : E_R \geq 0.95\}. \tag{29}$$

Normalized spectra and captured energies are averaged over the four components within each background. Comparisons between groupings are paired by background.

### B.2 Reference symmetry and finite-rank approximation

The following results give sufficient conditions for a compact geometry–condition kernel in a horizontally invariant reference model.

**Lemma 3** (Translation symmetry of the reference sensitivity)**.** *Assume that the background material depends only on depth, the horizontal domain and boundary conditions are invariant under translations, and the excitation is horizontally uniform. Assume a unique forward solution, a differentiable response, and receiver observations that translate with the receiver. If its sensitivity has an integral kernel $K_0(p,\xi,z,f)$, then*

$$K_0(p+a,\xi+a,z,f) = K_0(p,\xi,z,f), \qquad K_0(p,\xi,z,f) = \kappa(p-\xi,z,f) \tag{30}$$

*almost everywhere.*

*Proof.* Let $(T_a h)(\xi,z) = h(\xi - a, z)$. Translating the perturbed boundary-value problem and using uniqueness gives $DF(\mathbf{X}_0)[T_a h](p+a,f) = DF(\mathbf{X}_0)[h](p,f)$. Substitution of the integral representation and the change of variable $\xi \mapsto \xi + a$ prove the first equality for all admissible $h$. Invariance under diagonal translations gives dependence on the difference $p - \xi$. □

**Proposition 4** (Separation from transverse regularity)**.** *Let $p, \xi \in \mathbb{T}^2$ with normalized Haar measure, and let $Z$ and $I$ be bounded depth and frequency intervals. Suppose $K_0(p,\xi,z,f) = \kappa(p-\xi,z,f)$ and*

$$\kappa(r,z,f) = \sum_{n\in\mathbb{Z}^2} a_n(z,f)e^{2\pi i n\cdot r}, \qquad M_s^2 := \sum_n (1+|2\pi n|^2)^s \|a_n\|^2_{L^2(Z\times I)} < \infty \tag{31}$$

*for some $s > 0$. For every $Q \geq 0$, there is a kernel $K_Q$ whose separation rank across $(p,\xi) \mid (f,z)$ is at most $N(Q) = \#\{n : |2\pi n| \leq Q\} = O((1+Q)^2)$ and such that*

$$\|K_0 - K_Q\|_{L^2((\mathbb{T}^2)^2\times Z\times I)} \leq M_s(1+Q^2)^{-s/2}. \tag{32}$$

*Proof.* Retain the modes $|2\pi n| \leq Q$. Each term is $e^{2\pi i n\cdot(p-\xi)}a_n(z,f)$, the product of a geometry factor and a frequency–depth factor, so the separation rank is at most $N(Q)$. Translation invariance of the normalized measure and Parseval's identity give

$$\|K_0 - K_Q\|^2_{L^2} = \sum_{|2\pi n|>Q} \|a_n\|^2_{L^2(Z\times I)} \leq (1+Q^2)^{-s} M_s^2. \tag{33}$$

The number of lattice points in a disk is $O((1+Q)^2)$. For real kernels, pairing conjugate modes gives a real sine–cosine expansion with the same rank bound. □

**Corollary 5** (Perturbation of the reference kernel)**.** *Under Proposition 4, if another kernel $K_{\mathbf{X}}$ satisfies $\|K_{\mathbf{X}} - K_0\|_{L^2} \le \eta$, then*

$$\inf_{\mathrm{rank}_{\mathrm{sep}}(L) \le N(Q)} \|K_{\mathbf{X}} - L\|_{L^2} \le \eta + M_s(1+Q^2)^{-s/2}. \tag{34}$$

*Proof.* Choose $L = K_Q$ and apply the triangle inequality and Eq. (32). □

### B.3 Preserving condition-dependent depth information

**Proposition 6** (Depth compression and exact condition-dependent readout)**.** *Let $x \in \mathbb{R}^{N_z}$ denote a depth profile and let its linear response at frequency $f \in I$ be $w(f)^\top x$. For a fixed linear compression $P \in \mathbb{R}^{d_c \times N_z}$, a decoder $g$ satisfying $g(Px, f) = w(f)^\top x$ for every $x, f$ exists if and only if*

$$w(f) \in \mathrm{range}(P^\top) \quad \textit{for all } f \in I. \tag{35}$$

*Consequently, exact recovery requires $d_c \ge \dim \mathrm{span}\{w(f) : f \in I\}$. Moreover, every decoder satisfies*

$$\sup_{f \in I,\, \|x\|_2 \le 1} |g(Px, f) - w(f)^\top x| \ge \sup_{f \in I} \|\Pi_{\ker P} w(f)\|_2, \tag{36}$$

*where $\Pi_{\ker P}$ is the Euclidean orthogonal projection.*

*Proof.* If $h \in \ker P$, exact recovery from $Px = P(x+h)$ implies $w(f)^\top h = 0$ for every $f$. Thus $w(f) \in (\ker P)^\perp = \mathrm{range}(P^\top)$. Conversely, write $w(f) = P^\top a(f)$ and set $g(y, f) = a(f)^\top y$. This proves Eq. (35) and the dimension bound. For Eq. (36), fix $f$ with nonzero projected weight and take $h = \Pi_{\ker P} w(f) / \|\Pi_{\ker P} w(f)\|_2$. The profiles $h$ and $-h$ both encode to zero and have target values $\pm\|\Pi_{\ker P} w(f)\|_2$. At least one incurs this absolute error. Taking the supremum over $f$ completes the proof; frequencies with zero projection contribute a zero lower bound. □

Proposition 6 characterizes the information that a depth representation must preserve for frequency-dependent readout. Keeping depth-resolved features makes these condition-specific linear functionals available before compression.

