DIAGONAL-PRESERVING POWER-COHERENCE PARAMETERIZATION FOR LEARNED COVARIANCE SOURCE IMAGING

Anonymous authors
Paper under double-blind review

ABSTRACT

Learned estimators for coherent acoustic imaging predict a source power map from microphone cross-spectral data, but the measurements depend on the full source covariance, of which the map is only the diagonal. We study how a learned estimate should be refined against the measurements within a fixed optimization budget, and show that the choice of covariance coordinates matters. The conventional diagonal-plus-low-rank form couples power and coherence, so updating its coherent factor also changes the map. We adapt a bounded factor-correlation model to complex covariance: the map remains the covariance diagonal, and positive semidefiniteness holds without a dense projection. Both forms describe the same trace-normalized covariance family and can start from the same physical covariance, so matched controls isolate the effect of coordinates from learned initialization and learning rate selection. With separate development-selected rates for power and factors, power–coherence (PC) refinement lowers shape error by 26.5–30.0% and raises spatial overlap by 11.3–14.6 percentage points after 80 updates without a map anchor, across eight paired seeds and three initializations. The largest gains from CNN starts occur on coherent lines and arcs, with consistent gains on synthetic blade mixtures. With an anchor, PC lowers the joint objective by 17.4–18.6%. An independent time-domain simulator also shows spatial gains without retuning, while Gaussian and Fourier operators show objective reductions. These results identify optimization coordinates as a design choice for refining learned estimates in covariance inverse problems, even when the conventional form attains a lower measurement residual.

1 INTRODUCTION

Microphone array imaging often models sources as incoherent monopoles with a nonnegative power map. Aeroacoustic fields can also contain coherent and extended sources (Merino-Martínez et al., 2019; 2020). With coherence, measurements depend on the full source covariance. Its diagonal gives source power, and its off-diagonal entries describe mutual coherence.

Low-rank factors compactly represent coherence. A covariance parameterization specifies the variables used to construct and optimize it. In the usual form $\mathrm{diag}(p) + LL^*$, p is not the covariance diagonal: pixel i has power $p_i + \|L_i\|_2^2$. Updating L can change the map at fixed p. This coupling differs from measurement ambiguity, where different source fields produce the same measurements.

Our power–coherence (PC) parameterization adapts a classical factor-correlation model to complex covariance. The normalized map p ($p_i \geq 0$, $\sum_i p_i = 1$) scales a bounded unit-diagonal correlation matrix. Figure 1(a,b) shows the difference: Conventional factor updates can change power, but PC keeps it fixed at a given p. Both forms stay positive semidefinite and allow efficient projection. In Figure 1(c), a CNN initializes power and coherence for explicit CSM refinement. A map anchor retains spatial information from the initial estimate.

Learned initialization followed by explicit, physics-based refinement is a common design for inverse problems: a network supplies a fast estimate, and a few optimization steps enforce measurement consistency. Our question is whether the coordinates used for those steps can improve the source map when the covariance family, initial covariance and refinement objective are held fixed. This matters

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.

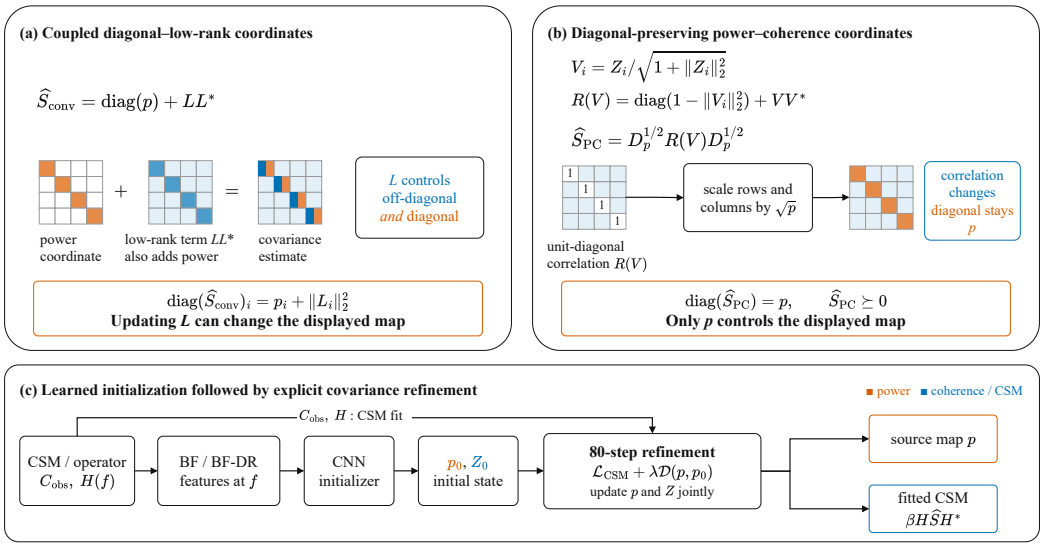


Figure 1: Power–coherence parameterization and inference. (a) Conventional factor updates can change the displayed power map. (b) PC preserves $\mathrm{diag}(\widehat{\mathbf{S}}) = p$ when coherence changes at fixed p. Both forms are positive semidefinite. (c) A CNN initializes power and coherence. An 80-step explicit refinement fits the microphone CSM with a map anchor weighted by $\lambda \geq 0$.

because a small measurement residual need not identify the right spatial distribution. We separate the CNN's spatial information from the refiner's contribution, then compare both parameterizations after giving each the same power and factor rate search. The main evidence is improved shape and overlap for complex sources.

Our contributions are as follows:

- We adapt factor-correlation structure to complex covariance, separating power and coherence while keeping the covariance PSD and projection efficient. Both forms represent the same covariance family after trace normalization, and in PC changing coherence at fixed power does not change the map-anchor penalty (Section 4).
- Matched controls separate learned initialization, parameterization and rate selection. After symmetric power and factor rate searches, PC improves source shape and overlap without a map anchor, and joint objectives and overlap with an anchor (Section 5).
- Source-family and power-distribution analyses locate the main spatial gains on coherent lines, arcs and blade mixtures from CNN starts. Further comparisons show spatial gains on an independent time-domain simulator without retuning, objective reductions with Gaussian and Fourier operators, and lower regularized objectives on measured wind turbine proxy CSMs (Sections 5 and 6).

Code, configurations, model states and numerical results are available at Power Coherence.

2 Related Work

Classical acoustic source mapping. DAMAS deconvolves beamforming maps (Brooks & Humphreys, 2006), with wind turbine applications (Xue & Yang, 2024) and sparse variants (Yardibi et al., 2008). CLEAN-SC removes coherent contributions (Sijtsma, 2007). MACS fits covariance for uncorrelated and correlated sources (Yardibi et al., 2010). Generalized inverse beamforming handles distributed and multipole fields (Suzuki, 2011).

Covariance parameterizations. Earlier work separates scale and correlation (Barnard et al., 2000) and studies diagonal-plus-low-rank identifiability (Saunderson et al., 2012). Borsdorf et al. (2010) study real factor-correlation matrices and nearest correlation matrix algorithms. We adapt this

structure to Hermitian covariance and compare refinement from matched starts within the same covariance family.

