RESCO: RESIDUAL-CONDITIONED SUBSPACE CORRECTION FOR HELMHOLTZ NEURAL OPERATORS

Anonymous authors
Paper under double-blind review

ABSTRACT

Helmholtz discretizations can share Poisson's sparsity pattern while exhibiting indefiniteness and near-resonant sensitivity, motivating neural-operator designs informed by matrix properties. We propose **residual-conditioned subspace correction** (RESCO), which combines a pretrained multigrid neural operator with a fixed POD error basis and a learned coefficient-recovery module. The module uses physical residuals and projected features to predict global wavefield corrections. To improve generalization, we train it on samples excluded from backbone training. On the Caltech Helmholtz validation benchmark across three seeds, the method reduces mean relative L^2 error from 0.00998 to 0.00619, achieving a mean per-seed improvement of 37.27%, compared with 6.2% when correction training uses backbone-seen samples. These results demonstrate the value of combining matrix-informed correction with training on generalization errors.

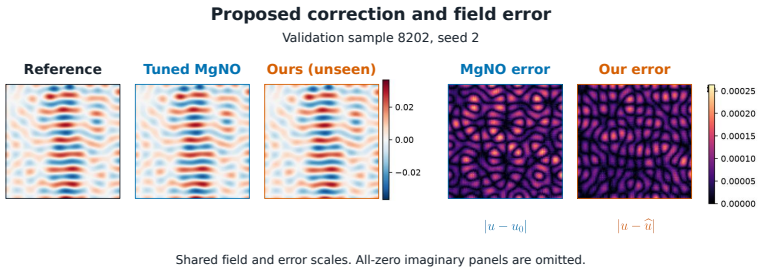


Figure 1: **Residual-conditioned subspace correction (RESCO) for Helmholtz wavefields.** From left to right: reference solution, tuned MgNO prediction, our corrected prediction, and their pointwise absolute errors for validation sample 8202 (seed 2). Our method refines the frozen backbone prediction using physical residuals and a reusable error basis. The correction head is trained on samples excluded from backbone training, denoted by "unseen." The error maps reveal structured discrepancies that are difficult to distinguish in the wavefield plots. Field panels share one color scale, and error panels share another; identically zero imaginary parts are omitted.

1 INTRODUCTION

Repeated wavefield prediction is a central computational task in seismic imaging, inverse scattering and acoustic modeling. Neural operators amortize this cost by learning maps from PDE inputs to solution fields (Kovachki et al., 2023; Li et al., 2021). Helmholtz prediction requires separating field accuracy, physical residual and iterative-solver cost.

The discrete inverse explains this distinction. For constant wavenumber and homogeneous Dirichlet conditions, the illustrative relation $A_{\mathrm{H}} = A_{\mathrm{P}} - k^2 I$ preserves the Poisson sparsity pattern while shifting its eigenvalues. Excited modes near the shifted spectrum's zero can be strongly amplified, making a small residual compatible with substantial solution error. Such spectral difficulties also motivate specialized Helmholtz preconditioners (Erlangga et al., 2006). Our experiments use a different Neumann Q1 finite-element operator, $A_h = K_h - \omega^2 M_h(c^{-2})$; the constant spectral shift is motivational, not an identity for this variable-velocity benchmark.

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.

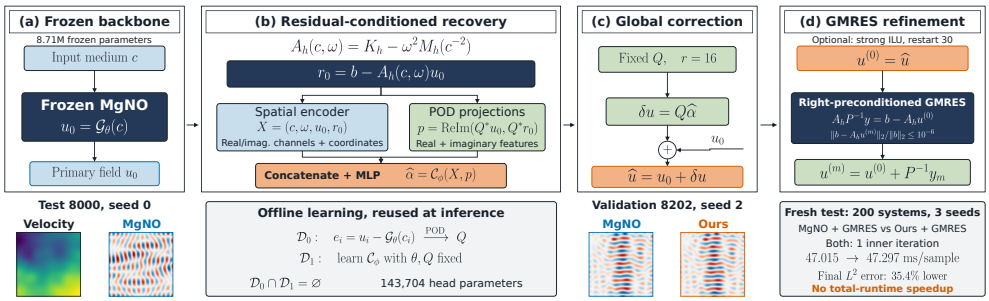


Figure 2: **Residual-conditioned subspace correction (RESCO) and optional solver refinement.** (a) Frozen MgNO predicts u_0. (b) Spatial features and POD projections of u_0 and $r_0 = b - A_h u_0$ condition the coefficient head, trained offline on backbone-unseen data. (c) The fixed rank-16 basis lifts $\widehat{\alpha}$ to $\widehat{u} = u_0 + Q\widehat{\alpha}$. (d) Right-preconditioned GMRES refines this warm start with shared ILU. Both neural warm starts use one inner iteration in the recorded fresh test, with no total-runtime speedup from the learned head. Illustrations are distinct cases: test 8000/seed 0 in (a), validation 8202/seed 2 in (c).

We study correction of a pretrained predictor within a reusable low-dimensional error subspace. A basis may represent an error accurately while its coefficients are recovered poorly: residual minimization need not minimize field error. Errors on backbone-training inputs can also differ from errors on new media. We therefore distinguish representation, coefficient recovery and correction-training provenance.

We propose ***Re**sidual-Conditioned **S**ubspace **Co**rrection* (RESCO), with RESCO-SEEN as the seen-data ablation and RESCO-OP as the operator-feature extension. In RESCO, we utilize MgNO (He et al., 2024) as the backbone to predict the initial field, with a fixed rank POD error basis for global correction directions:

$$u_0 = \mathcal{G}_\theta(c), \qquad r_0 = b - A_h u_0, \qquad \widehat{u} = u_0 + Q\mathcal{C}_\phi(X, p). \tag{1}$$

The coefficient head combines spatial features of the medium, frequency, prediction and residual with projections $Q^* u_0$ and $Q^* r_0$. It is supervised on an additional labeled set excluded from backbone training, with the backbone and basis fixed and validation kept separate. We compare this protocol with backbone-seen correction training at matched head-training counts and budgets. Basis–coefficient prediction is an established model-reduction construction (Bhattacharya et al., 2021); our contribution is its residual-conditioned use for backbone errors and the controlled assessment of correction-training provenance. Additional labels remain a confound in comparisons with the uncorrected backbone.

Our contributions are threefold:

1. **A correction architecture with an explicit error decomposition.** We separate representation error outside the fixed POD subspace from coefficient error within it. The reduced-solve identities explain how unrepresented error can contaminate residual-based coefficient recovery, motivating supervision of the field-optimal coefficients.
2. **Controlled accuracy and online-cost evidence.** In the reported original-test comparison, the unseen-trained correction achieves a mean per-seed relative L^2 reduction of 31.20%, versus 6.84% for the seen-data ablation. Separately, validation study reduces mean relative L^2 from 0.009979 to 0.006190, a 37.27% mean per-seed gain, with 143,704 trainable head parameters and 0.388 ms/sample additional batch-amortized online cost. Matched ablations support contributions from the POD basis, residual inputs and unseen-error training, without identifying a unique dominant component.
3. **Spectral and converged-solver characterization.** On eight validation media across three seeds, the correction reduces geometric-Q1 field error by 38.8% and error in the 24 Helmholtz modes nearest zero by 45.8%, while agreement remains weak in the highest spatial bands. On 200 shared fresh-test systems across three seeds, all iterative arms meet the prescribed true-residual tolerance. Relative to MgNO-initialized GMRES, the corrected

initialization improves endpoint field accuracy but saves neither an inner iteration nor total online time.

2 Related Work

Structured neural operators and pretrained PDE models. Neural operators learn maps between function spaces, with architectures that encode assumptions about the associated solution operators (Kovachki et al., 2023). Boullé & Townsend (2024) connect operator learning to low-rank, convolutional, local, and hierarchical matrix approximation, providing a numerical-linear-algebra perspective on architectural design. Representative constructions include Fourier neural operators (Li et al., 2021), U-shaped neural operators (Rahman et al., 2022), multipole graph neural operators (Li et al., 2020), and multigrid neural operators (He et al., 2024). Poseidon extends this landscape through large-scale pretraining and transfer across PDE families, including a downstream Helmholtz task (Herde et al., 2024). Our work applies this structural perspective to the errors of an existing predictor: a frozen MgNO produces the primary field, and the discrete Helmholtz operator supplies residual features for a separate correction module. The relevant structure concerns the inverse response and the distribution of prediction errors; identical sparsity of Poisson and Helmholtz matrices does not establish equivalent approximation behavior.

Reduced representations and correction of pretrained operators. Model reduction with neural coefficient maps (Bhattacharya et al., 2021), DeepONet (Lu et al., 2021), and POD-DeepONet (Lu et al., 2022) establish basis–coefficient representations for operator learning. Residual-based correctors instead improve surrogate predictions by solving a linear variational correction problem (Cao et al., 2023; Jha, 2024). More recently, LatentDDM learns a lightweight composition module around frozen local neural operators (Chen et al., 2026). These approaches motivate separating a reusable predictor from an adaptation or correction mechanism. Our construction uses a fixed POD basis of *backbone prediction errors*, with coefficients inferred from the medium, initial prediction, physical residual, and projected residual. We further train the coefficient module on a disjoint labeled set excluded from backbone training, while keeping validation separate, to investigate the discrepancy between backbone-seen and backbone-unseen errors. Thus, our contribution lies in the combination of error-subspace representation, residual-conditioned coefficient recovery, and this training protocol; neither POD coefficient prediction nor correction of a frozen predictor is individually new.

Classical and learned Helmholtz solvers. Complex-shifted multigrid preconditioning addresses the spectral difficulties of heterogeneous Helmholtz systems (Erlangga et al., 2006). HINTS and DeepONet-based preconditioners combine learned operators with classical iteration or learned-basis coarse correction (Zhang et al., 2024; Kopaničáková & Karniadakis, 2025). Helmholtz-specific designs include Neumann-series neural operators (Chen et al., 2024), multigrid-augmented preconditioners with implicit coarse layers (Lerer et al., 2024), and McMg, whose medium-conditioned multichannel V-cycle preserves residual linearity and represents unresolved wave information during coarsening (Jia et al., 2026). Complementary analysis identifies input projection and source linearity as important architectural choices for learned coarse solves (Melchers et al., 2026). Our present objective is a single correction that improves the field accuracy of a frozen surrogate. We learn coefficients targeting its solution error within a reusable global subspace, without asserting that the resulting module is a convergent preconditioner. Our validation results support improved prediction accuracy; reduction of the physical residual and acceleration to a prescribed solver tolerance require separate evidence.

3 Theoretical Framework

We use the Neumann Q1 operator $A = A_h(c, \omega) = K_h - \omega^2 M_h(c^{-2})$, where velocity enters the reaction mass. For invertible A, reference u, prediction $\widehat{u}$, and defect $\delta = b - Au$,

$$\|u - \widehat{u}\|_2 \le \sigma_{\min}(A)^{-1}(\|b - A\widehat{u}\|_2 + \|\delta\|_2). \tag{2}$$

For the illustrative constant-k Dirichlet matrix $A_H = A_P - k^2 I$, $\sigma_{\min}(A_H) = \min_j |\lambda_j(A_P) - k^2|$. This eigenvalue shift does not hold for the general Q1 matrix. Residual reduction alone need not improve field accuracy, especially when the inverse amplifies weak residual directions.