## C Conditional Statistical Benefits of Structured Local Estimation

Low-rank estimation balances approximation bias and estimation variance (Koltchinskii et al., 2011; Candès & Plan, 2011). We quantify this balance for geometry–condition unfoldings, compare variable partitions, and relate local operator estimation to prediction.

### C.1 Observation Model

Fix a background $\mathbf{X}_0$ and one response component. In vector coordinates, write its derivative as $J = DF(\mathbf{X}_0) \in \mathbb{R}^{q \times d}$, where $d = N_\xi N_z$ and $q = N_p N_f$. The geometry–condition unfolding is $K = \mathcal{U}_{\mathrm{gc}}(J) \in \mathbb{R}^{a \times b}$, with $a = N_p N_\xi$, $b = N_f N_z$, and $ab = qd$. Reshaping preserves the Frobenius norm. We consider

$$Z = K + \tau W, \qquad W_{ij} \overset{\mathrm{iid}}{\sim} \mathcal{N}(0,1), \qquad \tau > 0. \tag{37}$$

The unrestricted estimator $\widehat{K}_{\mathrm{full}} = Z$ has exact risk

$$\mathbb{E}\|\widehat{K}_{\mathrm{full}} - K\|_F^2 = \tau^2 ab. \tag{38}$$

Equation (37) also follows from balanced multi-output linear regression. Let $y_i = Jh_i + \varepsilon_i$, where the design vectors $h_i \in \mathbb{R}^d$ are fixed, $N \ge d$, $\sum_{i=1}^N h_i h_i^\top = N I_d$, and $\varepsilon_i \sim \mathcal{N}(0, \sigma_\varepsilon^2 I_q)$ independently across samples. Then

$$\widetilde{J} = \frac{1}{N} \sum_{i=1}^N y_i h_i^\top = J + \frac{\sigma_\varepsilon}{\sqrt{N}} G, \qquad G_{kl} \overset{\mathrm{iid}}{\sim} \mathcal{N}(0,1). \tag{39}$$

Indeed, each noise row has covariance $(\sigma_\varepsilon^2/N)I_d$, and the rows are independent. Unfolding $\widetilde{J}$ therefore gives Eq. (37) with $\tau^2 = \sigma_\varepsilon^2/N$.

### C.2 Estimating Unknown Geometry and Condition Factors

Fix an integer $1 \le R \le \min(a, b)$. Let $K_R$ and $\widehat{K}_R$ be the best rank-at-most-$R$ approximations of $K$ and $Z$, respectively. Define

$$\beta_R^2 = \|K - K_R\|_F^2 = \sum_{j>R} s_j(K)^2, \qquad c_{ab}^2 = (\sqrt{a} + \sqrt{b})^2 + 1, \tag{40}$$

where $s_j(K)$ are the singular values in decreasing order.

**Proposition 7** (Risk of structured local estimation)**.** *Under Eq.* (37)*,*

$$\mathbb{E}\|\widehat{K}_R - K\|_F^2 \le B_R := \left(\beta_R + \tau\sqrt{8R}\, c_{ab}\right)^2 \le 2\beta_R^2 + 16\tau^2 R c_{ab}^2. \tag{41}$$

*Consequently, $B_R < \tau^2 ab$ is sufficient for $\widehat{K}_R$ to have strictly smaller expected risk than $\widehat{K}_{\mathrm{full}}$.*

*Proof.* Optimality of the truncated SVD gives $\|Z - \widehat{K}_R\|_F^2 \le \|Z - K_R\|_F^2$. Set $e = \|\widehat{K}_R - K\|_F$ and $\beta = \beta_R$. Expanding both sides yields

$$e^2 \le \beta^2 + 2\tau\langle W, \widehat{K}_R - K_R\rangle. \tag{42}$$

The difference $\widehat{K}_R - K_R$ has rank at most $2R$. Spectral–nuclear norm duality and the triangle inequality imply

$$e^2 \le \beta^2 + 2\tau\sqrt{2R}\,\|W\|_{\mathrm{op}}(e + \beta). \tag{43}$$

Dividing by $e + \beta$ when it is positive gives $e \le \beta + \tau\sqrt{8R}\|W\|_{\mathrm{op}}$; the same bound holds when $e+\beta = 0$. Gaussian comparison and the Gaussian Poincaré inequality give $\mathbb{E}\|W\|_{\mathrm{op}} \le \sqrt{a}+\sqrt{b}$ and $\mathrm{Var}(\|W\|_{\mathrm{op}}) \le 1$, hence $\mathbb{E}\|W\|_{\mathrm{op}}^2 \le c_{ab}^2$ (Vershynin, 2018). Taking the $L^2$ norm of the preceding pointwise bound proves the first inequality in Eq. (41). The second follows from $(u+v)^2 \le 2u^2+2v^2$. Equation (38) gives the risk comparison. □

The bound accounts for estimating both factors: its variance term scales with $R(a + b)$, while $\beta_R^2$ measures the cost of imposing the partition's rank constraint. For an exact-rank matrix with $\mathrm{rank}(K) = R < \min(a, b)$, the small-noise limit gives the sharper expression

$$\mathbb{E}\|\widehat{K}_R - K\|_F^2 = \tau^2 R(a + b - R) + o(\tau^2), \qquad \tau \to 0. \tag{44}$$

To see this, the rank-$R$ projection is differentiable at $K$ because $s_R(K) > s_{R+1}(K) = 0$. Its derivative is the orthogonal projection onto the rank-$R$ tangent space, whose dimension is $R(a + b - R)$. Thus $(\widehat{K}_R - K)/\tau$ converges to the projection of $W$ onto that space. Optimality also gives $\|\widehat{K}_R - K\|_F \le 2\tau\|W\|_F$, so dominated convergence yields Eq. (44).