Learning-based acoustic imaging. Learned estimators map CSMs to source distributions (Ma & Liu, 2019), infer grid-free coordinates and strengths (Kujawski et al., 2019; Kujawski & Sarradj, 2022), accelerate deconvolution (Lobato et al., 2023), and combine beamforming with frequency and position information (Jia et al., 2025). Our CNN initializes power and coherence for explicit refinement.

Learned optimization and initialization. Unrolled solvers learn updates (Gregor & LeCun, 2010; Monga et al., 2021; Xiao et al., 2020), with learning guarantees for unrolled approximate message passing under Gaussian sensing and product-prior assumptions (Karan et al., 2024). Plug-and-play methods use denoisers in ADMM (Venkatakrishnan et al., 2013); pseudo-contractive denoisers give fixed-point convergence (Wei et al., 2024), and learned proximal networks exactly realize proximal operators of learned regularizers (Fang et al., 2024). Neural inverse operators map observed response operators to unknown fields (Molinaro et al., 2023), learned inverse observation operators initialize variational solvers (Frerix et al., 2021), learned starts support fixed-budget optimization (Sambharya et al., 2024), null-space learning preserves measurements (Schwab et al., 2019), and learning across incomplete operators addresses identifiability (Tachella et al., 2022).

Liu et al. (2023) gradually increase the weight of a pretrained generative prior and initialize each optimization from the previous solution. ReSample enforces hard data consistency by explicit optimization during latent diffusion sampling (Song et al., 2024), PETAL embeds physical linearizations in a learned forward emulator (Jin et al., 2023), and InverseBench compares plug-and-play diffusion methods across scientific inverse problems (Zheng et al., 2025). Our question is how covariance parameterization affects finite-budget refinement from learned or deterministic starts under the same objective and fixed forward model.

3 Problem Formulation

Let M microphones observe N candidate source locations. At frequency f, $\mathbf{C}_{\mathrm{obs}}(f) \in \mathbb{C}^{M\times M}$ is the microphone CSM and $\mathbf{H}(f) \in \mathbb{C}^{M\times N}$ the steering matrix. With noise CSM $\mathbf{N}(f)$, coherent source imaging uses a source covariance,

$$\mathbf{C}_{\mathrm{obs}}(f) \approx \mathbf{H}(f)\mathbf{S}(f)\mathbf{H}(f)^* + \mathbf{N}(f), \qquad \mathbf{S}(f) \succeq 0, \tag{1}$$

where $\mathrm{diag}(\mathbf{S})$ is the source power map. The incoherent approximation restricts $\mathbf{S}$ to a diagonal matrix. We fit the observed CSM without separately estimating or subtracting $\mathbf{N}$. Noise and forward model mismatch can both affect the fitted covariance and its residual. To avoid a dense $N \times N$ covariance, normalize a nonnegative map as $p = \mathbf{x}/(\mathbf{1}^\top \mathbf{x})$ and use a low-rank factor:

$$\widehat{\mathbf{S}}_{\mathrm{conv}} = \mathrm{diag}(p) + \mathbf{L}\mathbf{L}^*, \qquad \mathbf{L} \in \mathbb{C}^{N\times r}, \quad r \ll N. \tag{2}$$

This covariance is positive semidefinite and inexpensive to project, but its diagonal is $p_i + \|L_i\|_2^2$, so the same factor controls coherence and reconstructed power. Every CSM objective includes a nonnegative best-fit scale for absolute amplitude.

With more source locations than microphones ($N > M$), different power maps can produce the same microphone CSM (Appendix A). The learned estimate supplies spatial information, and its map anchor guides refinement among measurement-compatible solutions. PC makes this anchored map identical to the covariance diagonal.

4 Method

At each frequency, $\mathbf{H}$ encodes spherical spreading and narrowband propagation between microphones and scan points. The CNN takes conventional and diagonal-removal beamforming maps with a frequency encoding. Softplus output and sum normalization yield $p \in \mathbb{R}_+^N$ with $\mathbf{1}^\top p = 1$. A parameter-matched convolutional head produces complex rank-r factors $F_\theta(p) \in \mathbb{C}^{N\times r}$. PC and conventional training give $Z_0 = F_\theta(p)$ and $L_0 = D_p^{1/2} F_\theta(p)$, respectively.

4.1 DIAGONAL-PRESERVING POWER–COHERENCE COVARIANCE

The factor-correlation form below is a complex-valued version of the real construction studied by Borsdorf et al. (2010). We smoothly rescale network output rows to have norms below one:

$$V_i = \frac{Z_i}{\sqrt{1 + \|Z_i\|_2^2}}, \qquad R(V) = \operatorname{diag}(1 - \|V_i\|_2^2) + VV^*. \tag{3}$$

The source covariance is then

$$\widehat{\mathbf{S}}_{\mathrm{PC}} = D_p^{1/2} R(V) D_p^{1/2}, \qquad D_p = \operatorname{diag}(p). \tag{4}$$

Because $1 - \|V_i\|_2^2 > 0$, $R(V)$ is the sum of a nonnegative diagonal matrix and VV^*, so it is PSD. Its ith diagonal is $(1 - \|V_i\|_2^2) + \|V_i\|_2^2 = 1$. This gives $\widehat{\mathbf{S}}_{\mathrm{PC}} \succeq 0$ and $\operatorname{diag}(\widehat{\mathbf{S}}_{\mathrm{PC}}) = p$. With $W = D_p^{1/2} V$ and $d_i = p_i(1 - \|V_i\|_2^2)$, we write $\widehat{\mathbf{S}}_{\mathrm{PC}} = \operatorname{diag}(d) + WW^*$ and obtain

$$\mathbf{H}\widehat{\mathbf{S}}_{\mathrm{PC}}\mathbf{H}^* = (\mathbf{H}W)(\mathbf{H}W)^* + \mathbf{H}\operatorname{diag}(d)\mathbf{H}^*, \tag{5}$$

so projection costs $O(MNr + M^2N)$ without forming an $N \times N$ matrix. This is the same asymptotic cost as the conventional diagonal-plus-low-rank form.

Equivalence after trace normalization. For $\mathbf{S}_{\mathrm{L}} = \operatorname{diag}(q) + LL^*$ with $q_i > 0$ and $t = \operatorname{tr}(\mathbf{S}_{\mathrm{L}})$, set $p_i = (q_i + \|L_i\|_2^2)/t$ and $Z_i = L_i/\sqrt{q_i}$. Equations 3–4 give $\mathbf{S}_{\mathrm{PC}} = \mathbf{S}_{\mathrm{L}}/t$. The converse transformation and boundary cases are derived in Appendix A. Both forms represent the same covariance family, including limits, with nonunique factors. We fit CSM amplitude separately and compare both forms from the same initial covariance with equal update counts.