Representation and recovery. Let u_0 be a frozen prediction and $Q \in \mathbb{C}^{N\times r}$ satisfy $Q^*Q = I$. Writing $e = u - u_0 = Q\alpha_* + e_\perp$, with $\alpha_* = Q^*e$ and $Q^*e_\perp = 0$, gives exactly

$$\|u - (u_0 + Q\alpha)\|_2^2 = \underbrace{\|e_\perp\|_2^2}_{\text{representation error}} + \underbrace{\|\alpha - \alpha_*\|_2^2}_{\text{coefficient error}}. \tag{3}$$

The oracle $u_0 + Q\alpha_*$ is a label-dependent diagnostic. The leading left singular vectors of normalized training errors minimize their rank-r projection error; this POD optimality does not ensure coefficient recovery or generalization.

Why learn coefficients? Set $Z = AQ$, $H = Q^*AQ$ and $r_0 = b - Au_0$. Ridge minimizes $\|Z\alpha - r_0\|_2^2 + \lambda\|\alpha\|_2^2$, while shifted Galerkin solves $(H - i\beta I)\alpha_G = Q^*r_0$. For $\lambda > 0$ and Hermitian A with $\beta > 0$, their coefficient errors satisfy

$$\alpha_R - \alpha_* = (Z^*Z + \lambda I)^{-1}[Z^*(Ae_\perp + \delta) - \lambda\alpha_*], \tag{4}$$

$$\alpha_G - \alpha_* = (H - i\beta I)^{-1}[Q^*(Ae_\perp + \delta) + i\beta\alpha_*]. \tag{5}$$

Thus unrepresented error couples into physical coefficient recovery even when $Q^*e_\perp = 0$. Reduced conditioning, alignment and regularization jointly matter; proximity to resonance alone does not determine accuracy. We instead supervise the field-optimal coefficients α_*.

4 RESCO: RESIDUAL-CONDITIONED SUBSPACE CORRECTION

Figure 2 connects discrete residual information, a reusable error basis and learned coefficients. Panels (a)–(c) define one-shot RESCO; panel (d) optionally refines its prediction.

(a) Frozen backbone and discrete physics. For velocity c and frequency ω, the Neumann Q1 system is $A_h(c,\omega)u = b$, with $A_h = K_h - \omega^2 M_h(c^{-2})$. Here $M_h(c^{-2})$ is the weighted mass matrix and b the boundary load. Frozen MgNO gives $u_0 = \mathcal{G}_\theta(c)$ with 8.71M fixed parameters. We compute $r_0 = b - A_h u_0$ without reference labels, incorporating the discrete reaction term through residual evaluation.

(b) Residual-conditioned coefficient recovery. The spatial branch encodes $X = (c, \omega, u_0, r_0)$ as six real channels: $c, \omega, \mathrm{Re}\, u_0, \mathrm{Im}\, u_0, \mathrm{Re}\, r_0, \mathrm{Im}\, r_0$. Training-only normalization and two coordinate channels precede a three-layer convolutional encoder and adaptive pooling. The POD branch forms

$$p = [\mathrm{Re}(Q^*u_0), \mathrm{Im}(Q^*u_0), \mathrm{Re}(Q^*r_0), \mathrm{Im}(Q^*r_0)] \in \mathbb{R}^{4r}. \tag{6}$$

An MLP concatenates encoded spatial and normalized projection features. Its $2r$ real outputs, decoded with saved coefficient scales, define $\widehat{\alpha} = \mathcal{C}_\phi(X, p) \in \mathbb{C}^r$. The complete head has 143,704 trainable parameters.

(c) Global subspace correction. A fixed rank-16 basis $Q \in \mathbb{C}^{N_h\times r}$, $Q^*Q = I_r$, lifts the recovered coefficients and adds them to the backbone field:

$$\delta u = Q\widehat{\alpha}, \qquad \widehat{u} = u_0 + Q\mathcal{C}_\phi(X, p). \tag{7}$$

No reference solution is needed at inference.

Offline learning and reuse. Let $\mathcal{D}_0$ be backbone-training data and $\mathcal{D}_1$ disjoint labeled correction-training data. Construct $e_i = u_i - \mathcal{G}_\theta(c_i)$ on $\mathcal{D}_0$ and take the leading left singular vectors of $E = [e_i/\|e_i\|_2]_{e_i\neq 0}$ as Q. Keep θ, Q fixed and train ϕ on $\mathcal{D}_1$:

$$\mathcal{L}(\phi) = \frac{1}{|\mathcal{D}_1|s_0}\sum_{i\in\mathcal{D}_1}\frac{\|\mathcal{C}_\phi(X_i, p_i) - Q^*(u_i - u_{0,i})\|_2^2}{\|u_i\|_2^2}. \tag{8}$$

Here $s_0 > 0$ is the mean squared relative backbone error on $\mathcal{D}_1$. By Eq. 3, this objective equals squared relative field loss up to scale and a fixed additive term. Residuals condition the head; the primary objective has no residual penalty. Validation and test are excluded from both training sets. Validation selects the checkpoint. RESCO-SEEN instead trains the head on backbone-seen errors.

Algorithm 1 Offline learning of reusable correction components

1 Freeze $\mathcal{G}_\theta$; construct rank-16 normalized error POD Q on $\mathcal{D}_0$.
2 On disjoint $\mathcal{D}_1$, compute u_0, r_0, X, p and targets $Q^*(u - u_0)$.
3 Fit feature/coefficient scales on $\mathcal{D}_1$; zero-initialize the head output layer.
4 Optimize Eq. 8 with Adam/cosine decay; keep θ, Q fixed.
5 Select by separate validation relative L^2, including epoch zero; save Q, head and scales.

(d) Optional right-preconditioned GMRES. Use $u^{(0)} = \widehat{u}$ as the initial guess and form the same per-system ILU preconditioner P for every solver arm:

$$A_h P^{-1} y = b - A_h u^{(0)}, \qquad u^{(m)} = u^{(0)} + P^{-1} y_m. \tag{9}$$

The illustrated protocol uses strong ILU, restart 30, and the true-residual rule $\|b - A_h u^{(m)}\|_2 / \|b\|_2 \leq 10^{-6}$. RESCO supplies the initial guess once; ILU supplies the preconditioner. Their costs and convergence are evaluated separately.

Algorithm 2 Online correction and optional GMRES refinement

1 Predict $u_0 = \mathcal{G}_\theta(c)$; compute $r_0 = b - A_h u_0$ and normalized X, p.
2 Recover $\widehat{\alpha} = \mathcal{C}_\phi(X, p)$; form $\delta u = Q\widehat{\alpha}$, $\widehat{u} = u_0 + \delta u$.
3 Return $\widehat{u}$ for one-shot prediction; otherwise set $u^{(0)} = \widehat{u}$ and form ILU P.
4 Run GMRES(30) on $A_h P^{-1} y = b - A_h u^{(0)}$; set $u^{(m)} = u^{(0)} + P^{-1} y_m$.
5 Stop at $\|b - A_h u^{(m)}\|_2 / \|b\|_2 \leq 10^{-6}$ or the iteration budget; report convergence.

Cost and extension. Beyond MgNO, prediction adds operator assembly/evaluation, one head, and $O(N_h r)$ projections, lifting and basis storage. The separately evaluated RESCO-OP adds $H = Q^* A_h Q$, $g = (A_h Q)^* r_0$ and $S = (A_h Q)^* (A_h Q)$; it is not the primary head in panel (b). Field gains do not guarantee residual reduction, acceleration or transfer.

5 EXPERIMENTS

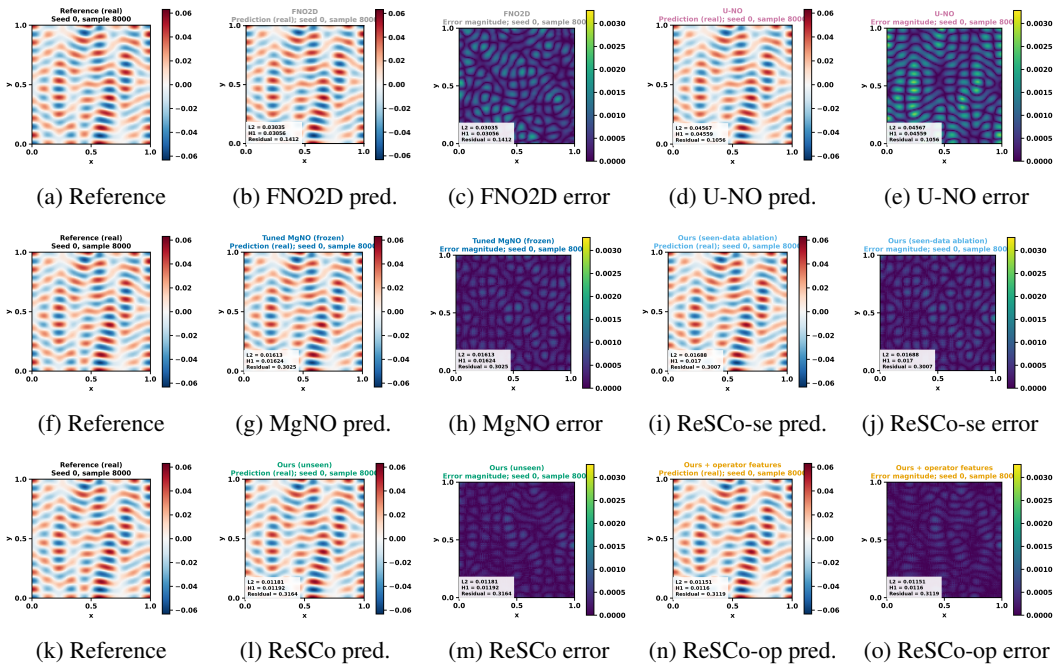
In this section, we present experimental results for our proposed methods and design experiments to demonstrate the significance of the improvements.

To guide our experiments, we address the following four research questions:

- **R1. Accuracy & Efficiency Gains** How much do the proposed methods increase accuracy and computational efficiency when solving the Helmholtz equation?
- **R2. Ablation & Component Analysis** How much do the POD basis, residual inputs, learned coefficients, backbone-unseen training, and operator features each contribute to accuracy and online cost?
- **R3. Spectral Comparison** How closely does the learned spectral behavior of the neural operator match that of traditional discrete spectral analysis?
- **R4. Hybrid Solver Efficiency** Does integrating the proposed methods with GMRES iterative solvers yield significant reductions in runtime or computational cost compared to baseline approaches?