### C.3 Comparing Variable Partitions

Let $P$ and $Q$ denote two unfoldings of the same derivative tensor, with matrix dimensions $a_P \times b_P$ and $a_Q \times b_Q$. Both preserve the Frobenius metric and satisfy $a_P b_P = a_Q b_Q = qd$. Fix an admissible rank $R$ for both unfoldings and define

$$\beta_{P,R} = \inf_{\mathrm{rank}(M)\le R} \|\mathcal{U}_P(J) - M\|_F, \qquad \beta_{Q,R} = \inf_{\mathrm{rank}(M)\le R} \|\mathcal{U}_Q(J) - M\|_F. \tag{45}$$

Apply both unfoldings to the same noisy derivative observation. Let $\widehat{J}_P$ be obtained by truncating the $P$ unfolding to rank $R$ and reversing the reshaping.

**Corollary 8** (Benefit of partition alignment)**.** *If*

$$\beta_{P,R} + \tau\sqrt{8R}\, c_{a_P b_P} < \beta_{Q,R}, \tag{46}$$

*then $\widehat{J}_P$ has strictly smaller expected Frobenius risk than every estimator $\widehat{J}_Q$ satisfying $\mathrm{rank}(\mathcal{U}_Q(\widehat{J}_Q)) \le R$ almost surely.*

*Proof.* Every such $Q$ estimator satisfies $\|\widehat{J}_Q - J\|_F^2 \geq \beta_{Q,R}^2$ for each realization. Proposition 7 bounds the expected risk of $\widehat{J}_P$ by $(\beta_{P,R} + \tau\sqrt{8R}\, c_{a_P b_P})^2$. Condition (46) places this upper bound below the $Q$ approximation lower bound. □

At matched dimensions and rank, both matrix classes have rank-$R$ manifold dimension $R(a+b-R)$. The corollary therefore identifies an advantage due to variable alignment at equal representation dimension: a smaller spectral tail leaves more room for estimation error.

### C.4 From Operator Estimation to Local Prediction

Suppose $F$ is twice differentiable on a neighborhood of the closed ball of radius $t_0$ around $\mathbf{X}_0$, with $\|D^2F(X)[v,w]\|_2 \leq L\|v\|_2\|w\|_2$ there. For $h = tU$, where $0 < t \leq t_0$ and $U$ is uniform on the unit sphere in $\mathbb{R}^d$, Taylor expansion gives

$$F(\mathbf{X}_0 + h) = F(\mathbf{X}_0) + Jh + r(h), \qquad \|r(h)\|_2 \leq Lt^2/2. \tag{47}$$

Take $U$ independent of $W$, and use the known value $F(\mathbf{X}_0)$ in the predictor $\widehat{F}(\mathbf{X}_0 + h) = F(\mathbf{X}_0) + \widehat{J}h$. Since $\mathbb{E}[UU^\top] = I_d/d$, reshaping invariance gives

$$\mathbb{E}_{U,W}\|(\widehat{J} - J)h\|_2^2 = \frac{t^2}{d}\mathbb{E}_W\|\widehat{K} - K\|_F^2. \tag{48}$$

Let $\mathcal{R}_{\mathrm{str}}(t)$ and $\mathcal{R}_{\mathrm{full}}(t)$ be the expected squared prediction errors against $F(\mathbf{X}_0 + h)$ for the structured and unrestricted estimators. Equation (41) and the $L^2$ triangle inequality give

$$\mathcal{R}_{\mathrm{str}}(t) \leq \left(t\sqrt{B_R/d} + Lt^2/2\right)^2. \tag{49}$$

For the unrestricted estimator, the estimation noise is centered and independent of $U$, so its cross term with $r(tU)$ vanishes. Hence

$$\mathcal{R}_{\mathrm{full}}(t) = \tau^2 ab\, t^2/d + \mathbb{E}_U\|r(tU)\|_2^2. \tag{50}$$

It follows that

$$\sqrt{B_R} + \frac{Lt\sqrt{d}}{2} < \tau\sqrt{ab} \quad \Longrightarrow \quad \mathcal{R}_{\mathrm{str}}(t) < \mathcal{R}_{\mathrm{full}}(t). \tag{51}$$

This condition balances separation bias, factor-estimation error, and local nonlinearity against unrestricted estimation variance.

## D Experimental Protocols

### D.1 Synthetic MT Benchmark

The MT benchmark comprises independently generated three-dimensional resistivity models. Random volumetric fields are mapped to resistivity distributions, and numerical forward solutions provide the supervised responses. Figure 7 summarizes this construction; Table 3 specifies the spatial discretization, solver configuration, and data split. The network receives the log-resistivity field $\widetilde{\mathbf{X}} = \log_{10}\mathbf{X}$. The auxiliary density field is used during sample generation.

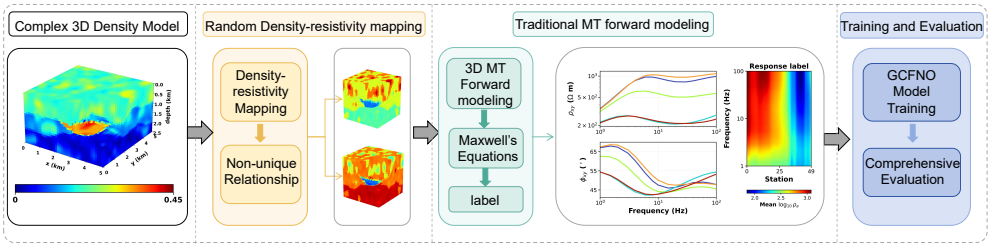


Figure 7: **Synthetic MT data generation.** Random volumetric fields are mapped to resistivity models and passed to the numerical solver to generate multi-frequency receiver responses.