4.2 TRAINING AND COVARIANCE REFINEMENT

For a real map u, $\nu(u) = \max(u, 0)/\mathbf{1}^\top \max(u, 0)$ clamps negative entries and normalizes the sum, with a 10^{-12} denominator floor. With $\tilde{p} = \nu(p)$ and $\tilde{q} = \nu(q)$, the map loss is $\mathcal{L}_{\mathrm{map}}(p, q) = 0.60\|\tilde{p} - \tilde{q}\|_1 - \sum_i \tilde{q}_i \log\big(\max\{\tilde{p}_i, 10^{-8}\}\big)$. The training objective is

$$\mathcal{L}_{\mathrm{map}}(\operatorname{diag}(\widehat{\mathbf{S}}), p_{\mathrm{true}}) + 0.8\mathcal{L}_{\mathrm{CSM}}, \quad \mathcal{L}_{\mathrm{CSM}} = \min_{\beta \geq 0} \frac{\|\mathbf{C}_{\mathrm{obs}} - \beta\mathbf{H}\widehat{\mathbf{S}}\mathbf{H}^*\|_F}{\|\mathbf{C}_{\mathrm{obs}}\|_F}. \tag{6}$$

At inference, the CNN weights remain fixed. We use a complex valued variant of Adam (Kingma & Ba, 2015) to jointly update each observation's power logits and complex factor parameters (L or Z) for 80 iterations. For each complex factor entry, the real and imaginary components share a second moment based on the squared gradient magnitude $|g|^2$. With $\pi(\mathbf{S}) = \nu(\operatorname{diag}(\mathbf{S})) = \operatorname{diag}(\mathbf{S})/\mathbf{1}^\top \operatorname{diag}(\mathbf{S})$ for $\mathbf{S} \neq 0$, both parameterizations minimize

$$\mathcal{L}_{\mathrm{CSM}}(\mathbf{S}) + \lambda\mathcal{D}(\pi(\mathbf{S}), \pi(\mathbf{S}_0)), \qquad \mathcal{D}(a, b) = 0.60\|a - b\|_1 + \mathrm{KL}(b\|a). \tag{7}$$

Here, $\lambda \geq 0$ controls the strength of the power map anchor relative to CSM fitting. When corresponding map entries are equal, we choose the zero subgradient of the ℓ_1 term. We compare unanchored refinement ($\lambda = 0$) and refinement with the learned map anchor ($\lambda = 0.05$).

Proposition 1 (Separation of anchor and coherence)**.** *Let $s = \operatorname{diag}(\mathbf{S})$ be the unnormalized source power vector. In PC with sum-normalized power, $\pi(\mathbf{S}) = p$, so the anchor $\mathcal{D}(\pi(\mathbf{S}), \pi(\mathbf{S}_0))$ has zero derivative with respect to Z at fixed p. In the conventional parameterization, $s_j = q_j + \|L_j\|_2^2$, and the anchor's real Euclidean gradient with respect to factor row L_j is $2u_jL_j$ with $u_j = \partial\mathcal{D}/\partial s_j$, so the anchor rescales factor rows radially (Appendix A).*

CSM fitting still couples power and coherence, so anchor separation does not guarantee better reconstruction.

Why the update paths differ. Write the trace-normalized covariance in real coordinates $s = \phi(\theta)$, where θ collects the power logits and factor entries (L or Z), and s stacks the normalized diagonal d and off-diagonal entries o. For a smooth objective $\mathcal{L}(s)$, one gradient step on θ moves the covariance by $s_{t+1} = s_t - \eta A_\theta A_\theta^\top \nabla_s \mathcal{L} + O(\eta^2)$ with $A_\theta = D\phi(\theta)$ (Appendix A). At a matched covariance, both parameterizations share s_t and $\nabla_s\mathcal{L}$ but generally have different induced preconditioning operators

$A_\theta A_\theta^\top$, so their steps can differ in direction and scale. In PC, $\partial d/\partial Z = 0$, so the off-diagonal residual moves the map only through the power block. In Conventional, the normalized diagonal also depends on L, so this residual can move the map through the factor block. Proposition 1 is the case where $\mathcal{L}$ is the anchor. This explains the update structure, not the endpoint: Adam rescales the coordinates of θ, the paths are nonlinear, and the spatial consequences are established empirically in Section 5.

Paired refiners share trace-normalized $\mathbf{S}_0$, anchor, rank, objective and update budget. Rates are selected on development observations and fixed across evaluation geometries. We compare one rate per parameterization with separate power and factor rates under the selection rules in Section 5. The single-rate anchored reference selects 0.002 (Conventional) and 0.05 (PC) using three disjoint circular-array development seeds, both CNN origins and six families. Each CNN has 43,397 parameters (Appendix B).

5 Synthetic Benchmark

5.1 Experimental setup

We construct a synthetic acoustic benchmark with known source covariances. The 21×17 scan grid spans 1.2×0.9 m at 1.4 m range. Training uses a 32-microphone circle of radius 0.45 m. Evaluation uses both this circle and an unseen four-arm spiral with the same microphone count and outer radius. Six families cover one to three point-source centers, lines, arcs, coherent lines/arcs, and turbine-blade-like mixtures. We map continuous source supports onto the grid with a Gaussian kernel ($\sigma = 0.045$ m), then generate $C_{\text{obs}} = HS_{\text{true}}H^* + \eta I$ using the same grid operator. The observations are analytic grid-model covariances rather than finite-snapshot estimates or independently propagated continuous-source measurements.

For each coherent component with power map p_c, we mix an incoherent covariance and a phase-bearing rank-one term as $(1-\rho)\,\mathrm{diag}(p_c) + \rho vv^*$ with $|v_i|^2 = (p_c)_i$, so coherence changes off-diagonal structure at fixed component power; ρ is uniform on $[0.75, 0.98]$ for coherent lines and $[0.72, 0.96]$ for coherent arcs. Blade mixtures sum component covariances without cross-component coherence. S_{true} is trace-normalized, and white microphone noise power η is log-uniform on $[10^{-5}, 2\times10^{-3}]$.

Each parameterization-specific CNN receives 2,400 AdamW updates (Loshchilov & Hutter, 2019), with batch size 8 and factor width $r = 6$. r is neither the number of sources nor the total covariance rank. Frequency centers are 1.2, 2.2, and 3.4 kHz with $\pm8\%$ jitter. Each of eight paired seeds supplies 16 observations per family and geometry. Appendix B specifies generation, training and aggregation.

Performance is measured by normalized shape L1, best-scaled relative CSM residual, and the *false-positive active-pixel fraction* (FP-pix). Each map is thresholded at -20 dB relative to its own peak power. FP-pix is the fraction of predicted active pixels that are inactive in the reference map, recall is the fraction of reference active pixels recovered, and intersection-over-union (IoU) is the number of pixels active in both maps divided by the number active in either map. Metrics are averaged within seed before paired reductions and two-sided Student-t intervals are calculated over eight seeds. Exact two-sided sign-flip tests are also reported.

5.2 Learned Spatial Information and Matched Refinement

We compare deterministic beamforming/eigenfactor and two trained CNN starts on fresh observations. Paired refiners share $\mathbf{S}_0$, rank six and a measurement-only objective ($\lambda = 0$). The deterministic factor carries 30% of the covariance trace, allowing immediate coherence updates. Three disjoint development seeds select rates before evaluation on eight seeds, six families and both arrays at 0, 20 and 80 updates (Appendix C.1).

The CNN supplies spatial information before refinement. Its initial covariance map has shape error about 0.63, compared with 1.17 for the deterministic start, despite a higher initial CSM residual (Table 1). At 80 updates, PC retains and improves this spatial estimate, with shape error about 0.44 and IoU about 0.70 under the single-rate policy. The deterministic PC result is less accurate, so the parameterization does not replace learning. This motivates the next control: does PC still improve spatial reconstruction with separately tuned power and factor rates?

Table 1: Learned initialization and refinement without an anchor, with one rate per parameterization. Entries are Conventional/PC means over eight seeds, six families and two arrays. Lower CSM and shape errors and higher recall and IoU are better. Zero updates denotes the common covariance map, not necessarily the direct CNN output.