5.1 EXPERIMENTAL SETTING

Data. We use the Caltech paired real-velocity/solution release ($101 \times 101 \times 40000$, float64). We keep the earlier benchmark split, using zero-based indices: 4,000 training samples (0–3999), 200 validation samples (8200–8399), and 200 test samples (8000–8199). A physical adapter reconstructs the original Q1 finite-element discretization on 100×100 cells with 3×3 Gauss quadrature, $\omega = 1000$, and Neumann boundary forcing. The maximum reference residual across all 4,200 training and validation samples is reduced to 4.47×10^{-6}, well below the unchanged 10^{-3} tolerance. The first four float64 checks give residuals of approximately 3.48×10^{-14}–4.43×10^{-14}. Test references were excluded from operator calibration. The fresh test set consists of 200 held-out systems drawn at



(a) Reference (b) FNO2D pred. (c) FNO2D error (d) U-NO pred. (e) U-NO error

(f) Reference (g) MgNO pred. (h) MgNO error (i) ReSCo-se pred. (j) ReSCo-se error

(k) Reference (l) ReSCo pred. (m) ReSCo error (n) ReSCo-op pred. (o) ReSCo-op error

Figure 3: **Qualitative comparison of Helmholtz wavefield predictions.** Reference fields, predictions, and pointwise absolute errors are shown for FNO2D and U-NO (top row), tuned MgNO and our seen-data ablation (middle row), and our unseen-trained correction and operator-feature extension (bottom row). The proposed corrections reduce the structured wavefield errors remaining in the MgNO prediction. Reference fields are repeated for visual comparison; color bars indicate the scale of each panel.

random from the unused portion of the Caltech release (IDs 4123–39962, excluding the backbone-training 0–3999, test 8000–8199, and validation 8200–8399 indices). We provide the full list in the code supplementary.

Compared models. We benchmark FNO2D, U-NO, and a frozen tuned MgNO backbone, and evaluate correction variants built on the same frozen MgNO within each seed. Corrections are trained with a supervised squared relative L^2 loss plus an optional squared relative-residual term (zero weight in the architecture-only setting), using Adam with cosine learning-rate decay. Checkpoints are selected by validation L^2, with epoch zero eligible. We report hyper-parameter information in the code supplementary. We report per-sample relative L^2, full finite-difference H^1 (nodal spacing $1/100$), and the physical residual $\|A\widehat{u} - b\|_2/\|b\|_2$; complex predictions are compared against the real reference without discarding imaginary error. Means and sample SDs are computed from three seed-level means on identical samples.

Hardware and software. All experiments run on a single NVIDIA RTX 5090 with PyTorch `2.12.0+cu128`, using seeds 0, 1, and 2.

5.2 Accuracy and Efficiency Gains

Overall improvement. Table 1 and Figure 3 show that our correction methods consistently outperform backbones on the Caltech test split. While FNO2D and U-NO yield the highest L^2 errors and lower physical residuals than Tuned MgNO, this highlights that residual minimization does not guarantee solution accuracy. Tuned MgNO serves as the baseline with an L^2 error of 9.26×10^{-3}. Our corrections yield substantial gains: the seen-data ablation achieves a $+6.8\%$ improvement, the unseen-trained correction delivers $+31.2\%$ (0.00630), and operator features provide the best L^2 accuracy at 0.00625 ($+31.9\%$). Relative H^1 errors follow the same trend. Importantly, these accuracy improvements come with nearly unchanged physical residuals ($\approx$ 66–68) and similar computational costs, confirming that the gains stem from correction rather than residual fitting or increased run-

Table 1: Consolidated Caltech comparisons. Errors are displayed in units of 10^{-3}; lower is better. Uncertainty is sample SD across three seed means. Panels must not be pooled into one ranking.

Method	Params. (M)	Relative L^2 ($\times 10^{-3}$)	Relative H^1 ($\times 10^{-3}$)	Residual ($\times 10^{-3}$)	L^2 gain (%)
A. Backbone benchmark: test					
FNO2D	1.333	17.168 ± 0.263	17.077 ± 0.268	47.588 ± 3.387	-87.71 ± 26.59
U-NO	16.394	12.905 ± 0.032	12.851 ± 0.039	26.706 ± 1.121	-41.00 ± 18.83
Tuned MgNO	8.710	9.263 ± 1.236	9.267 ± 1.250	66.875 ± 3.715	0.00 ± 0.00
B. Learned-basis corrections: test					
Tuned MgNO	8.710	9.262 ± 1.237	9.266 ± 1.251	67.024 ± 3.753	≈ 0
Generic	8.730	9.262 ± 1.237	9.266 ± 1.251	67.026 ± 3.755	≈ 0
Galerkin	8.717	9.214 ± 1.252	9.266 ± 1.251	66.852 ± 3.791	0.519
RM	8.717	9.214 ± 1.252	9.266 ± 1.251	66.850 ± 3.790	0.520
C. Fixed POD, rank 16: validation					
Tuned MgNO	8.710	9.978 ± 1.340	9.985 ± 1.360	68.546 ± 3.720	0.00
POD constant	–	9.979 ± 1.340	9.987 ± 1.350	68.503 ± 3.680	-0.02 ± 0.07
POD learned (seen)	–	9.339 ± 1.180	9.374 ± 1.190	67.614 ± 3.810	6.34 ± 0.77
POD Galerkin (shifted)	–	19.849 ± 0.649	19.805 ± 0.621	74.988 ± 3.850	-101.82 ± 32.20
POD RM (ridge)	–	15.411 ± 1.620	15.412 ± 1.630	59.400 ± 2.880	-54.87 ± 4.55
POD oracle†	–	3.681 ± 0.125	3.705 ± 0.125	71.226 ± 4.790	62.57 ± 6.11
D. Common baseline, POD rank 16: validation					
Tuned MgNO	8.710	9.979 ± 1.340	9.983 ± 1.360	68.380 ± 3.670	0.00
ResCo-Seen	–	9.355 ± 1.180	9.389 ± 1.180	67.503 ± 3.780	6.18 ± 0.90
ResCo	–	6.203 ± 0.226	6.225 ± 0.225	70.357 ± 4.750	37.27 ± 8.04
POD oracle†	–	3.678 ± 0.125	3.699 ± 0.125	71.087 ± 4.730	62.61 ± 6.11
E. Our method: test					
ResCo-Seen	8.853	8.621 ± 1.061	8.652 ± 1.066	66.047 ± 3.801	6.84 ± 1.06
ResCo	8.853	6.302 ± 0.317	6.325 ± 0.315	67.752 ± 4.132	31.20 ± 8.99
ResCo-Op	8.993	6.247 ± 0.264	6.270 ± 0.261	67.740 ± 4.227	31.86 ± 7.99

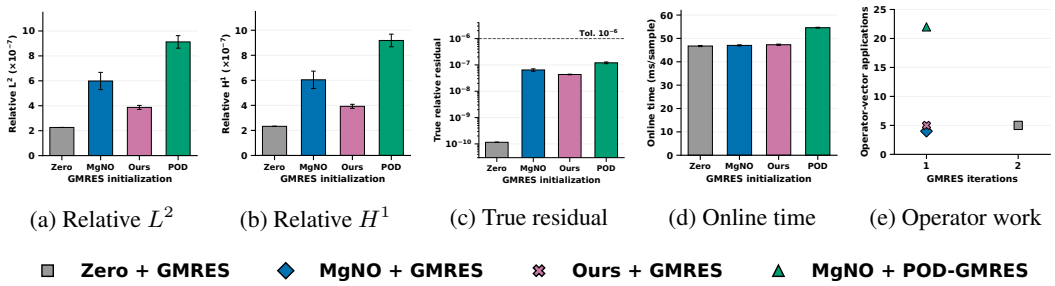
Notes. Panel B gains use the ratio of aggregate means relative to its own MgNO baseline; panels C–D report means and sample SDs of per-seed gains. RM denotes residual minimization. †Oracles use labels and are diagnostic only. Panels C–D use rank 16; no basis rank is inferred for panel B. Panel E reports the proposed variants on the original test split, using the corresponding MgNO baseline in panel A. It is separate from the validation component-ablation study.

time. Our residual-conditioned subspace correction also reduces structured prediction errors in the wavefield. Overall, our approach achieves meaningfully lower L^2 and H^1 for test data than MgNO at similar cost, trading minimal efficiency for significant accuracy gains.

Table 2: **Validation accuracy and online cost of the selected variants.** All errors are scaled by 10^{-3}; smaller is better. Entries are mean $\pm$ sample SD across three seed means. All rows share the frozen 8.71M-parameter MgNO; head counts exclude fixed POD storage.

Method	Head params. (k)	Relative L^2 ($\times 10^{-3}$)	Relative H^1 ($\times 10^{-3}$)	Residual ($\times 10^{-3}$)	Online time (ms/sample)
Frozen MgNO	0	9.979 ± 1.340	9.983 ± 1.360	68.381 ± 3.670	1.486 ± 0.006
Matched generic (POD + residual)	283.128	6.328 ± 0.235	6.348 ± 0.236	70.020 ± 4.450	1.925 ± 0.003
ResCo	143.704	6.190 ± 0.213	6.212 ± 0.213	70.272 ± 4.630	1.874 ± 0.019
ResCo-Op	283.128	6.051 ± 0.219	6.071 ± 0.220	70.476 ± 4.530	7.645 ± 0.015

Online time is synchronized and batch-amortized per sample, including backbone inference, applicable physics/features, head, and reconstruction; it excludes I/O, labels, metrics, and offline preparation. These values are not isolated single-query latencies.



(a) Relative L^2 (b) Relative H^1 (c) True residual (d) Online time (e) Operator work

Figure 4: **Accuracy and computational cost of GMRES pipelines.** Results are averaged over 200 test systems and three seeds; error bars show sample SD across seed means. (a,b) Relative field errors; (c) true residual (dashed line: 10^{-6} tolerance); (d) Batch-amortized online time; (e) Operator-vector applications vs. inner GMRES iterations. "Zero," "MgNO," "Ours," and "POD" are solver pipelines; standalone neural predictions are excluded. Ours+GMRES achieves higher accuracy than MgNO+GMRES with similar online cost.

5.3 Ablation and Component Analysis

Using the same frozen MgNO, we compare components on 200 validation samples across three seeds (Table 2). Basis controls have rank 16; operator/generic heads are parameter-matched. Gains are calculated per seed against the stated control. Residuals are inputs, with no residual-loss penalty.

Reduced-matrix conditioning. Adding $Q^* A_h Q$ and related operator features reduces L^2 by $4.38 \pm 0.13\%$ relative to the matched generic POD/residual head, but adds 5.720 ± 0.017 ms/sample. Without residual inputs, the effect is $1.03 \pm 3.81\%$.

Residuals, basis, and coefficient learning. For the generic POD head, adding residual inputs reduces L^2 by $33.81 \pm 9.27\%$. Replacing a random basis with POD gives $35.84 \pm 8.48\%$ at fixed head and residual inputs; residuals give essentially no benefit with a random basis. Learned coefficients outperform constant coefficients by $37.34 \pm 7.95\%$. Training the same primary head on backbone-unseen rather than seen inputs improves L^2 by $33.18 \pm 7.93\%$ at matched head-training counts and budgets.