Table 3: **MT benchmark and evaluation protocol.**

| Item | Setting |
|---|---|
| Input discretization | $50 \times 50 \times 50$ cells; $\Delta x = \Delta y = 100$ m and $\Delta z = 50$ m |
| Physical domain | $5 \times 5 \times 2.5$ km$^3$ |
| Dataset | 900 independent volumes |
| Split | 630 training, 135 validation, and 135 test volumes |
| Solver configuration | $100\ \Omega\,\mathrm{m}$ background; ten padding layers; padding expansion factor 1.4 |
| Receiver grid | $50 \times 50$ surface receivers |
| Frequencies | 15 predefined frequencies from 1 to 100 Hz |
| Response components | $\rho_{xy}$, $\rho_{yx}$, $\varphi_{xy}$, and $\varphi_{yx}$ |
| Stored output | $4 \times 15 \times 50 \times 50$, component-major ordering |
| Model selection | Lowest validation composite RMSE |
| Test evaluation | Held-out volumes evaluated after model selection |

The acquisition frequencies are

$$\mathcal{F} = \{100, 80, 60, 40, 30, 20, 15, 10, 8, 6, 4, 3, 2, 1.5, 1\}\ \mathrm{Hz}. \tag{52}$$

At each receiver–frequency pair, the transformed target contains the two off-diagonal apparent-resistivity components and their phases:

$$\widetilde{\mathbf{Y}} = [\log_{10}\rho_{xy}, \log_{10}\rho_{yx}, \varphi_{xy}/180, \varphi_{yx}/180]\,. \tag{53}$$

Phase is expressed in degrees before division by 180. Training uses this transformed representation directly.

### D.2 Head-EIT Benchmark

The head-EIT benchmark follows the synthetic construction released with BREIT (Abdelmoumene et al., 2026). It uses 193 healthy head anatomies from IXI to construct subject-specific finite-element models. Tissue admittivity spectra define the electrical-property input, and a finite-element solver generates complex electrode voltages. The benchmark therefore combines anatomical variability with numerically simulated EIT responses.

Each head model contains approximately 76,000 finite-element cells. The acquisition protocol comprises 17 frequencies from 5 to 2,000 Hz, 31 prescribed current-injection patterns, and 33 scalp electrodes. The output tensor has dimensions $17 \times 31 \times 33 \times 2$, corresponding to frequency, injection pattern, electrode, and real/imaginary voltage components. The subject-disjoint split contains 135 training subjects, 29 validation subjects, and 29 test subjects. All EIT models are trained independently on this benchmark using its own input and response representations.

### D.3 Losses and Evaluation Metrics

**MT training objective.** For a scalar prediction error $e = \widehat{y} - y$, the Huber penalty (Huber, 1964) is

$$\mathcal{H}_\delta(e) = \begin{cases} \frac{1}{2}e^2, & |e| \le \delta, \\ \delta\left(|e| - \frac{1}{2}\delta\right), & |e| > \delta, \end{cases} \qquad \delta = 1. \tag{54}$$

Let $\mathcal{L}_\rho$ and $\mathcal{L}_\varphi$ denote the meanHuber losses for the transformed apparent-resistivity and phase components.The MT objective is

$$\mathcal{L} = \lambda_\rho \mathcal{L}_\rho + \lambda_\varphi \mathcal{L}_\varphi, \qquad \lambda_\rho = \lambda_\varphi = 1. \tag{55}$$

Frequency-completion experiments evaluate this loss over the supervised frequency subset. Validation-based checkpoint selection uses the same supervised subset.

**MT evaluation.** We evaluate MT predictions using total RMSE, $\log_{10}\rho_a$ RMSE, and phase RMSE. Total RMSE is computed for each test volume over the four transformed response channels in Eq. (53) and averaged across test volumes. Table 1 reports $\log_{10}\rho_a$ RMSE in decades and phase RMSE in degrees.

**EIT evaluation.** We evaluate predicted electrode voltages against the numerical reference responses under the multi-frequency, multi-injection protocol described in Appendix D.2. Complex voltages are represented by their real and imaginary components. Voltage RMSE and relative $L_2$ error are computed separately for each held-out subject and then averaged over the 29 subjects with equal weight. Table 1 reports RMSE in mV and relative $L_2$ error as a percentage. Learned depth-aggregation profiles are provided in Appendix H.

# E Implementation and Training Details

## E.1 GCFNO Implementation

**MT implementation.** A shared three-dimensional encoder produces a latent volume, from which learned projections form surface-aligned features $\mathbf{S}$, global context $\mathbf{g}$, and depth-resolved features $\mathbf{D} = \{\mathbf{D}_d\}$. Receiver nodes are initialized using the surface-aligned representation and global context. Their interactions are conditioned on absolute coordinates, relative displacement, and pairwise distance. The condition pathway scores each depth slice using the numerical log-frequency embedding and encoded depth coordinate, then normalizes the scores over depth to aggregate $\mathbf{D}_d$.

For a given frequency, the aggregation weights are shared across samples and horizontal positions. Sample-specific information enters through the encoded depth features and the global context. The response readout combines the receiver feature, aggregated depth feature, global context, and frequency embedding. Frequency and global context also determine the FiLM scale and shift. The final prediction adds a learned condition residual to the base response. Zero initialization of the residual head's final weight matrix and bias gives an initial prediction equal to the base response.