Initializer	Updates	CSM	Shape	Recall	IoU
Deterministic	0	0.343/0.343	1.174/1.174	0.999/0.999	0.232/0.232
	80	0.024/0.016	1.110/0.948	0.999/0.997	0.222/0.316
Conv.-trained CNN	0	0.410/0.410	0.627/0.627	0.899/0.899	0.594/0.594
	80	0.018/0.011	0.677/0.443	0.984/0.905	0.465/0.698
PC-trained CNN	0	0.408/0.408	0.630/0.630	0.897/0.897	0.590/0.590
	80	0.018/0.011	0.676/0.442	0.984/0.902	0.463/0.698

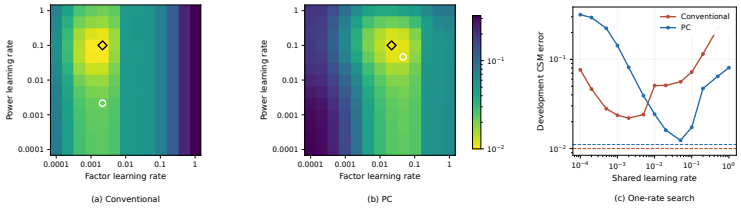


Figure 2: Development CSM error at 80 updates without an anchor over all power/factor rate pairs. (a,b) Conventional and PC on a shared logarithmic color scale. Circles mark the best shared rate and diamonds the best pair. (c) Shared-rate scores along the diagonal; dashed lines mark each method's best two-rate score.

5.3 Spatial Reconstruction after Separate Rate Selection

Without an anchor, each method searches all power/factor pairs on the same 13-rate grid from 10^{-4} to 1. Development selection minimizes the CSM objective, not a source-map metric, and gives rates $(0.1, 0.002)$ for Conventional and $(0.1, 0.02)$ for PC. Evaluation reuses the eight seeds and two array streams of the single-rate control, so this is a paired tuning control, not an independent evaluation (Appendix C.1). Figure 2 shows the development scores behind this selection. Separate rates lower the best development score by 54.5% for Conventional but by 10.3% for PC, so the single-rate policy constrained Conventional more. On the evaluation streams, the corresponding CSM reductions are 43.2–45.6% and 2.4–6.9%.

PC improves the spatial map after this added tuning freedom: shape error falls by 26.5–30.0%, FP-pix by 23.4–38.4%, and IoU rises by 11.3–14.6 percentage points across the three initializations (Table 2). For CNN starts, the largest gains occur on coherent lines and arcs, where shape error falls by 39.8–43.5% and IoU rises by 26.5–28.2 points. Synthetic blade mixtures also improve from either CNN, with 19.9–20.4% lower shape error and 13.7–14.7 points higher IoU. Incoherent point, line and arc gains from CNN starts are small or mixed. Appendix Table 6 lists all five metrics by family.

Figure 3 shows how the gain varies with initialization and source structure. With either CNN, shape and IoU improve in all eight paired seeds for coherent lines, coherent arcs and blade mixtures. The pooled shape reductions are 30.0% (95% CI: 28.8–31.2%) and 29.6% (28.3–30.9%) and the IoU gains 11.7 pp (11.0–12.3) and 11.3 pp (10.6–11.9) for the Conventional-trained and PC-trained CNNs, so the main spatial result does not depend on which parameterization trained the CNN.

The deterministic start improves shape and IoU across all six families, showing that the coordinate effect is not limited to learned starts. Learning still improves the final reconstruction: PC's pooled shape error is 0.409 from either CNN and 0.797 from the deterministic start, with IoU about 0.73 versus 0.52. The CNN supplies the stronger spatial estimate, and PC refines it more effectively under this budget.

The benefit is better spatial reconstruction, not uniformly better measurement fit: Conventional has lower pooled CSM error for all three starts and higher recall for CNN starts. Because both forms represent the same family and start from the same covariance, the difference arises along their finite-step optimization paths, not from added representational capacity. A power decomposition characterizes these paths. We write each fitted covariance as $\mathrm{diag}(d)+WW^*$, divide both components by the total trace, and sum the residual-diagonal power d_i and low-rank power $\|W_i\|_2^2$ outside the

Table 2: PC relative to Conventional with separately selected power and factor rates, no anchor and 80 updates. Positive values favor PC. Error reductions are calculated within seed before averaging. Recall and IoU differences are percentage points (pp).

Initializer	CSM (%)	Shape (%)	FP-pix (%)	Recall (pp)	IoU (pp)
Conv.-trained CNN	-4.9	+30.0	+38.4	-2.4	+11.7
PC-trained CNN	-5.3	+29.6	+37.7	-2.5	+11.3
Deterministic	-16.1	+26.5	+23.4	+0.4	+14.6

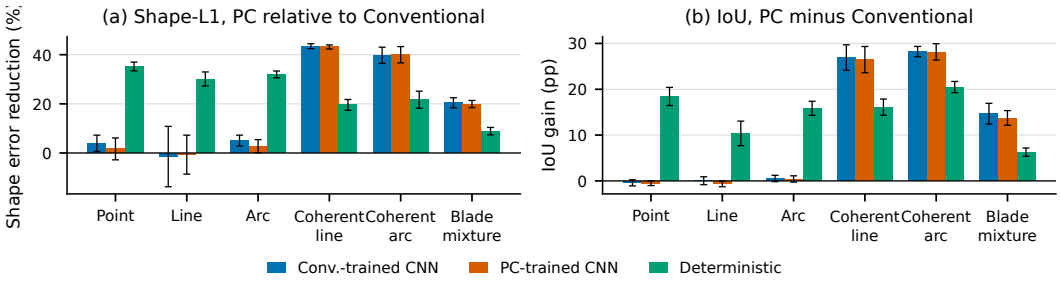


Figure 3: Spatial gains by source family with separate power and factor rates, $\lambda = 0$ and 80 updates. Bars are eight-seed means of (a) the within-seed relative shape-L1 reduction and (b) the IoU gain in percentage points, PC relative to Conventional. Error bars are 95% Student-t intervals. Positive values favor PC. Both arrays are included.

reference map's -20 dB support. Unlike FP-pix, this weights leakage by power rather than counting pixels. For coherent lines and arcs from CNN starts, outside-support power falls from 36.7–39.5% to 14.0–17.2% with PC, even though its total low-rank share is larger; across all families it falls from 16.5–16.7% to 7.3–7.4%, with both components contributing (Figure 4). The observed maps thus reflect a different spatial distribution of power, not simply suppression of the low-rank component. This is a descriptive diagnostic, not a causal explanation.

This matches the update structure in Section 4: the off-diagonal residual can move the Conventional map through the factor but reaches the PC map only through the power block, so the two refiners follow different paths within the same covariance family.

5.4 Refinement with a Map Anchor

With $\lambda = 0.05$, both methods search the same 9×9 grid for power and factor rates. At 80 updates, PC lowers the joint objective by 18.6% and 17.4% for the two CNN origins (95% intervals 13.9–23.3% and 13.0–21.9%), raises IoU by 3.6–4.1 points and reduces FP-pix by 18.7–19.7% (Table 3). Conventional has lower CSM and shape errors and higher recall, and its objective is 2.0–3.2% lower at 20 updates. The anchored trade-off differs: PC leads in joint objective and overlap.