Accuracy and online cost. Our primary correction improves L^2 by $37.27 \pm 7.95\%$ over MgNO with 1.65% additional parameters and 0.388 ± 0.023 ms/sample overhead. The operator extension is slightly more accurate but takes $4.08\times$ the online time. POD, residuals, and unseen-error training jointly contribute; these contrasts identify no unique dominant component.

5.4 Spectral Comparison

We compare frozen MgNO and ReSCo, its backbone-unseen-trained, residual-conditioned rank-16 POD correction, on eight validation media with three neural seeds. The reference is a complex128 sparse-LU solution of the actual Neumann Q1 Helmholtz system. All spectral field norms use the geometric consistent mass matrix $G = M_h(1)$, distinct from the coefficient-weighted reaction matrix and the nodal L^2/H^1 metrics. Spatial bands use $K_h \phi_j = \mu_j G \phi_j$ and wavenumber $\sqrt{\mu_j}$; Helmholtz modes instead satisfy $A_h v_j = \lambda_j G v_j$. For G-orthonormal modes, the discrete field obeys $v_j^* G u = (v_j^* b)/\lambda_j$. Thus small $|\lambda_j|$ and low spatial wavenumber are different notions.

Field spectra agree most closely in the energetic bands. Both models reproduce the overall spatial energy distribution (Figure 5). However, overlapping energy spectra alone conceal coefficient errors. Band 2, with $\sqrt{\mu} \in [40.82, 61.24)$, contains 97.15% of the reference field energy. ReSCo reduces its relative coefficient error from 0.00866 to 0.00516 (40.4%), and increases coherence from 0.999950 to 0.999987. This band accounts for 99.5% of the reduction in mean normalized squared field error, computed per medium before aggregation. Agreement remains weaker at the highest spatial frequencies: in band 23, relative error is 1.65 for MgNO and 1.70 for Ours, amplitude ratios are 1.95 and 1.98, and coherence is approximately 0.53 for both. That band carries only 3.27×10^{-9} of reference energy, explaining why large relative errors there coexist with accurate total fields.

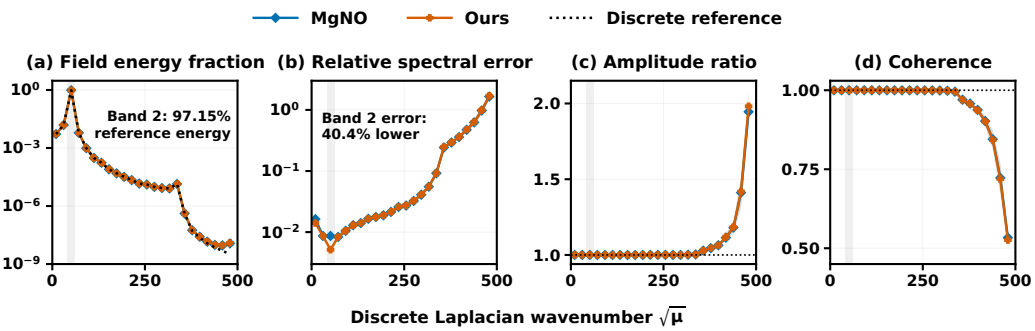


Figure 5: Spatial spectral agreement on eight validation media. Curves and shading show means and sample SD across three seed-level means. The gray strip marks the dominant spatial band. Ours improves dominant-band accuracy, while high-wavenumber mismatch remains.

The correction improves selected Helmholtz components. Table 4 (App.C) shows a 38.8% reduction in Q1 relative L^2 error and a 45.8% reduction in relative error within the 24 Helmholtz modes nearest zero. These modes contain 79.3% of reference energy on average. True residual is essentially unchanged ($0.06403 \rightarrow 0.06432$), so the field improvement does not imply improved PDE consistency. POD alignment supports capacity in this subspace: 84.96% of the mass-orthonormalized basis energy lies in it, versus 0.52% for an equal-rank random control (Table 6). This does not identify POD vectors with Helmholtz eigenvectors or establish learned coefficient accuracy by itself.

5.5 Hybrid solver efficiency

We evaluate whether the proposed correction reduces the cost of converged GMRES solves on 200 shared fresh-test systems across three seeds. Each solver uses restarted GMRES with right preconditioning and the same ILU configuration (drop tolerance 10^{-6}, fill factor 20), restart length 30, and relative true-residual tolerance 10^{-6}; all converge. Our proposed method combined with GMRES achieves markedly lower final relative L^2, H^1, and physical residual errors compared to MgNO+GMRES (see Figure 4). Specifically, it reduces the L^2 error by 35.4% and the H^1 error by 35.0% (see Figure 4a and 4b). Similar benefits appear in the physical residual (see Figure 4c). Both approaches use a single inner iteration (Figure 4e), and the total online computational cost rises only slightly—from 47.015 ms to 47.297 ms per sample—resulting in a minor overhead of just 0.60% (Figure 4d). These results demonstrate that our method provides a significant improvement in solution accuracy while incurring only a negligible increase in computational expense, effectively trading a small amount of efficiency for a meaningful gain in accuracy.

6 Conclusion

We presented ResCo for a frozen Helmholtz neural operator. A fixed POD error basis represents global corrections, while a small head learns their coefficients from the prediction and physical residual. The exact error decomposition separates subspace capacity from coefficient recovery; the experiments support combining a suitable basis, residual inputs and backbone-unseen correction training, rather than attributing the gains to one component alone. The reported original-test comparison yields a 31.20% mean per-seed relative L^2 improvement over MgNO. The validation study yields 37.27% with 0.388 ms/sample additional batch-amortized online cost.

The benefits have several limits. Spectral diagnostics on eight media show improved agreement in energetic bands and the selected near-zero Helmholtz subspace, without uniform high-frequency improvement or a lower physical residual. With strong ILU, all fresh-test iterative arms converge, but adding the correction to MgNO-initialized GMRES preserves one inner iteration and increases total online time from 47.015 to 47.297 ms/sample. Its 35.4% lower endpoint L^2 error is an accuracy benefit, not solver acceleration. Future work should prioritize equal-total-label controls, larger correction-training sets, and training across physical settings, alongside corrections that reduce time to tolerance while retaining field accuracy.

AI assistance. AI-assisted tools were used to support manuscript drafting and language editing, code development and debugging, and figure preparation. The authors reviewed and revised the assisted outputs and take full responsibility for the methodology, experimental results, interpretations, citations, and final manuscript. AI-generated suggestions were treated as development aids rather than scientific evidence.

REFERENCES

Kaushik Bhattacharya, Bamdad Hosseini, Nikola B. Kovachki, and Andrew M. Stuart. Model reduction and neural networks for parametric PDEs. *The SMAI Journal of Computational Mathematics*, 7:121–157, 2021. doi: 10.5802/smai-jcm.74. URL https://smai-jcm.centre-mersenne.org/articles/10.5802/smai-jcm.74/.

Nicolas Boullé and Alex Townsend. A mathematical guide to operator learning. In *Handbook of Numerical Analysis*, volume 25, pp. 83–125. Elsevier, 2024. doi: 10.1016/bs.hna.2024.05.003. URL https://arxiv.org/abs/2312.14688.

Lianghao Cao, Thomas O'Leary-Roseberry, Prashant K. Jha, J. Tinsley Oden, and Omar Ghattas. Residual-based error correction for neural operator accelerated infinite-dimensional Bayesian inverse problems. *Journal of Computational Physics*, 486:112104, 2023. doi: 10.1016/j.jcp.2023.112104. URL https://arxiv.org/abs/2210.03008.

Fukai Chen, Ziyang Liu, Guochang Lin, Junqing Chen, and Zuoqiang Shi. NSNO: Neumann series neural operator for solving Helmholtz equations in inhomogeneous medium, 2024. URL https://arxiv.org/abs/2401.13494. arXiv preprint arXiv:2401.13494.

Zituo Chen, Baiming Zhang, and Sili Deng. Learnable composition for neural operators, 2026. URL https://arxiv.org/abs/2609.03069. arXiv preprint arXiv:2609.03069.

Y. A. Erlangga, C. W. Oosterlee, and C. Vuik. A novel multigrid based preconditioner for heterogeneous Helmholtz problems. *SIAM Journal on Scientific Computing*, 27(4):1471–1492, 2006. doi: 10.1137/040615195. URL https://doi.org/10.1137/040615195.

Juncai He, Xinliang Liu, and Jinchao Xu. MgNO: Efficient parameterization of linear operators via multigrid. In *International Conference on Learning Representations*, 2024. URL https://arxiv.org/abs/2310.19809.

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. arXiv preprint arXiv:2405.19101.

Prashant K. Jha. Residual-based error corrector operator to enhance accuracy and reliability of neural operator surrogates of nonlinear variational boundary-value problems. *Computer Methods in Applied Mechanics and Engineering*, 419:116595, 2024. doi: 10.1016/j.cma.2023.116595. URL https://doi.org/10.1016/j.cma.2023.116595.

Jiwei Jia, Xinliang Liu, Juntao Wang, and Jinchao Xu. McMg: A learned phase-space multi-channel multigrid preconditioner for Helmholtz equation, 2026. URL https://arxiv.org/abs/2606.30495. arXiv preprint arXiv:2606.30495.

Alena Kopaničáková and George Em Karniadakis. DeepONet based preconditioning strategies for solving parametric linear systems of equations. *SIAM Journal on Scientific Computing*, 47(1): C151–C181, 2025. doi: 10.1137/24M162861X. URL https://epubs.siam.org/doi/10.1137/24M162861X.

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 https://jmlr.org/papers/v24/21-1524.html.

Bar Lerer, Ido Ben-Yair, and Eran Treister. Multigrid-augmented deep learning preconditioners for the Helmholtz equation using compact implicit layers. *SIAM Journal on Scientific Computing*, 46 (5):S123–S144, 2024. doi: 10.1137/23M1583302. URL https://arxiv.org/abs/2306.17486.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Andrew Stuart, Kaushik Bhattacharya, and Anima Anandkumar. Multipole graph neural operator for parametric partial differential equations. In *Advances in Neural Information Processing Systems*, volume 33, pp. 6755–6766, 2020. URL https://arxiv.org/abs/2006.09535.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021.

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, 2021. doi: 10.1038/s42256-021-00302-5. URL https://www.nature.com/articles/s42256-021-00302-5.

Lu Lu, Xuhui Meng, Shengze Cai, Zhiping Mao, Somdatta Goswami, Zhongqiang Zhang, and George Em Karniadakis. A comprehensive and fair comparison of two neural operators (with practical extensions) based on FAIR data. *Computer Methods in Applied Mechanics and Engineering*, 393:114778, 2022. doi: 10.1016/j.cma.2022.114778. URL https://arxiv.org/abs/2111.05512.

Hugo Melchers, Victorita Dolean, and Michael Abdelmalik. When can a neural operator replace a coarse solve? architectural principles for two-level preconditioning, 2026. URL https://arxiv.org/abs/2605.19867. arXiv preprint arXiv:2605.19867.