Specifically, the MT readout fuses the representations defined in Section 3.3 and applies frequency-dependent modulation:

$$\begin{aligned} &\mathbf{u}_{i,f} = \psi_\theta\left([\mathbf{h}_i, \mathbf{d}_{i,f}, \mathbf{g}, E_f(f)]\right), \\ &[\boldsymbol{\gamma}_f, \boldsymbol{\beta}_f] = W_{\mathrm{film}}\left([E_f(f), \mathbf{g}]\right), \qquad \widetilde{\mathbf{u}}_{i,f} = \boldsymbol{\gamma}_f \odot \mathrm{LN}(\mathbf{u}_{i,f}) + \boldsymbol{\beta}_f, \end{aligned} \tag{56}$$

where $\psi_\theta$ is a nonlinear fusion network, $E_f(f)$ is the frequency embedding, $W_{\mathrm{film}}(\cdot)$ predicts the FiLM (Perez et al., 2018) scale and shift $(\boldsymbol{\gamma}_f, \boldsymbol{\beta}_f)$, and $\mathrm{LN}(\cdot)$ denotes layer normalization (Ba et al., 2016).

The final response combines a base prediction with a condition-dependent residual:

$$\widehat{\mathbf{y}}_i(f) = R_{\mathrm{base}}\left([\mathbf{h}_i, \mathbf{g}, E_f(f)]\right) + R_{\mathrm{cond}}\left(\widetilde{\mathbf{u}}_{i,f}\right), \tag{57}$$

where $R_{\mathrm{base}}$ maps the receiver feature, global context, and frequency embedding to the base response, while $R_{\mathrm{cond}}$ maps the fused geometry–condition representation to a condition-dependent residual.

**EIT implementation.** GCFNO operates directly on the unstructured finite-element representation. Each cell is encoded using its full admittivity spectrum. Geometry-dependent message passing uses source and target coordinates, relative displacement, distance, and shared-face geometry. The condition pathway groups cells into 16 bins according to normalized distance from the head boundary and combines the resulting representations with a frequency embedding. A zero-initialized residual supplies condition-dependent context to an electrode decoder conditioned on frequency and current-injection pattern. The MT implementation organizes the latent volume by vertical depth; the EIT implementation uses boundary-distance bins adapted to the head geometry.

## E.2 Dense and Coordinate-Query Readouts

Dense prediction uses the fixed $50 \times 50$ receiver lattice and arranges outputs as grid-indexed response maps. Coordinate-query prediction takes an explicit set $\mathcal{P}_q = \{\mathbf{p}_i\}_{i=1}^{N_q}$ and evaluates the response at those locations. The two interfaces use separately trained, mode-specific checkpoints. Geometry-generalization and query-cardinality results use coordinate-query models trained on the regular 625-receiver layout. Within these experiments, each model is evaluated at the requested test coordinates using its query interface.

# F Generalization Protocols and Full Results

## F.1 Held-Out Frequency Completion

Frequency completion withholds selected groups from the 15-frequency set in Eq. (52). Training and checkpoint selection use only supervised frequencies. The held-out frequencies are evaluated after training. Table 4 summarizes the four completion settings, which cover interior interpolation and directional extrapolation within the predefined acquisition set.

Table 4: **Held-out-frequency completion settings.** Each setting specifies the relation between supervised and excluded frequencies within the predefined MT frequency set.

| Setting | Completion task |
|---|---|
| Interpolation | Predict excluded interior frequencies between supervised frequencies. |
| High-to-low | Predict a lower-frequency group using supervision at higher frequencies. |
| Low-to-high | Predict a higher-frequency group using supervision at lower frequencies. |
| Mid-to-edges | Predict both frequency-range edges using supervision at central frequencies. |

The comparison uses a common GNO backbone with three treatments of frequency: frequency-specific output channels, an explicit numerical frequency input, and the structured GCFNO condition pathway. Their macro-averaged held-out-frequency RMSEs are 2.419, 1.264, and 0.546, respectively. GCFNO lowers the error in three of the four settings. For low-to-high completion, GCFNO obtains 0.563, compared with 0.478 for the numerical-frequency-input baseline. These settings evaluate completion over the specified discrete acquisition frequencies.

## F.2 Receiver-Geometry Generalization

All coordinate-query models are trained on one regular $25 \times 25$ receiver grid. Geometry transfer is evaluated on five coordinate-disjoint layouts containing 625 receivers each: a half-cell shift, affine warp, Poisson-disk sampling, clustered sampling, and anisotropic transects. The solver-field experiment uses 24 held-out volumes and selects receiver locations directly from the corresponding $50 \times 50$ numerical response fields. Every target response in this experiment is taken at a stored solver station. GINO (Li et al., 2023) is evaluated using the same training layout and test station selections.

Table 5: **Solver-field receiver-geometry generalization.** Total RMSE on 24 held-out MT volumes. Each layout contains 625 stations selected from the common $50 \times 50$ solver field. Train denotes the training receiver layout evaluated on held-out volumes.

| Model | Train | Shift | Warp | Poisson | Cluster | Aniso. |
|---|---|---|---|---|---|---|
| 3D U-Net (Çiçek et al., 2016) | 0.061137 | 0.128586 | 0.065994 | 0.113496 | 0.071209 | 0.072339 |
| FNO (Li et al., 2021) | 0.051752 | 0.053847 | 0.053761 | 0.053242 | 0.050358 | 0.059601 |
| DeepONet (Lu et al., 2021) | 0.076068 | 0.099546 | 0.095934 | 0.091744 | 0.096232 | 0.106153 |
| Perceiver IO (Jaegle et al., 2022) | 0.130902 | 0.130763 | 0.145028 | 0.129896 | 0.125329 | 0.158004 |
| GINO (Li et al., 2023) | 0.038008 | 0.043349 | 0.043452 | 0.041410 | 0.041148 | 0.048212 |
| GNO (Li et al., 2020) | 0.023803 | **0.035930** | 0.033192 | 0.032597 | 0.033301 | 0.036705 |
| GCFNO | **0.022797** | 0.036063 | **0.032778** | **0.032205** | **0.031870** | **0.035478** |

GCFNO obtains an unseen-layout macro-average RMSE of 0.033679, compared with 0.034345 for GNO and 0.043514 for GINO. Relative RMSE reductions against GNO are $-0.37\%$, $1.25\%$, $1.20\%$, $4.30\%$, and $3.34\%$ for shift, warp, Poisson, cluster, and anisotropic layouts, respectively. Positive values indicate improvement. The corresponding reductions against GINO are $16.8\%$, $24.6\%$, $22.2\%$, $22.6\%$, and $26.4\%$.