Figure 5 illustrates the spatial effect under the anchored single-rate protocol, using the first observation of each source family from the first circular-array evaluation seed, not a best-case selection. PC reduces spreading around points, arcs and the blade mixture; some weak support is lost, and the coherent-line artifact remains. Across budgets and optimizers with the same $\mathbf{S}_0$ and $\lambda = 0.05$, PC lowers the joint objective in all eight seeds at every 20–320-update Adam and SGD budget, by 28.6% and 8.8% at 320 updates and by 24.6% (95% CI: 22.1–27.1%) at matched single-thread CPU time (Figure 6). These are anchored single-rate results; Tables 2 and 3 give the separate-rate comparisons.

At 80 updates, single-thread refinement takes 1.38–1.43 s for Conventional and 1.53–1.57 s for PC per batch of 16 (Appendix B), a modest runtime cost.

5.5 Comparison with Other Reconstruction Methods

Reference methods place the resulting maps in context. With matched observations and 2,400 training updates, retrained DAMAS-FISTA-Net (Liang et al., 2024) gives shape-L1 of 0.613–0.615 and IoU of 0.566–0.567, versus PC's 0.373–0.382 and 0.729–0.731 across analytic and finite-snapshot training; direct CNN shape-L1 is 0.432–0.445. PC takes 1.25–1.26 s per 16-observation GPU batch,

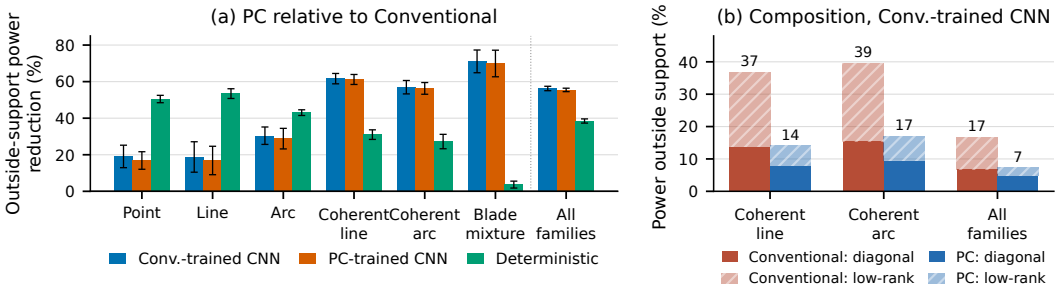


Figure 4: Power outside the reference map's -20 dB support after 80 updates with two rates and no anchor. (a) Seed-wise reduction with PC relative to Conventional by source family and start; bars are eight-seed means with 95% Student-t intervals, and positive values favor PC. (b) Composition for the Conventional-trained CNN as a percentage of total reconstructed power: solid, residual diagonal; hatched, low-rank. Appendix Table 7 lists the values.

Table 3: Effect of symmetric learning rate tuning at 80 updates, $\lambda = 0.05$. Positive values favor PC. Error/objective reductions are computed within seed before averaging. Recall and IoU differences are percentage points (pp).

Initializer	Policy	J (%)	CSM (%)	Shape (%)	FP (%)	Recall (pp)	IoU (pp)
Conv.-trained CNN	One rate	+27.6	+31.4	+16.3	+42.1	-5.8	+16.1
	Two rates	+18.6	-15.2	-2.7	+19.7	-5.5	+4.1
PC-trained CNN	One rate	+28.0	+33.1	+16.3	+42.4	-5.8	+16.2
	Two rates	+17.4	-17.8	-3.5	+18.7	-5.5	+3.6

versus 3.0–4.4 ms for the network; this compares complete pipelines, not an isolated coordinate effect (Appendix E.1). Against a strengthened LowRank-CF baseline with rank six, 80 updates and a development-selected rate, PC has 56.0–63.8% lower CSM error and 62.3–64.7% lower shape error, with higher IoU but lower recall (Appendix Table 8; Appendix D compares eight non-learning methods). The matched comparisons remain the evidence for PC itself.

5.6 Cross-Simulator Reconstruction and Non-Acoustic Fitting

Acoular 26.04 (Sarradj & Herold, 2017; Sarradj et al., 2026) replaces analytic CSMs with time-domain propagation and finite-snapshot Welch estimation (Welch, 1967) on the same circular array and grid. Eight paired seeds give 128 recordings, half with three independent points and half with a coherent pair plus a separate point. Within each CNN origin, refiners share rank six, $\mathbf{S}_0$, objective and 80 updates, without Acoular training or tuning. We also retrain on sample covariances from 256 complex Gaussian snapshots.

Across both CNN origins and training distributions, PC lowers shape-L1 by 7.8–8.6% and CSM error by 11.7–11.9% versus Conventional on new paired Acoular observations. PC has higher IoU than Conventional (0.044–0.047 versus 0.021–0.022), showing less spreading without Acoular tuning. Low overlap and recall above 0.99 indicate broad supports, and finite-snapshot training changes this little. Appendix Figure 8 shows median-error examples, and Table 12 gives all metrics.

To test whether the coordinate effect depends on acoustic steering structure, complex Gaussian and Fourier-sampling operators (32×357) replace H, keeping HSH^*, the six source families and a deterministic start. With separately developed rates, PC lowers the objective in all eight seeds by 76.2–78.3% without an anchor and 44.7–46.1% with $\lambda = 0.05$ (Appendix E.3). Absolute shape errors remain 0.89–1.49: these are fitting gains beyond acoustics, not learned-model transfer.

6 Regularized Fitting of Wind Turbine Measurements

We next assess covariance refinement on the processed pairwise spectral exports accompanying Xue & Yang (2024), from a 32-microphone ground array observing a full-scale wind turbine. The exports contain autopowers and relative phases but no independent cross-spectral magnitudes, so we use PSD-corrected proxy CSMs, not synchronous covariance estimates. There is no measured

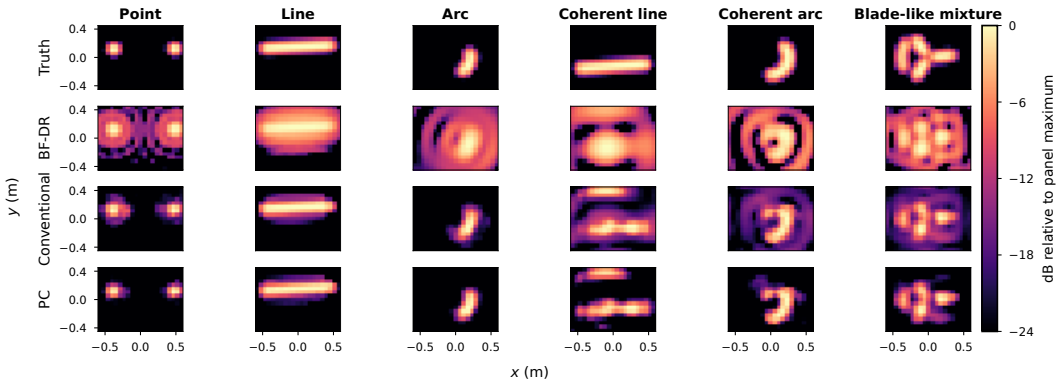


Figure 5: Matched source-map refinement across six synthetic families under the anchored single-rate protocol: Conventional-trained initial covariance, $r = 6$, 80 updates, $\lambda = 0.05$ and rates 0.002/0.05. Each panel uses its own peak over $[-24, 0]$ dB to show spatial structure rather than absolute power.