Md Ashiqur Rahman, Zachary E. Ross, and Kamyar Azizzadenesheli. U-NO: U-shaped neural operators, 2022. URL https://arxiv.org/abs/2204.11127. arXiv preprint arXiv:2204.11127.

Enrui Zhang, Adar Kahana, Alena Kopaničáková, Eli Turkel, Rishikesh Ranade, Jay Pathak, and George Em Karniadakis. Blending neural operators and relaxation methods in PDE numerical solvers. *Nature Machine Intelligence*, 6(11):1303–1313, 2024. doi: 10.1038/s42256-024-00910-x. URL https://arxiv.org/abs/2208.13273.

A NOTATION TABLE

Table 3 summarizes the notation for the discrete Helmholtz problem, the components of RESCO, and the training and evaluation protocols. Bold symbols denote discrete vectors, $(\cdot)^*$ denotes the conjugate transpose, and Q denotes the fixed orthonormal correction basis. We distinguish the physical residual $\mathbf{r}_0 = \mathbf{b}_h - A_h\mathbf{u}_0$, which is available at inference, from the solution error $\mathbf{e}_0 = \mathbf{u} - \mathbf{u}_0$, which requires a reference solution. Oracle quantities are used only for diagnosis. Dependence on the medium, frequency, and source configuration is omitted when clear from context.

Table 3: **Notation used throughout the paper.** The discrete operator, right-hand side, and solution depend on the medium, frequency, and source configuration; these dependencies are suppressed when clear from context.

Symbol	Definition
Physical problem and discretization	
$\Omega,\ \partial\Omega$	Physical domain and its boundary.
$\mathbf{x},\ c(\mathbf{x})$	Spatial coordinate and spatially varying wave velocity.
$\omega,\ k(\mathbf{x})$	Angular frequency and local wavenumber, $k(\mathbf{x}) = \omega/c(\mathbf{x})$.
$f,\ g$	Volume forcing and prescribed boundary forcing, respectively.
$u,\ \mathbf{u}$	Continuous wavefield and its discrete reference vector.
$h,\ N_h$	Mesh spacing and number of discrete degrees of freedom.
$A_h,\ \mathbf{b}_h$	Discrete Helmholtz matrix and assembled forcing vector; $A_h\mathbf{u} = \mathbf{b}_h$.
$K_h,\ M_h(c^{-2})$	Stiffness and velocity-weighted mass matrices; $A_h = K_h - \omega^2 M_h(c^{-2})$ for the real FEM setting considered.
$A_{\mathrm{P}},\ A_{\mathrm{H}}$	Poisson and Helmholtz matrices used in the motivating comparison.
$\lambda_j(A),\ \sigma_j(A)$	Eigenvalues and singular values of a matrix A.
Neural prediction and subspace correction	
$G_\theta,\ \theta$	Pretrained neural-operator backbone and its frozen parameters.
$\mathbf{u}_0$	Backbone prediction, $\mathbf{u}_0 = G_\theta(c)$.
$\mathbf{e}_0,\ \mathbf{r}_0$	Backbone error $\mathbf{e}_0 = \mathbf{u} - \mathbf{u}_0$ and physical residual $\mathbf{r}_0 = \mathbf{b}_h - A_h\mathbf{u}_0$.
$r,\ Q$	Correction rank and fixed POD basis $Q \in \mathbb{C}^{N_h \times r}$, with $Q^*Q = I_r$.
P_Q	Orthogonal projector onto the correction subspace, $P_Q = QQ^*$.
$\mathbf{z}$	Inference-time features constructed from the medium, backbone prediction, residual, and their subspace projections.
$C_\phi,\ \phi$	Coefficient-recovery network and its trainable parameters.
$\widehat{\boldsymbol{\alpha}}$	Predicted correction coefficients, $\widehat{\boldsymbol{\alpha}} = C_\phi(\mathbf{z})$.
$\widehat{\mathbf{u}}$	Corrected prediction, $\widehat{\mathbf{u}} = \mathbf{u}_0 + Q\widehat{\boldsymbol{\alpha}}$.
$\boldsymbol{\alpha}^\star$	Oracle coefficients, $\boldsymbol{\alpha}^\star = Q^*\mathbf{e}_0$; require reference labels.
$\mathbf{u}_{\mathrm{or}},\ \mathbf{e}_\perp$	Oracle field $\mathbf{u}_{\mathrm{or}} = \mathbf{u}_0 + P_Q\mathbf{e}_0$ and unrepresented error $\mathbf{e}_\perp = (I - P_Q)\mathbf{e}_0$.
$Z,\ H,\ S,\ \mathbf{g}_r$	Operator features: $Z = A_hQ$, $H = Q^*A_hQ$, $S = Z^*Z$, and $\mathbf{g}_r = Z^*\mathbf{r}_0$.
Training, diagnostics, and evaluation	
$\mathcal{D}_{\mathrm{bb}},\ \mathcal{D}_{\mathrm{corr}}$	Backbone-training and correction-training datasets.
$\mathcal{D}_{\mathrm{val}},\ \mathcal{D}_{\mathrm{test}}$	Validation and designated test datasets.
$\mathcal{L}_{\mathrm{field}},\ \mathcal{L}_{\mathrm{coef}}$	Supervised field-error and coefficient-recovery objectives.
$\mathcal{L}_{\mathrm{res}},\ \lambda_{\mathrm{res}}$	Optional physical-residual loss and its nonnegative weight.
$\rho,\ \eta$	Unrepresented-error ratio $\rho = \|\mathbf{e}_\perp\|_2/\|\mathbf{e}_0\|_2$ and captured-energy fraction $\eta = 1 - \rho^2$.
$E_{L^2},\ E_{H^1},\ R$	Relative field errors in L^2 and H^1, and true relative residual $R = \|A_h\widehat{\mathbf{u}} - \mathbf{b}_h\|_2/\|\mathbf{b}_h\|_2$.
$G,\ F$	Percentage L^2 gain over the matched backbone and fraction of the available oracle L^2 gain recovered.
$B,\ M^{-1}$	Right-preconditioning matrix and its inverse action, with $B = A_hM^{-1}$.
$m,\ J_{\max}$	GMRES restart length and maximum total inner iterations.
$\varepsilon_{\mathrm{rel}},\ \varepsilon_{\mathrm{abs}}$	Relative and absolute true-residual stopping tolerances.
$N_A,\ T_{\mathrm{online}}$	Total operator-vector applications and measured online time, under the stated accounting convention.
$(\cdot)^*,\ \|\cdot\|_2,\ \|\cdot\|_{\mathrm{F}}$	Conjugate transpose, Euclidean vector or spectral matrix norm, and Frobenius matrix norm.

Physical-control definitions. The historical and robustness evaluations use different physical correction rules. Writing $H = Q^*AQ$, $Z = AQ$, and $r_0 = b - Au_0$, the historical controls use shifted Galerkin and ridge-regularized residual minimization:

$$(H - \mathrm{i}\beta I)\alpha_{\mathrm{G}} = Q^* r_0, \qquad \alpha_{\mathrm{RM}} = \arg\min_{\alpha} \left\{ \|Z\alpha - r_0\|_2^2 + \lambda \|\alpha\|_2^2 \right\}, \quad \beta, \lambda > 0. \tag{10}$$

In contrast, the robustness controls use unshifted Galerkin ($\beta = 0$) and residual least squares without a ridge penalty ($\lambda = 0$); the latter uses a numerical singular-value cutoff of `rcond`=10^{-12}. The robustness evaluation also uses regenerated backbone predictions and newly solved references. Consequently, these results are identified by both correction rule and evaluation protocol, rather than treated as repeated measurements of an identical method.

B Theoretical Framework

We develop the spectral and variational theory that motivates the correction architecture. All identities are for the discrete Helmholtz operator on a fixed grid with homogeneous Dirichlet conditions.

B.1 Discrete Helmholtz Operator and Spectral Regimes

On a uniform grid with n^2 interior nodes and spacing $h = 1/(n-1)$, let L be the five-point Laplacian and

$$A_P = h^{-2} L, \qquad A_H = A_P - k^2 I. \tag{11}$$

Both are symmetric; A_P is positive definite and A_H shares its eigenvectors. If $\{(\lambda_j, v_j)\}_{j=1}^{n^2}$ are the eigenpairs of A_P with $0 < \lambda_1 \le \cdots \le \lambda_{n^2}$, then

$$A_H v_j = (\lambda_j - k^2) v_j, \qquad \|A_H^{-1}\|_2 = \frac{1}{\min_j |\lambda_j - k^2|}. \tag{12}$$

We distinguish four regimes: *positive definite* ($k^2 < \lambda_1$), *indefinite* ($\lambda_1 < k^2 < \lambda_{n^2}$, off-spectrum), *near-resonant* ($k^2 \approx \lambda_j$ for some j), and *negative definite* ($k^2 > \lambda_{n^2}$). At $k^2 = \lambda_j$ the operator is singular; the adapter selects k off the discrete spectrum. The scalar $\|A_H^{-1}\|_2$ in equation 12 stratifies all subsequent results. For spatially varying velocity, $A_H = -\nabla\cdot(c\nabla\cdot) - k^2 I$ and the spectrum is no longer a uniform shift; absorbing media make A_H non-Hermitian. Both are treated as robustness variants.

B.2 Residual–Solution Relationship

For any invertible A and any $\widehat{u}$, with $\delta = b - Au$,

$$\|u - \widehat{u}\|_2 \;\le\; \|A^{-1}\|_2 \big(\|b - A\widehat{u}\|_2 + \|\delta\|_2 \big).$$

The bound is not reversible: residual reduction implies solution reduction only up to $\|A_H^{-1}\|_2$, which is large near resonance. Hence small residual improvements can coexist with large or negligible solution changes, the theoretical basis for the observed residual–solution separation.

Two corrections are used. The ridge problem $a_R = \arg\min_a \{\|Za - r_0\|_2^2 + \lambda\|a\|_2^2\}$ with $Z = AQ$, $r_0 = b - Au_0$, yields

$$a_R = (Z^*Z + \lambda I)^{-1} Z^* r_0, \qquad \|r_0 - Z a_R\|_2^2 + \lambda \|a_R\|_2^2 \le \|r_0\|_2^2, \tag{13}$$

so residual cannot increase, while no analogous bound holds for $\|u - (u_0 + Qa_R)\|_2$. The shifted Galerkin problem $(H - i\beta I)a_G = Q^* r_0$ with $H = Q^*AQ$, $\beta > 0$, gives $a_G = (H - i\beta I)^{-1} Q^* r_0$, which inherits the shifted spectrum of A_H restricted to $\mathrm{range}(Q)$; for small β and near-singular H, a_G can be large and the correction unstable.

B.3 Exact Error Decomposition

Let $Q \in \mathbb{C}^{N \times r}$ have orthonormal columns. For $e = u - u_0$,

$$a_* = Q^* e, \quad e_\perp = (I - QQ^*)e \implies \|u - (u_0 + Qa)\|_2^2 = \|e_\perp\|_2^2 + \|a - a_*\|_2^2$$

for all $a \in \mathbb{C}^r$. The first term is the *representation error* (outside $\mathrm{range}(Q)$), the second the *coefficient error*. The oracle minimizer is a_*, giving $u_{\mathrm{oracle}} = u_0 + QQ^*(u - u_0)$ with error $\|e_\perp\|_2$; it is a label-dependent diagnostic floor, not a deployable predictor.