The reported paired bootstrap over the 24 volumes gives a clustered-layout gain of 0.001279 with a 95% interval of $[0.000168, 0.002453]$. The intervals for the other layouts include zero. The evidence therefore comprises modest mean improvements on most layouts and a positive paired interval for the clustered configuration. The evaluation domain is the common discrete solver station field.

A separate sampled-field experiment evaluates the same classes of receiver layouts on all 135 test volumes, with labels obtained by bilinear sampling from the stored $50 \times 50$ response field. Its summary is reported alongside query-cardinality transfer in Table 6. The two protocols retain their own test populations and label-construction procedures.

### F.3 Query-Cardinality Transfer

We evaluate receiver-query generalization on 135 held-out MT volumes. All models are trained on a regular $25 \times 25$ grid containing 625 receivers. Test labels are obtained by bilinear sampling from the stored $50 \times 50$ numerical response fields, allowing receiver locations and query-set sizes to vary within a common response representation.

Table 6 summarizes four evaluation settings. *Seen receiver layout* reports test error at the training receiver coordinates. *Unseen receiver layouts* summarizes five new layouts, each containing 625 receivers. *Unseen receiver counts* evaluates new coordinate sets containing 144, 2,500, 5,625, or 10,000 receivers. *All unseen settings* combines the five layout changes and four coordinate-and-cardinality changes. The final two columns provide high-gradient RMSE and 95th-percentile pointwise error diagnostics.

GCFNO achieves the lowest total RMSE in all four summary columns. Across unseen receiver counts, it reduces mean RMSE from $0.027517$ for GNO to $0.024992$, a $9.18\%$ reduction. Figure 8 shows that its RMSE remains between $0.0243$ and $0.0253$ across the tested query-set sizes. Over all nine unseen configurations, GCFNO reduces the combined error from $0.027323$ to $0.025916$, corresponding to a $5.15\%$ improvement.

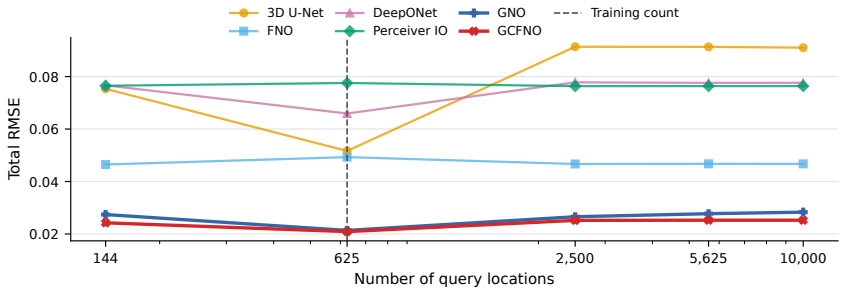


Figure 8: **Coordinate-and-cardinality transfer.** Models are trained on the regular 625-receiver grid and evaluated on unseen query sets containing 144, 2,500, 5,625, or 10,000 receivers. The 625-receiver point uses the training layout.

Table 6: **Receiver-query generalization on sampled MT response fields.** Results are evaluated on 135 held-out volumes using bilinearly sampled response labels. Seen layout uses the training receiver grid. Unseen layouts summarizes five new layouts with 625 receivers each; unseen counts summarizes four settings with new coordinates and different receiver counts. All unseen settings combines these nine configurations.

| | Total RMSE ↓ | | | | Error diagnostics ↓ | |
|---|---|---|---|---|---|---|
| Model | Seen receiver layout | Unseen receiver layouts | Unseen receiver counts | All unseen settings | High-gradient RMSE | 95th-percentile pointwise error |
| 3D U-Net (Çiçek et al., 2016) | 0.051649 | 0.087682 | 0.087258 | 0.087494 | 0.067276 | 0.132433 |
| FNO (Li et al., 2021) | 0.049300 | 0.047290 | 0.046661 | 0.047010 | 0.065090 | 0.087879 |
| DeepONet (Lu et al., 2021) | 0.065916 | 0.080673 | 0.077424 | 0.079229 | 0.106491 | 0.150943 |
| Perceiver IO (Jaegle et al., 2022) | 0.077552 | 0.076114 | 0.076400 | 0.076241 | 0.094298 | 0.134332 |
| GINO (Li et al., 2023) | 0.034985 | 0.035425 | 0.034930 | 0.035205 | 0.055085 | 0.066331 |
| GNO (Li et al., 2020) | 0.021367 | 0.027168 | 0.027517 | 0.027323 | 0.044587 | 0.052462 |
| GCFNO | **0.020914** | **0.026656** | **0.024992** | **0.025916** | **0.042500** | **0.050810** |

## G Data Efficiency and Robustness

### G.1 Data Efficiency

The MT training set is subsampled at fractions $p \in \{0.05, 0.10, 0.20, 0.50, 1.00\}$, with validation and test sets fixed. Each fraction uses three subset seeds. Figure 9 reports the resulting means and one-standard-deviation error bars for Total RMSE and the two response components.