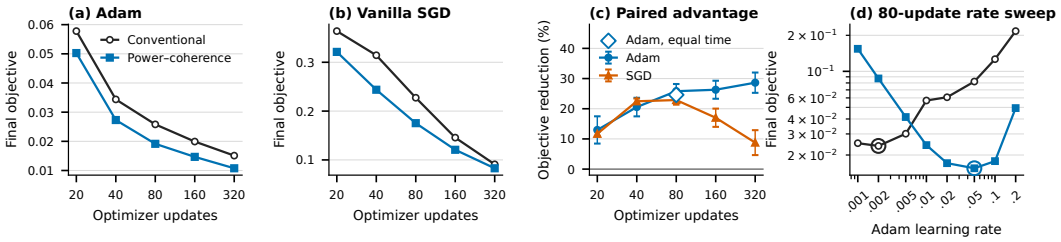


Figure 6: Anchored refinement from the same $\mathbf{S}_0$ with $r = 6$, $\lambda = 0.05$ and one rate per parameterization. (a,b) Mean joint objectives under Adam and SGD. (c) Mean seed-wise reductions favoring PC with 95% Student-t intervals over eight seeds; the diamond uses matched CPU time (mean target 2.143 s per batch of 16), plotted at 80 for reference. (d) The 80-update Adam rate search on three disjoint development seeds; rings mark selected rates.

source-position ground truth, so this experiment assesses regularized fitting, not localization accuracy (Appendix F).

Both rank-16 refiners share each initial covariance, rotor support and an objective combining off-diagonal CSM error, square-root sparsity (weight 0.03) and total variation (weight 0.02) (Rudin et al., 1992). We compare deterministic and measurement-adapted starts under two PSD corrections and five initial coherence levels that preserve each power map. For the regularized objective, 59 frequency rows select among 14 rates (0.0005–5), with evaluation on 16 disjoint rows.

PC lowers the regularized objective by 52.1–53.2% at 80 updates across the five coherence settings, with gains also at 20 and 320 updates (Appendix Figure 9). The main reduction is in the sparsity term. At incoherent fraction $\epsilon = 0.01$, its weighted value falls from 0.455 to 0.167, while CSM error rises from 0.055 to 0.077. CSM-only fitting favors Conventional at 20 and 80 updates. This supports PC for spatially regularized fitting, without establishing more accurate turbine locations (Appendix F.1). Injection diagnostics use amplitude-aligned map differences at three source levels and a 4.24 m success threshold (Appendix F.2).

7 Conclusion

PC separates power and coherence within the same normalized covariance family. With separately selected power and factor rates, PC improves shape and overlap without an anchor, most on coherent lines, arcs and blade mixtures from CNN starts. With an anchor, it lowers joint objectives and improves overlap at 80 updates. Spatial gains also appear on an independent simulator without retuning, while Gaussian and Fourier experiments show objective reductions. These findings support optimization coordinates as a design choice for refining learned estimates in covariance inverse problems, where spatial accuracy and measurement fit need not improve together.

REPRODUCIBILITY STATEMENT

The Power Coherence repository provides the observation generator, CNN training and refinement code, frozen configurations and seeds, trained synthetic model states, saved evaluation records, recorded stage timings, and the scripts and numerical inputs that regenerate the experimental figures and tables. It also contains the learned-baseline and Acoular workflows, the proxy-CSM construction and adaptation code, and the derived objective records used in the measured tables and figures. The wind turbine pair-spectrum files and adapted model states are not included; the pair files are obtained from the repository cited in Appendix F. Appendices A–B give derivations, architecture and experimental settings.

AI USE STATEMENT

The authors led the research, including conceptualization, literature review, methodology, experimental design, initial mathematical derivations, and manuscript preparation. Under their direction, generative AI tools assisted with implementing the method, providing feedback on experimental design, and performing supplementary checks of mathematical derivations. These tools also supported literature retrieval, language refinement throughout the manuscript, and development of plotting scripts and intermediate formatting for all figures, under the authors' supervision. They assisted the authors with preparing materials for the anonymous research artifact. The authors reviewed all AI-assisted work and take responsibility for the final content of this work and its accompanying artifacts.

REFERENCES

John Barnard, Robert McCulloch, and Xiao-Li Meng. Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 10(4): 1281–1311, 2000.

Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009. doi: 10.1137/080716542.

Rüdiger Borsdorf, Nicholas J. Higham, and Marcos Raydan. Computing a nearest correlation matrix with factor structure. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2603–2622, 2010. doi: 10.1137/090776718.

Thomas F. Brooks and William M. Humphreys. A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays. *Journal of Sound and Vibration*, 294(4–5):856–879, 2006. doi: 10.1016/j.jsv.2005.12.046.

Emmanuel J. Candès, Michael B. Wakin, and Stephen P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *Journal of Fourier Analysis and Applications*, 14(5–6):877–905, 2008. doi: 10.1007/s00041-008-9045-x.

Zhenghan Fang, Sam Buchanan, and Jeremias Sulam. What's in a prior? learned proximal networks for inverse problems. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/f4757db82a02eea015670ecca605d5cc-Abstract-Conference.html.

Thomas Frerix, Dmitrii Kochkov, Jamie Smith, Daniel Cremers, Michael Brenner, and Stephan Hoyer. Variational data assimilation with a learned inverse observation operator. In *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pp. 3449–3458, 2021. URL https://proceedings.mlr.press/v139/frerix21a.html.

Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In *Proceedings of the 27th International Conference on Machine Learning*, pp. 399–406, 2010.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs), 2016. URL https://arxiv.org/abs/1606.08415. arXiv:1606.08415.

Nicholas J. Higham. Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and its Applications*, 103:103–118, 1988. doi: 10.1016/0024-3795(88)90223-6.

Nicholas J. Higham. Computing the nearest correlation matrix—a problem from finance. *IMA Journal of Numerical Analysis*, 22(3):329–343, 2002. doi: 10.1093/imanum/22.3.329.

Haobo Jia, Feiran Yang, Xiaoqing Hu, and Jun Yang. A dual-encoder U-net architecture with prior knowledge embedding for acoustic source mapping. *Journal of the Acoustical Society of America*, 158(3):1767–1782, 2025. doi: 10.1121/10.0039104.

Jihui Jin, Etienne Ollivier, Richard Touret, Matthew McKinley, Karim Sabra, and Justin Romberg. PETAL: Physics emulation through averaged linearizations for solving inverse problems. In *Advances in Neural Information Processing Systems*, volume 36, 2023. doi: 10.52202/075280-2129. URL https://papers.neurips.cc/paper_files/paper/2023/hash/991c9324ca71aa85ab4dd11146b35fc3-Abstract-Conference.html.

Aayush Karan, Kulin Shah, Sitan Chen, and Yonina C. Eldar. Unrolled denoising networks provably learn to perform optimal Bayesian inference. In *Advances in Neural Information Processing Systems*, volume 37, 2024. doi: 10.52202/079017-4299. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/f40b522ef32bbc050d37a2a1ab7ffc2b-Abstract-Conference.html.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations*, 2015. URL https://arxiv.org/abs/1412.6980.