Define sample-mean errors B, C, O of the backbone, corrected prediction, and oracle, and let

$$G = 100\,\frac{B - C}{B}, \qquad F = 100\,\frac{B - C}{B - O}. \tag{14}$$

Since $B = \|e\|^2$ and $B - O = \|a_*\|^2$ at the sample-mean level,

$$G = F \cdot \frac{\|a_*\|^2}{\|e\|^2}. \tag{15}$$

A small G therefore arises from either small F (coefficient recovery fails) or small $\|a_*\|^2/\|e\|^2$ (the POD subspace captures little error energy), separating the two hypotheses tested below.

B.4 Coefficient Error Bounds

With $r_0 = Ae + \delta = Za_* + Ae_\perp + \delta$,

$$a_R - a_* = (Z^*Z + \lambda I)^{-1}\big(\underbrace{Z^*Ae_\perp}_{\text{coupling}} - \underbrace{\lambda a_*}_{\text{bias}} + \underbrace{Z^*\delta}_{\text{noise}}\big),$$

and similarly

$$a_G - a_* = (H - i\beta I)^{-1}\big(Q^*Ae_\perp + Q^*\delta + i\beta a_*\big).$$

The coupling $Z^*Ae_\perp$ is generally nonzero even though $Q^*e_\perp = 0$, because A^*A mixes the unrepresented subspace back into $\mathrm{range}(Q)$. Writing

$$Z^*Ae_\perp = Q^*A_H^*A_H(I - QQ^*)e, \tag{16}$$

and noting $A_H^*A_H$ has eigenvalues $(\lambda_j - k^2)^2$, the coupling is governed by the alignment between the singular vectors of A_H and the POD basis. Coefficient error thus depends on three spectral quantities: the singular values of $Z = AQ$ (conditioning of $Z^*Z + \lambda I$), the projection of $Ae_\perp$ onto $\mathrm{range}(Q)$ (coupling), and the eigenvalues of $H = Q^*AQ$ (Galerkin bias). All three degrade near resonance, motivating stratification by $\|A_H^{-1}\|_2$.

B.5 POD Basis Optimality

Collecting normalized training errors into $E = U\Sigma V^*$, the rank-r POD basis $Q_r = U_{:,1:r}$ satisfies, by Eckart–Young–Mirsky,

$$Q_r \in \arg\min_{Q^*Q = I} \|E - QQ^*E\|_F^2, \qquad \rho_r = \frac{\sum_{j \le r}\sigma_j^2}{\sum_j \sigma_j^2}. \tag{17}$$

This minimizes normalized training projection error, not validation performance or coefficient recovery. The complementary projection satisfies $Q_r^*e_\perp = 0$ and $\|e_\perp\|_2^2 = \|e\|_2^2 - \|Q_r^*e\|_2^2$: $e_\perp$ is invisible to $a_* = Q_r^*e$ yet remains visible to the residual through equation 16.

B.6 Distribution-Shift Bound

Let $\mu_{\mathrm{train}}, \mu_{\mathrm{eval}}$ be the correction-training and evaluation distributions, with $a_*(x) = Q^*e(x)$. For any hypothesis class $\mathcal{H}$ and integral probability metric d,

$$\mathbb{E}_{\mu_{\mathrm{eval}}}\|a_\theta - a_*\|_2^2 \;\le\; \mathbb{E}_{\mu_{\mathrm{train}}}\|a_\theta - a_*\|_2^2 + d(\mu_{\mathrm{train}}, \mu_{\mathrm{eval}}) + \mathfrak{C}(\mathcal{H}), \tag{18}$$

decomposing into fitting error, distribution shift, and complexity. The distributions inherit their structure from (c, ω, k) through the spectrum $\{\lambda_j - k^2\}$: near-resonant samples have larger coefficient magnitudes and different per-mode structure than well-conditioned samples. Training on backbone-unseen errors replaces μ_{train} by a distribution closer to μ_{eval}, reducing the second term without changing the backbone or basis.

Table 4: **Agreement with the discrete Helmholtz solution.** Eight validation media and three neural seeds; means ± sample SD across seed-level means. Ours is frozen MgNO with the backbone-unseen-trained, residual-conditioned rank-16 POD correction. Lower error is better.

Method	Q1 relative L^2 $\times 10^{-3}$	Near-zero error $\times 10^{-3}$	True residual $\times 10^{-3}$	Error in near subspace (%)
MgNO	8.77 ± 0.61	9.11 ± 0.83	64.03 ± 3.75	80.35 ± 5.59
Ours	$\mathbf{5.37 \pm 0.37}$	$\mathbf{4.94 \pm 0.44}$	64.32 ± 4.40	65.66 ± 2.44

Near-zero error uses the 24 modes of $A_h v = \lambda G v$ nearest zero, normalized by reference energy in that subspace. The last column is a fraction of total squared field error, not an accuracy metric. G is the geometric Q1 mass. These errors do not replace nodal L^2/H^1 metrics.

B.7 Architectural Consequences

The reduced matrices $H = Q^*AQ$, $g = Z^*r_0$, $S = Z^*Z$ with $Z = AQ$ carry the operator action on the basis; by Eq. 4–Eq. 5 they are exactly the quantities appearing in the coefficient error. Feeding H, g, S to the head is therefore the natural way to expose the correction to the same spectral information that governs its error. Conditioning the head on $\kappa(A_H) = \|A_H\|_2\|A_H^{-1}\|_2$ (or equivalently $\min_j |\lambda_j - k^2|$) targets the regime dependence of Eq. 2 and Eq. 4.

A complementary branch

$$\widehat{u} = u_0 + Q\widehat{a}_\theta + (I - QQ^*)\mathcal{R}_\phi(c, \omega, u_0, r_0) \tag{19}$$

splits the remaining error into two orthogonal terms,

$$\|u - \widehat{u}\|_2^2 = \|a_* - \widehat{a}_\theta\|_2^2 + \|e_\perp - (I - QQ^*)\mathcal{R}_\phi\|_2^2, \tag{20}$$

so each branch addresses exactly one term of Eq. 3. A multi-level variant $\widehat{u} = u_0 + \sum_{\ell=1}^{L} Q_\ell a_\ell$ with mutually orthogonal Q_ℓ extends the decomposition to per-level coefficient errors plus a final complementary term; we treat it as exploratory. In all cases, any architectural component must be ablated with matched parameter count, data, rank, seeds, and training budget.

C Spectral Diagnostic Definitions and Additional Results

Evaluation and aggregation. The diagnostic uses validation IDs 8200–8207 and seeds 0, 1, and 2; local-response probes use IDs 8200–8203. The same physical media recur across seeds. We compute each metric per medium, average within each seed, and report the mean and sample SD across three seed means. SD measures seed variation conditional on these media, not a confidence interval or variation over 24 independent systems. Percentage reductions in the main text are ratios of method means. No fresh-test arrays are read, and no checkpoint or evaluation protocol is changed for this analysis.

Spatial metrics. Let $c_B = \Phi_B^* G u$ and $\widehat{c}_B = \Phi_B^* G \widehat{u}$ be the reference and predicted coefficients in band B. We report

$$E_B = \frac{\|\widehat{c}_B - c_B\|_2}{\|c_B\|_2}, \qquad R_B = \frac{\|\widehat{c}_B\|_2}{\|c_B\|_2}, \qquad C_B = \frac{|c_B^* \widehat{c}_B|}{\|c_B\|_2 \|\widehat{c}_B\|_2}. \tag{21}$$

The complete spatial expansion partitions $\|u\|_G^2 = u^* G u$ into 24 bands. Figure 5 reports each prediction's band energy divided by its own total field energy, and the analogous reference fraction. Matching normalized energy alone therefore does not establish matching coefficients. Table 5 makes the contrast between dominant and weak bands explicit. Bands 17–23 together contain only 4.96×10^{-7} of reference energy on average.

Near-zero subspace and POD coverage. Let V contain the 24 computed G-orthonormal Helmholtz modes with smallest $|\lambda|$. The near-zero error is $\|V^* G(\widehat{u} - u)\|_2 / \|V^* G u\|_2$; its error-energy fraction is $\|V^* G(\widehat{u} - u)\|_2^2 / \|\widehat{u} - u\|_G^2$. The fraction decreases from 80.35% to 65.66%, but fractions alone cannot measure improvement because the denominator also changes. The associated

Table 5: **Dominant and highest spatial bands.** Means ± sample SD across three seed-level means on the same eight media.

Band	Metric	MgNO	**Ours**
2	Relative error	0.00866 ± 0.00067	0.00516 ± 0.00037
2	Amplitude ratio	1.0002 ± 0.0007	1.0006 ± 0.0004
2	Coherence	0.999950 ± 0.000005	0.999987 ± 0.000001
23	Relative error	1.65461 ± 0.09877	1.69523 ± 0.12810
23	Amplitude ratio	1.9450 ± 0.0883	1.9809 ± 0.1131
23	Coherence	0.533345 ± 0.016201	0.525872 ± 0.022559

Band 2 spans $\sqrt{\mu} \in [40.82, 61.24)$ and contains 97.15% of reference energy; band 23 spans $[469.49, 489.90]$ and contains only 3.27×10^{-9} of reference energy on average. Ideal amplitude ratio and coherence equal one.

Table 6: **POD capacity for the near-zero Helmholtz subspace.** Rank-16 POD versus an equal-rank random basis, measured against 24 mass-orthonormal modes. Means ± sample SD across seed-level means.

Basis	Mode coverage (%)	Basis coverage (%)	$\theta_{\min}$ (°)	$\theta_{\max}$ (°)
POD	56.64 ± 2.08	84.96 ± 3.12	3.53 ± 0.27	87.00 ± 1.03
Random (rank 16)	0.34 ± 0.01	0.52 ± 0.01	83.06 ± 0.02	89.02 ± 0.01

The coverage denominators are 24 and 16, respectively; mode coverage cannot exceed 66.67% at this rank. Angles are averaged after computing the per-system minimum or maximum. Alignment measures capacity, not successful learned coefficient recovery.

absolute normalized squared error decreases from 9.79×10^{-5} to 2.18×10^{-5}, whereas error outside this subspace increases slightly from 9.05×10^{-6} to 9.46×10^{-6}. These are means of per-case squared quantities, not squares or products of reported means.

For the rank-16 basis, form $Q_G = Q(Q^*GQ)^{-1/2}$ and $S = V^*GQ_G$. Principal angles satisfy $\cos\theta_i = \sigma_i(S)$; mode coverage is $\|S\|_F^2/24$, and basis coverage is $\|S\|_F^2/16$. Table 6 shows substantial but incomplete alignment. The maximum possible mode coverage at this rank is $16/24 = 66.67\%$.