For a test-error curve $E(p)$, the normalized area on the logarithmic training-data axis is

$$\mathrm{AUC}_{\log}(E) = \frac{1}{\log p_{\max} - \log p_{\min}} \int_{\log p_{\min}}^{\log p_{\max}} E(e^u)\, \mathrm{d}u, \qquad p_{\min} = 0.05, \quad p_{\max} = 1. \tag{58}$$

Here $u = \log p$. Lower values indicate lower error averaged over the logarithmic data-fraction range. The same summary is applied to Total RMSE, apparent-resistivity RMSE, and phase RMSE.

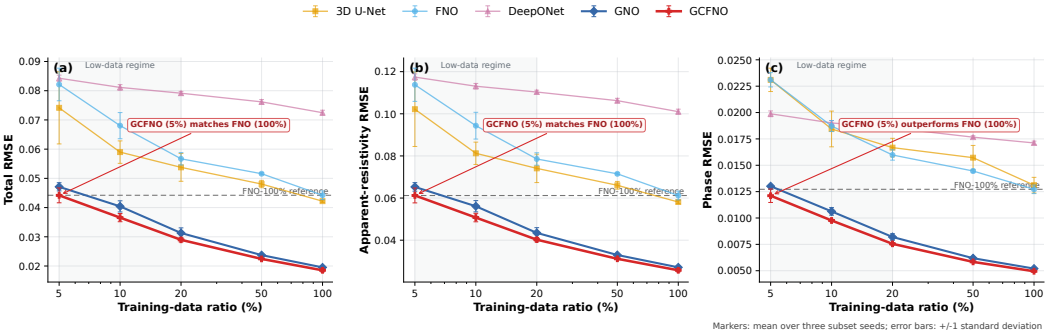


Figure 9: **Learning curves under reduced MT training data.** Markers show means over three subset seeds; error bars show one standard deviation. Dashed lines indicate the full-data FNO reference for the corresponding response metric.

Table 7: **Full data-efficiency summaries.** The first three metric columns report Total RMSE at selected training fractions. Mean summarizes Total RMSE across the training fractions; the final three columns report log-ratio AUC for each response metric.

| Model | 5% | 20% | 100% | Mean | Total AUC | $\rho_a$ AUC | Phase AUC |
|---|---|---|---|---|---|---|---|
| 3D U-Net (Çiçek et al., 2016) | 0.074150 | 0.053764 | 0.042189 | 0.055427 | 0.054456 | 0.074993 | 0.017159 |
| FNO (Li et al., 2021) | 0.082136 | 0.056693 | 0.044212 | 0.060535 | 0.059448 | 0.082385 | 0.016652 |
| DeepONet (Lu et al., 2021) | 0.084270 | 0.079145 | 0.072473 | 0.078640 | 0.078629 | 0.109615 | 0.018407 |
| GNO (Li et al., 2020) | 0.047109 | 0.031296 | 0.019526 | 0.032414 | 0.031840 | 0.044207 | 0.008431 |
| GCFNO | **0.044185** | **0.028965** | **0.018528** | **0.030135** | **0.029522** | **0.040982** | **0.007829** |

### G.2 Robustness to Input Degradation

The target response is held fixed while the input electrical-property volume is degraded. Let $\mathbf{X}$ be the clean input, $\widetilde{\mathbf{X}}_\eta = \mathcal{T}_\eta(\mathbf{X})$ its version at degradation level $\eta$, and $\mathbf{Y} = \mathcal{F}_{\mathrm{solver}}(\mathbf{X})$ the clean solver response. The prediction is $\widehat{\mathbf{Y}}_\eta = \mathcal{G}_\theta(\widetilde{\mathbf{X}}_\eta)$. We report

$$E_\eta = \mathrm{RMSE}(\widehat{\mathbf{Y}}_\eta, \mathbf{Y}), \qquad D_\eta = \frac{E_\eta - E_0}{E_0} \times 100\%, \tag{59}$$

where $E_0$ is the clean-input error for the same model. The experiments cover log-resistivity noise, boundary smoothing, and coarse-grid reconstruction.

GCFNO retains lower absolute error at all reported noise levels and coarse-grid reconstruction factors. Under boundary smoothing, its error is lower at level 0.25, while GNO has lower error at levels 0.50 and 0.75. The relative-degradation metric additionally shows how each model's error changes from its own clean-input reference.

## H Additional Head-EIT Analyses

### H.1 Learned Frequency–Depth Conditioning in Head EIT

Figure 10 shows learned aggregation weights over normalized boundary distance, from the head surface toward the interior. Lower frequencies emphasize interior features, while higher frequencies favor boundary-adjacent features, as summarized by the weighted-depth and peak-depth profiles.

Table 8: **Robustness to input degradation.** Entries report Total RMSE. Parentheses give the percentage increase relative to each model's clean-input error.

| Input | Level | GNO (Li et al., 2020) | GCFNO |
|---|---|---|---|
| Clean | – | 0.019072 (0.00%) | **0.017285** (0.00%) |
| Log-resistivity noise | 0.01 | 0.019078 (+0.03%) | **0.017294** (+0.05%) |
| | 0.03 | 0.019104 (+0.17%) | **0.017328** (+0.24%) |
| | 0.05 | 0.019171 (+0.52%) | **0.017425** (+0.81%) |
| | 0.10 | 0.019430 (+1.87%) | **0.017775** (+2.83%) |
| Boundary smoothing | 0.25 | 0.034397 (+80.35%) | **0.033771** (+95.37%) |
| | 0.50 | **0.057436** (+201.15%) | 0.057723 (+233.94%) |
| | 0.75 | **0.079959** (+319.24%) | 0.082582 (+377.75%) |
| Coarse-grid reconstruction | Factor 2 | 0.047613 (+149.64%) | **0.046915** (+171.41%) |
| | Factor 4 | 0.069229 (+262.98%) | **0.069078** (+299.63%) |
| | Factor 5 | 0.073718 (+286.52%) | **0.073615** (+325.88%) |