Adam Kujawski and Ennes Sarradj. Fast grid-free strength mapping of multiple sound sources from microphone array data using a Transformer architecture. *Journal of the Acoustical Society of America*, 152(5):2543–2556, 2022. doi: 10.1121/10.0015005.

Adam Kujawski, Gert Herold, and Ennes Sarradj. A deep learning method for grid-free localization and quantification of sound sources. *Journal of the Acoustical Society of America*, 146(3):EL225–EL231, 2019. doi: 10.1121/1.5126020.

Hao Liang, Guanxing Zhou, Xiaotong Tu, Andreas Jakobsson, Xinghao Ding, and Yue Huang. Learning an interpretable end-to-end network for real-time acoustic beamforming. *Journal of Sound and Vibration*, 591:118620, 2024. doi: 10.1016/j.jsv.2024.118620.

Tianci Liu, Tong Yang, Quan Zhang, and Qi Lei. Optimization for amortized inverse problems. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 22289–22319. PMLR, 2023. URL https://proceedings.mlr.press/v202/liu23au.html.

Thiago Lobato, Roland Sottek, and Michael Vorländer. Deconvolution with neural grid compression: A method to accurately and quickly process beamforming results. *Journal of the Acoustical Society of America*, 153(4):2073–2089, 2023. doi: 10.1121/10.0017792.

Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. In *7th International Conference on Learning Representations*, 2019. URL https://openreview.net/forum?id=Bkg6RiCqY7.

Wei Ma and Xun Liu. Phased microphone array for sound source localization with deep learning. *Aerospace Systems*, 2(2):71–81, 2019. doi: 10.1007/s42401-019-00026-w.

Roberto Merino-Martínez, Pieter Sijtsma, Mirjam Snellen, Thomas Ahlefeldt, Jérôme Antoni, Christopher J. Bahr, Daniel Blacodon, Daniel Ernst, Arthur Finez, Stefan Funke, et al. A review of acoustic imaging methods using phased microphone arrays. *CEAS Aeronautical Journal*, 10:197–230, 2019. doi: 10.1007/s13272-019-00383-4.

Roberto Merino-Martínez, Salil Luesutthiviboon, Riccardo Zamponi, Alejandro Rubio Carpio, Daniele Ragni, Pieter Sijtsma, Mirjam Snellen, and Christophe Schram. Assessment of the accuracy of microphone array methods for aeroacoustic measurements. *Journal of Sound and Vibration*, 470, 2020. doi: 10.1016/j.jsv.2020.115176.

Roberto Molinaro, Yunan Yang, Björn Engquist, and Siddhartha Mishra. Neural inverse operators for solving PDE inverse problems. In *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25105–25139. PMLR, 2023. URL https://proceedings.mlr.press/v202/molinaro23a.html.

Vishal Monga, Yuelong Li, and Yonina C. Eldar. Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44, 2021. doi: 10.1109/MSP.2020.3016905.

Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1–4):259–268, 1992. doi: 10.1016/0167-2789(92)90242-F.

Rajiv Sambharya, Georgina Hall, Brandon Amos, and Bartolomeo Stellato. Learning to warm-start fixed-point optimization algorithms. *Journal of Machine Learning Research*, 25(166):1–46, 2024.

Ennes Sarradj and Gert Herold. A python framework for microphone array data processing. *Applied Acoustics*, 116:50–58, 2017. doi: 10.1016/j.apacoust.2016.09.015.

Ennes Sarradj, Gert Herold, Adam Kujawski, Simon Jekosch, Art J. R. Pelling, Mikolaj Czuchaj, Tom Gensch, Sebastian Oertwig, Jakob Tschavoll, Jan Ostendorf, and Timo Schumacher. Acoular: Acoustic testing and source mapping software. Zenodo, version 26.04, 2026. URL https://doi.org/10.5281/zenodo.19892324.

James Saunderson, Venkat Chandrasekaran, Pablo A. Parrilo, and Alan S. Willsky. Diagonal and low-rank matrix decompositions, correlation matrices, and ellipsoid fitting. *SIAM Journal on Matrix Analysis and Applications*, 33(4):1395–1416, 2012. doi: 10.1137/120872516.

Johannes Schwab, Stephan Antholzer, and Markus Haltmeier. Deep null space learning for inverse problems: Convergence analysis and rates. *Inverse Problems*, 35(2), 2019. doi: 10.1088/1361-6420/aaf14a.

Pieter Sijtsma. CLEAN based on spatial source coherence. *International Journal of Aeroacoustics*, 6 (4):357–374, 2007. doi: 10.1260/147547207783359459.

Bowen Song, Soo Min Kwon, Zecheng Zhang, Xinyu Hu, Qing Qu, and Liyue Shen. Solving inverse problems with latent diffusion models via hard data consistency. In *International Conference on Learning Representations*, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/1f09e1ee5035a4c3fe38a5681cae5815-Abstract-Conference.html.

Takao Suzuki. l_1 generalized inverse beam-forming algorithm resolving coherent/incoherent, distributed and multipole sources. *Journal of Sound and Vibration*, 330(24):5835–5851, 2011. doi: 10.1016/j.jsv.2011.05.021.

Julián Tachella, Dongdong Chen, and Mike Davies. Unsupervised learning from incomplete measurements for inverse problems. In *Advances in Neural Information Processing Systems*, pp. 4983–4995, 2022. doi: 10.52202/068431-0360.

Singanallur V. Venkatakrishnan, Charles A. Bouman, and Brendt Wohlberg. Plug-and-play priors for model based reconstruction. In *2013 IEEE Global Conference on Signal and Information Processing*, pp. 945–948, 2013. doi: 10.1109/GlobalSIP.2013.6737048.

Deliang Wei, Peng Chen, and Fang Li. Learning pseudo-contractive denoisers for inverse problems. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 52500–52524. PMLR, 2024. URL https://proceedings.mlr.press/v235/wei24b.html.

Peter D. Welch. The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. *IEEE Transactions on Audio and Electroacoustics*, AU-15(2):70–73, 1967. doi: 10.1109/TAU.1967.1161901.

Peng Xiao, Bin Liao, and Nikos Deligiannis. DeepFPC: A deep unfolded network for sparse signal recovery from 1-bit measurements with application to DOA estimation. *Signal Processing*, 176, 2020. doi: 10.1016/j.sigpro.2020.107699.

Weicheng Xue. LabVIEW frequency-domain pairwise spectra. GitHub repository, 2026. Commit d65bf632.

Weicheng Xue and Bing Yang. Aeroacoustic source localization using the microphone array method with application to wind turbine noise. *International Journal of Aeroacoustics*, 23(5–6):533–557, 2024. doi: 10.1177/1475472X241259122.

Tarik Yardibi, Jian Li, Petre Stoica, and Louis N. III Cattafesta. Sparsity constrained deconvolution approaches for acoustic source mapping. *Journal of the Acoustical Society of America*, 123(5): 2631–2642, 2008. doi: 10.1121/1.2896754.

Tarik Yardibi, Jian Li, Petre Stoica, Nikolas S. Zawodny, and Louis N. III Cattafesta. A covariance fitting approach for correlated acoustic source mapping. *Journal of the Acoustical Society of America*, 127(5):2920–2931, 2010. doi: 10.1121/1.3365260.