Restricted local response. Write the admissible input map as $a \mapsto c(a) = 20 + \tanh(a)$. For each of six mass-normalized real latent cosine directions ψ_j, we compare central finite differences of the neural and discrete maps:

$$y_j^{\mathcal{F}}(\epsilon) = \frac{\mathcal{F}(c(a + \epsilon\psi_j)) - \mathcal{F}(c(a - \epsilon\psi_j))}{2\epsilon}, \qquad Y_{\mathcal{F}} = [y_1^{\mathcal{F}}, \ldots, y_6^{\mathcal{F}}]. \tag{22}$$

Here $\mathcal{F}$ is either the neural predictor or the sparse-LU solution map with the same source. The selected latent indices are $(1,0), (0,1), (2,2), (4,0), (4,4), (8,0)$. Restricted gains are the square roots of the eigenvalues of $\mathrm{Re}(Y_{\mathcal{F}}^* G Y_{\mathcal{F}})$, ordered largest first; the real part restricts input combinations to real coefficients. For each medium and seed, we average $|\widehat{\sigma}_j - \sigma_j|/\sigma_j$ equally over the six gains, then aggregate over media and seeds. Singular gain indices do not label individual latent probe directions.

At $\epsilon = 0.001$, Ours improves four of six mean per-gain errors, but the weakest gain's error increases from 1.49% to 3.27%.

Scope of the answer. The local probe examines a six-dimensional real input subspace at four media; the field diagnostic uses eight media and one fixed source policy. It does not measure global neural eigenvalues or a full approximation to A_h^{-1}. The agreement and improvements are descriptive results for this validation subset, with unresolved derivative-step and eigenvalue-cutoff checks. They do not establish statistical significance, generalization across physical settings, or a cause for the historical Galerkin/residual-minimization failures.

Table 7: **Restricted response-gain error across all recorded steps.** Mean relative absolute error (%) over six singular gains and four validation media within each seed, then mean ± sample SD across three seeds. Lower is better.

Step ϵ	MgNO	**Ours**
0.01	4.17 ± 0.65	$\mathbf{2.76 \pm 0.38}$
0.003	2.87 ± 0.46	$\mathbf{2.06 \pm 0.35}$
0.001	2.77 ± 0.45	$\mathbf{2.03 \pm 0.37}$

Each learned response is compared with the discrete response at the same step and along the same six real latent directions. These are finite-step diagnostics, not verified full Jacobian spectra.

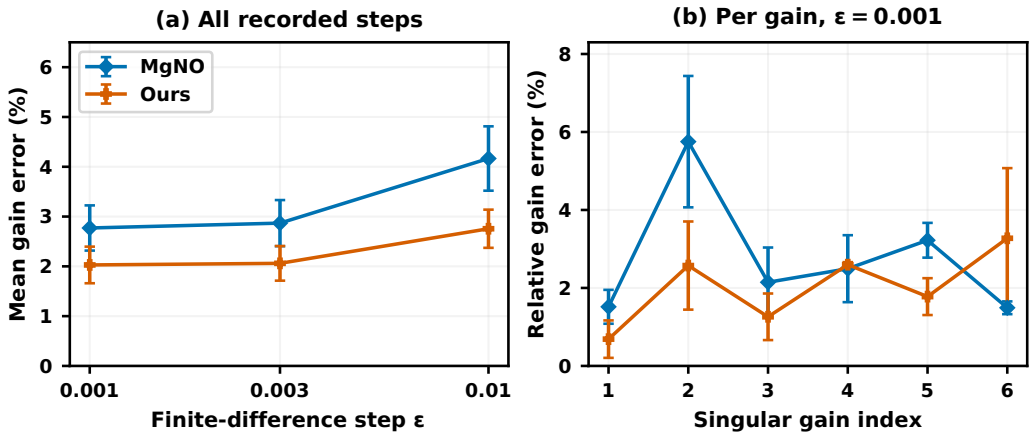


Figure 6: **Restricted local response gains.** Error bars are sample SD across three seed means on four media. (a) All recorded steps are shown; the mean gain error of Ours is lower at each step. (b) At $\epsilon = 0.001$, improvements are not uniform: Ours has lower average error in four of six gains, while the weakest gain is less accurate. This panel describes that recorded step, not a selected or certified Jacobian estimate.

D HYBRID SOLVER EFFICIENCY

On the locked fresh test (200 shared systems; three seeds), we ask whether our residual-conditioned POD correction accelerates GMRES (Table 8; Figure 8). With the validation-selected strong ILU, all iterative arms satisfy the relative true-residual tolerance 10^{-6}. MgNO initialization reduces GMRES from two iterations to one, but total online time increases from 46.740 to 47.015 ms/sample. **Ours + GMRES** also takes one iteration, uses five rather than four operator-vector applications, and raises total time to 47.297 ms/sample (+0.60%). Its final L^2 error is 35.4% lower than that of MgNO + GMRES, but this is an endpoint-accuracy benefit, not acceleration. ILU setup accounts for 84.0% of the MgNO-initialized pipeline, leaving little iterative work to save. **Ours (one-shot)** costs 1.995 ms/sample but has residual 6.78×10^{-2} and does not meet the solver tolerance. Thus, **adding the proposed correction does not reduce total runtime or computational work relative to MgNO-initialized GMRES**; its main benefit remains inexpensive approximate field prediction.

E ROBUSTNESS ACROSS SETTINGS: SCOPE, RESULTS, AND LIMITATIONS

Scope and frozen-model protocol. We test transfer from $(n, \omega, s) = (101, 1000, 1)$, where n is grid nodes per direction, ω angular frequency, and s source count. The tuned MgNO, rank-16 POD correction space, and coefficient head remain fixed; **Ours (unseen)** denotes training on backbone-unseen media, not adaptation during evaluation. We vary $n \in \{81, 101, 151, 201\}$, $\omega \in \{800, 900, 1000, 1100, 1200\}$, or $s \in \{1, 2, 4\}$, changing one axis at a time: ten distinct settings, not a full factorial study. Resolution transfer uses the reported explicit grid transport and newly solved Q1 references. No setting-specific retraining is performed.

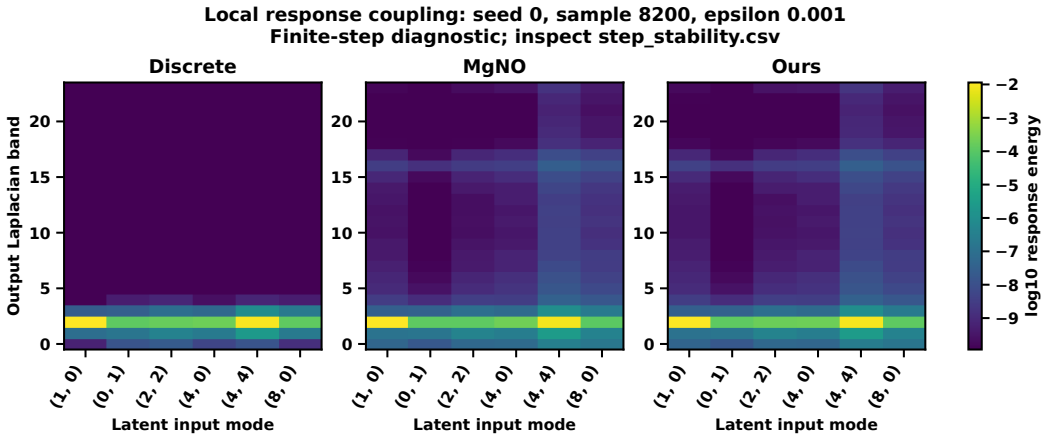


Figure 7: **Local-response coupling example.** Seed 0, validation medium 8200, $\epsilon = 0.001$; columns are latent input probes and rows are output spatial bands. The reference response is concentrated near band 2. Both neural methods reproduce that dominant response but retain excess energy in weak spatial bands.

Table 8: **Locked fresh-test accuracy and computational cost.** Results are mean ± sample SD across three seed-level means on 200 shared systems; SDs are not confidence intervals. Panel A reports converged solver pipelines, while Panel B reports one-shot prediction. Error scales differ between panels. Bold method names identify the proposed methods.

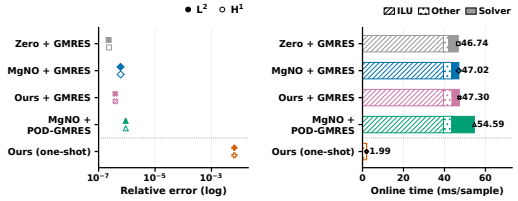
Method	Relative L^2 ($\times 10^{-7}$)	Relative H^1 ($\times 10^{-7}$)	True residual ($\times 10^{-7}$)	Inner iters.	Op. vecs.	Online (ms/sample)	Solver (ms/sample)	Conv. (%)
A. Solver pipelines: true-residual tolerance 10^{-6}								
Zero + GMRES	2.259 ±0	2.327 ±0	0.001152 ±0	2	5	46.740 ±0.241	4.656 ±0.017	100
MgNO + GMRES	5.987 ±0.692	6.042 ±0.692	0.6421 ±0.0681	1	4	47.015 ±0.239	3.447 ±0.021	100
ReSCo + GMRES	3.869 ±0.161	3.925 ±0.160	0.4331 ±0.0112	1	5	47.297 ±0.265	3.558 ±0.005	100
MgNO + POD–GMRES	9.123 ±0.503	9.184 ±0.504	1.201 ±0.092	1	22	54.590 ±0.181	11.022 ±0.077	100
B. One-shot prediction: solver tolerance not attained								
Method	Relative L^2 ($\times 10^{-3}$)	Relative H^1 ($\times 10^{-3}$)	True residual ($\times 10^{-3}$)	Inner iters.	Op. vecs.	Online (ms/sample)	Solver (ms/sample)	Conv. (%)
ReSCo (one-shot)	6.451 ±0.182	6.469 ±0.183	67.796 ±4.031	0	1	1.995 ±0.029	0	0

Notes. Online time includes assembly, ILU setup, neural inference where applicable, transfers, and the solve; solver time is part of this total. Times are synchronized and batch-amortized. Operator-vector counts include the correction work. Zero-start GMRES is independent of the neural seed, explaining its zero across-seed accuracy SD.

Reporting. Validation and the subsequent fresh robustness evaluation remain separate; settings were fixed before the latter. With $\overline{E}_{2,k,m}(t)$ the mean relative L^2 error for seed k, method m, and setting t, the gain is

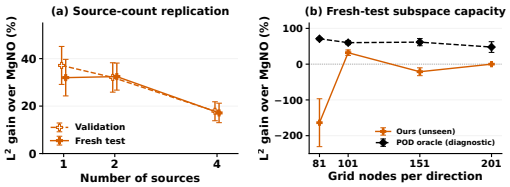
$$G_{k,m}(t) = 100\left(1 - \frac{\overline{E}_{2,k,m}(t)}{\overline{E}_{2,k,\mathrm{MgNO}}(t)}\right), \qquad G_m(t) = \frac{1}{3}\sum_{k=0}^{2} G_{k,m}(t). \tag{23}$$

Reported SDs describe three seed-level quantities, not confidence intervals; gains are not ratios of the displayed aggregate errors. Repeated cases across seeds are not independent systems. The physical



(a) Field accuracy (b) Online cost

Figure 8: **Accuracy and cost at the recorded stopping points.** (a) Relative L^2 (filled) and H^1 (hollow), on a log axis. (b) ILU setup, other online costs, and solver time; labels show mean totals. Whiskers indicate sample SD across seed means. The separated one-shot row is an approximation, not a converged solve. Physical residuals appear in Table 8.