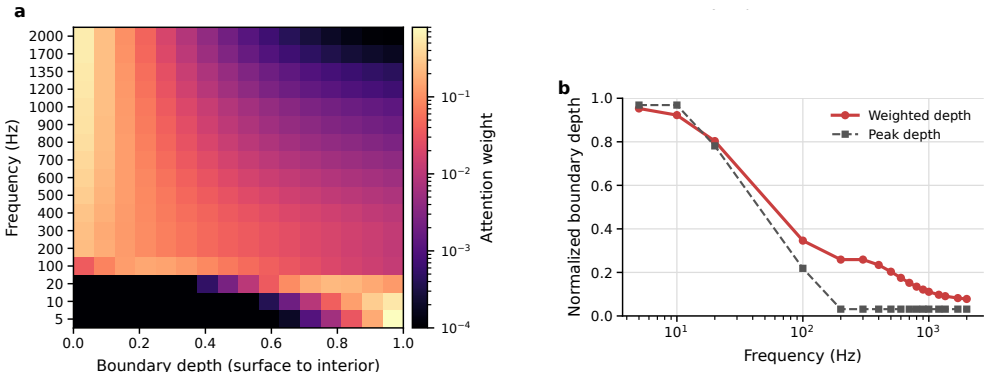


Figure 10: **Learned depth aggregation on head EIT.** (a) Aggregation weights over normalized boundary depth and frequency. (b) Weighted-depth and peak-depth profiles across frequency.

## I GCFNO-Driven Inverse Reconstruction

**MT inversion.** As an additional downstream evaluation, we use the learned GCFNO forward surrogate to drive a 3-D MT inversion. In this setting, GCFNO serves as the forward mapping between a candidate volumetric resistivity model and its multi-frequency surface responses, allowing the inverse procedure to search for a subsurface model that is consistent with the observed MT measurements. Figure 11 shows a representative reconstruction together with the corresponding ground truth.

The recovered model reproduces the dominant subsurface structures and their spatial locations in both the 3-D representation and the representative horizontal and vertical cross-sections. In particular, the major anomalous regions appearing in the reference model remain identifiable in the inversion, and their approximate lateral extent and depth distribution are preserved. The reconstruction is nevertheless smoother than the ground truth, especially around sharp structural boundaries and smaller-scale heterogeneities, which is consistent with the ill-posed nature of the inverse problem. Despite these local discrepancies, the principal large-scale resistivity contrasts are retained across the displayed sections. This auxiliary experiment therefore indicates that the learned GCFNO forward operator can remain useful when embedded in a downstream inversion workflow

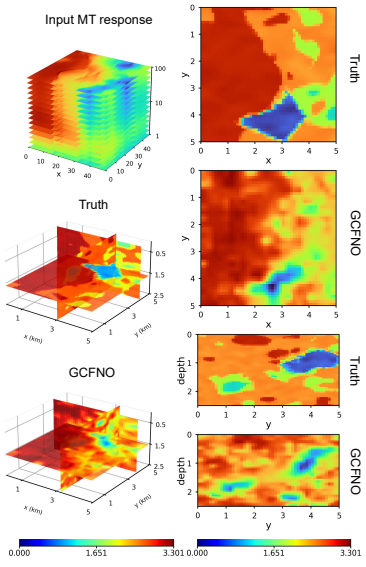


Figure 11: **GCFNO-driven 3-D MT inversion.** Representative ground-truth and reconstructed structures.

**Head-EIT inversion.** We further evaluate downstream inverse use in head EIT using the learned GCFNO forward surrogate to drive reconstruction from electrode-voltage measurements.

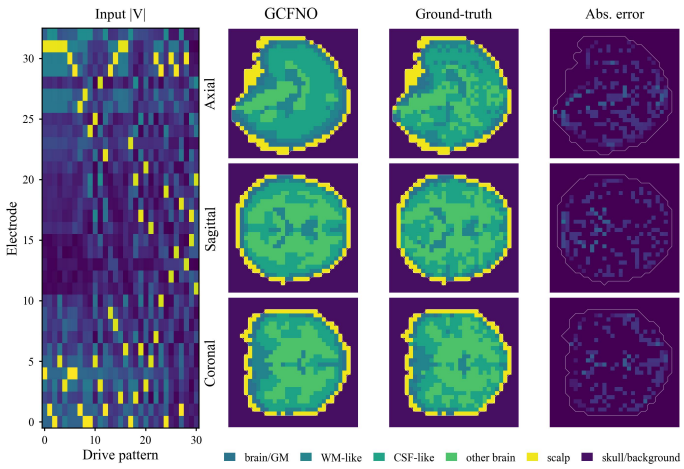


Figure 12: **GCFNO-driven head-EIT inverse reconstruction.** Ground truth, reconstruction, and absolute error in three orthogonal views.

Figure 12 shows a representative example. The reconstructed volume preserves the overall head geometry and the major internal tissue organization across axial, sagittal, and coronal views, with the remaining errors mainly localized within the interior tissue regions.

For the final reconstruction, the Dice scores for brain/GM, WM-like, CSF-like, other brain, scalp, and skull/background are $60.74\%$, $50.69\%$, $69.52\%$, $76.06\%$, $99.92\%$, and $100.00\%$, respectively. The corresponding IoU scores are $43.62\%$, $33.95\%$, $53.28\%$, $61.37\%$, $99.84\%$, and $99.99\%$. The higher agreement for exterior structures than for the internal tissue classes reflects the greater difficulty of resolving fine interior boundaries from boundary-voltage measurements. Together with the MT example, this result provides an auxiliary indication that GCFNO can support downstream inverse use.