(a) Source sweep (b) Grid transfer

Figure 9: **Replication and limits of coefficient recovery.** (a) Within-seed gains for Ours: hollow/dashed validation and filled/solid fresh-test traces represent distinct evaluations; horizontal offsets separate overlapping markers. (b) Fresh-test learned and oracle gains across grid sizes. Whiskers indicate sample SD across seed means. A useful correction space does not guarantee coefficient transfer; the label-dependent oracle is diagnostic only.

residual is $R = \|A_h\widehat{u} - b\|_2/\|b\|_2$; H^1 retains the nodal finite-difference convention. No fresh-grid prediction failures were reported, but finite outputs do not establish accuracy or solver convergence.

RESCO retains a relative field-accuracy advantage over frozen MgNO as the source count increases, with mean per-seed L^2 gains of 37.13%, 32.07%, and 17.83% for one, two, and four sources. However, absolute errors increase substantially, and physical residuals exceed those of MgNO. Thus, the correction transfers partially to additional sources, but these results do not establish reliable multi-source solution accuracy or physical consistency.

Representation versus coefficient transfer. The label-dependent oracle projects the reference error onto the correction space; it is not an inference method. At $n = 81$, its 71.01% gain contrasts with failed learned correction; at 151 and 201, oracle gains remain 61.56% and 47.82% (Figure 9b). Thus representation alone does not ensure successful coefficient transfer. At changed frequencies, oracle gains fall to 0.45–18.18%, indicating an additional subspace limitation. Unregularized POD RM yields only 0.25–10.17% gains at the changed frequencies on the fresh robustness test, relative to an already inaccurate backbone. At the native setting $(n, \omega, s) = (101, 1000, 1)$, its mean per-seed L^2 gain is $-93.89 \pm 9.6\%$ on the fresh robustness test, compared with $-90.97 \pm 9.9\%$ on the robustness validation set (Table 9). The historical validation gain of $-54.87 \pm 4.55\%$ instead belongs to ridge-regularized POD RM under the earlier evaluation protocol; it is not another estimate for the same unregularized control. The very large validation Galerkin errors and seed spreads are retained in Table 9. These controls identify limitations, not a unique resonance mechanism.

F ADDITIONAL QUALITATIVE COMPARISON

Figure 10 extends Figure 1 to all correction variants and to the GMRES pipelines on the same validation medium.

Table 9: **Robustness to additional sources with frozen models.** Validation results at $n = 101$ and $\omega = 1000$, with no retraining. Each cell shows the mean and, below it, $\pm$ sample SD across three seed-level means; SDs are not confidence intervals. Errors are unscaled. Percentage L^2 gains are computed within each seed relative to the matched MgNO baseline. Bold names identify the proposed method.

Method	Relative L^2	Relative H^1	True residual	L^2 gain (%)
A. One source ($s = 1$): native setting				
Tuned MgNO (frozen)	0.009979 $\pm$0.0013	0.0099831 $\pm$0.0014	0.068381 $\pm$0.0037	0
RESCO-SEEN	0.0093552 $\pm$0.0012	0.0093888 $\pm$0.0012	0.067503 $\pm$0.0038	6.18 $\pm$0.90
RESCO	0.0062033 $\pm$0.00023	0.0062251 $\pm$0.00023	0.070357 $\pm$0.0048	37.13 $\pm$8.0
POD Galerkin (unshifted)	6.3505 ± 11	6.3459 ± 11	13.25 ± 22	-72987 ± 120000
POD RM (unregularized)	0.018968 ± 0.0016	0.018964 ± 0.0016	0.059061 ± 0.0029	-90.97 ± 9.9
POD oracle†	0.0036775 $\pm$0.00013	0.0036986 $\pm$0.00013	0.071087 $\pm$0.0047	62.61 $\pm$6.1
B. Two sources ($s = 2$)				
Tuned MgNO (frozen)	0.96097 $\pm$0.000097	0.9603 $\pm$0.000099	0.69286 $\pm$0.00068	0
RESCO-SEEN	0.92057 $\pm$0.029	0.92044 $\pm$0.029	0.84553 $\pm$0.15	4.20 $\pm$3.1
RESCO	0.65283 $\pm$0.060	0.65261 $\pm$0.060	1.9414 $\pm$0.43	32.07 $\pm$6.2
POD Galerkin	551.73 $\pm$950	551.25 $\pm$940	2526 $\pm$4300	-5.7308×10^4 $\pm 9.8 \times 10^4$
POD RM	0.95952 $\pm$0.00089	0.95886 $\pm$0.00087	0.69039 $\pm$0.00097	0.151 $\pm$0.098
POD oracle†	0.29501 $\pm$0.0068	0.29497 $\pm$0.0065	3.5086 $\pm$0.46	69.30 $\pm$0.71
C. Four sources ($s = 4$)				
Tuned MgNO (frozen)	1.0847 $\pm$0.00013	1.0823 $\pm$0.00013	0.94739 $\pm$0.00072	0
RESCO-SEEN	1.0363 $\pm$0.035	1.0352 $\pm$0.035	1.1784 $\pm$0.25	4.46 $\pm$3.3
RESCO	0.89133 $\pm$0.043	0.89133 $\pm$0.044	2.1269 $\pm$0.23	17.83 $\pm$4.0
POD Galerkin	562.11 $\pm$960	560.87 $\pm$960	3297.6 $\pm$5700	-5.1720×10^4 $\pm 8.9 \times 10^4$
POD RM	1.081 $\pm$0.0027	1.0786 $\pm$0.0026	0.94416 $\pm$0.0015	0.342 $\pm$0.26
POD oracle†	0.42781 $\pm$0.021	0.42906 $\pm$0.020	3.934 $\pm$0.76	60.56 $\pm$2.0

Notes. RESCO uses backbone-unseen correction-training data; RESCO-SEEN is the seen-data ablation. RM denotes reduced residual minimization. †The oracle uses reference labels and is diagnostic only. All 18 rows have zero reported prediction failures; this does not imply accurate solutions or solver convergence. Negative gains indicate deterioration relative to MgNO.

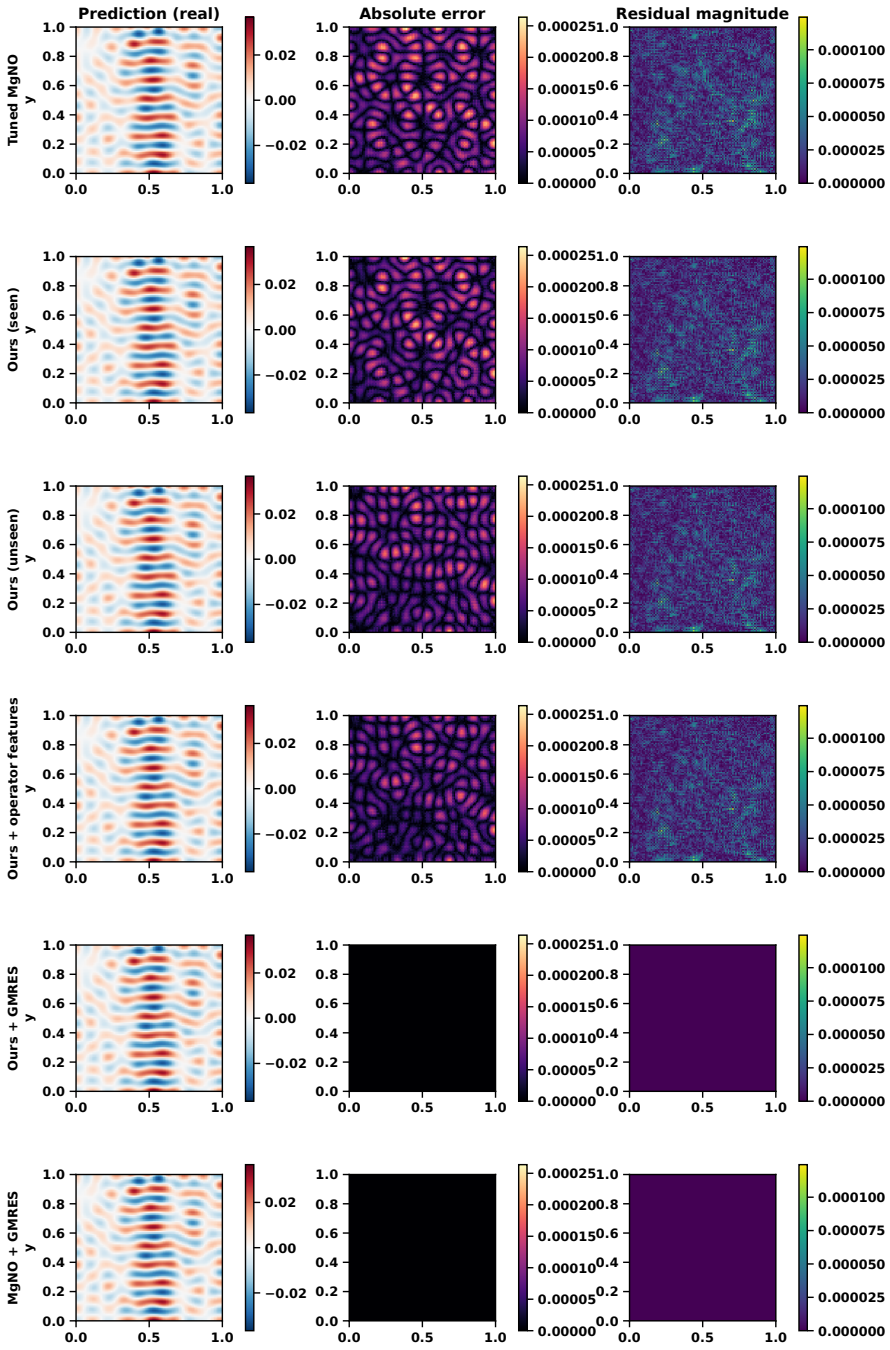


Figure 10: **Per-method comparison on validation sample 8202 (seed 2).** Rows: tuned MgNO, RESCO-SEEN, RESCO, RESCO-OP, RESCO + GMRES, and MgNO + GMRES. Columns: real part of the prediction, pointwise absolute error with respect to the reference, and pointwise magnitude of the physical residual $|\mathbf{b} - A_h\hat{\mathbf{u}}|$. Each column shares one color scale across rows. The one-shot corrections change the error pattern while leaving the residual map visually unchanged, consistent with the nearly constant residuals in Table 1. Both GMRES pipelines reach the 10^{-6} true-residual tolerance, so their error and residual maps lie below the display floor and appear uniform; their differences in endpoint accuracy are reported in Table 8.