Hongkai Zheng, Wenda Chu, Bingliang Zhang, Zihui Wu, Austin Wang, Berthy Feng, Caifeng Zou, Yu Sun, Nikola Kovachki, Zachary Ross, Katherine Bouman, and Yisong Yue. InverseBench: Benchmarking plug-and-play diffusion priors for inverse problems in physical sciences. In *International Conference on Learning Representations*, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/hash/e297068f5ec7f483a017948715b591fb-Abstract-Conference.html.

A Covariance Derivatives and Map Anchor Gradients

Correlation and row contraction. For positive p_i, p_j, the source coherence coefficient is

$$\frac{S_{ij}}{\sqrt{S_{ii}S_{jj}}} = V_i V_j^*, \qquad |V_i V_j^*| \le \|V_i\|_2 \|V_j\|_2 < 1 \quad (i \neq j). \tag{8}$$

Inside the unit row ball, the inverse is $Z_i = V_i/\sqrt{1-\|V_i\|_2^2}$. For one complex row, set $a = (1+\|Z\|_2^2)^{-1/2}$. Treating complex variables through their real and imaginary parts gives

$$\mathrm{d}V = a\,\mathrm{d}Z - a^3 Z\,\mathrm{Re}\langle Z, \mathrm{d}Z\rangle. \tag{9}$$

This gives

$$\mathrm{d}R = \mathrm{diag}(-2\,\mathrm{Re}\langle V_i, \mathrm{d}V_i\rangle) + \mathrm{d}V\,V^* + V\,\mathrm{d}V^*, \qquad \mathrm{diag}(\mathrm{d}R) = 0. \tag{10}$$

This cancellation preserves the diagonal under coherence updates.

Derivative of the anchor on the normalized diagonal. Let $s = \mathrm{diag}(S)$, $t = \mathbf{1}^\top s$, $a = s/t$, and let b be the fixed anchor map. Where the numerical floor is inactive, define $g_i = 0.60\,\mathrm{sign}(a_i - b_i) - b_i/a_i$. Differentiating the anchor in Eq. 7 through normalization yields

$$u_j := \frac{\partial \mathcal{D}}{\partial s_j} = \frac{g_j - \sum_i a_i g_i}{t}. \tag{11}$$

For the conventional parameterization $s_j = q_j + \|L_j\|_2^2$, its real Euclidean factor gradients are

$$\nabla_{\mathrm{Re}\,L_j}\mathcal{D} = 2u_j\,\mathrm{Re}\,L_j, \qquad \nabla_{\mathrm{Im}\,L_j}\mathcal{D} = 2u_j\,\mathrm{Im}\,L_j. \tag{12}$$

At fixed q, a perturbation $\dot{L}$ gives

$$\dot{s}_i = 2\,\mathrm{Re}\langle L_i, \dot{L}_i\rangle, \qquad \dot{\pi} = \frac{1}{t}(I - \pi\mathbf{1}^\top)\dot{s}. \tag{13}$$

In PC $s = p$ and $\mathbf{1}^\top p = 1$, so $D_Z\pi = 0$ and $\partial\mathcal{D}/\partial Z = 0$ at fixed p. When corresponding map entries are equal, we choose the zero subgradient of the ℓ_1 term. The numerical floor in Appendix B does not change this independence. This proves the separation of anchor and coherence in Proposition 1.

Best-fit observation scale. For $A = HSH^*$ with nonzero Frobenius norm,

$$\beta^* = \max\left(0, \frac{\mathrm{Re}\langle C_{\mathrm{obs}}, A\rangle_F}{\|A\|_F^2}\right). \tag{14}$$

This is the nonnegative minimizer of $\|C_{\mathrm{obs}} - \beta A\|_F^2$. Off-diagonal fitting first removes both diagonals. The scale is recomputed and differentiated through the covariance parameters. The squared predicted norm and observed norm are floored at 10^{-20}.

Boundary cases and factor nonuniqueness. For the reverse transformation after trace normalization, write $\mathbf{S}_{\mathrm{PC}} = \mathrm{diag}(d) + WW^*$, where $d_i = p_i(1-\|V_i\|_2^2)$ and $W = D_p^{1/2}V$. Let $c = \sum_i d_i > 0$, $q = d/c$ and $L = W/\sqrt{c}$. Then $\mathrm{diag}(q) + LL^* = \mathbf{S}_{\mathrm{PC}}/c$, whose trace normalization is $\mathbf{S}_{\mathrm{PC}}$. For a fixed factor representative, these transformations are inverse on the positive residual-diagonal interior. A zero residual diagonal can require unbounded $Z_i = L_i/\sqrt{q_i}$. Such states belong to the common covariance closure, not necessarily the finite parameter domain. The right-unitary freedom $WW^* = (WQ)(WQ)^*$ also remains. Matched initializations have the same physical covariance, normalized diagonal and microphone projection even when their factor entries differ.

The nonuniqueness of a covariance inferred from microphone measurements follows directly from the steering matrix. Let $\mathcal{A}(\mathbf{S}) = \mathbf{H}\mathbf{S}\mathbf{H}^*$. In our benchmark $N = 357 > M = 32$, so $\ker(\mathbf{H})$ is nontrivial. For any $\mathbf{v} \in \ker(\mathbf{H})$, $\tau \ge 0$, and $\mathbf{S} \succeq 0$,

$$\mathcal{A}(\mathbf{S} + \tau\mathbf{v}\mathbf{v}^*) = \mathcal{A}(\mathbf{S}), \qquad \mathrm{diag}(\mathbf{S} + \tau\mathbf{v}\mathbf{v}^*) = \mathrm{diag}(\mathbf{S}) + \tau|\mathbf{v}|^2. \tag{15}$$

CSM consistency alone does not determine the source power map.

Update geometry of the two parameterizations. Let $\theta \in \mathbb{R}^n$ collect the power logits u and the real and imaginary parts of the factor F (L or Z). Define $\bar{\mathbf{S}} = \mathbf{S}/\operatorname{tr}(\mathbf{S})$ and let $s = \phi(\theta) = (d, o) \in \mathbb{R}^m$ collect its real coordinates: $d = \operatorname{diag}(\bar{\mathbf{S}})$ and o contains the real and imaginary parts of its strict upper triangle. The compared refinement objectives depend on θ through s, $\mathcal{L}(\theta) = \ell(\phi(\theta))$. At differentiable points, with $A_\theta = D\phi(\theta) \in \mathbb{R}^{m\times n}$, the chain rule gives $\nabla_\theta \mathcal{L} = A_\theta^\top \nabla_s \ell$. A gradient step gives

$$s_{t+1} = \phi(\theta_{t+1}) = s_t - \eta\, G_\theta \nabla_s \ell + O(\eta^2), \qquad G_\theta = A_\theta A_\theta^\top \succeq 0,$$

where ϕ is smooth on the finite positive interior. G_θ is the induced positive-semidefinite preconditioning operator; it can be singular in covariance space. At a matched normalized covariance, the two parameterizations share s_t and $\nabla_s \ell$ but generally have different G_θ, so their steps can differ in direction and scale. At an ℓ_1 kink, the same expansion applies to the chosen subgradient direction.

Order $s = (d, o)$ and $\theta = (u, F)$. Then

$$A_\theta = \begin{pmatrix} \partial d/\partial u & \partial d/\partial F \\ \partial o/\partial u & \partial o/\partial F \end{pmatrix}.$$

In PC, $d = p = \operatorname{softmax}(u)$ does not depend on Z (Section 4), so $\partial d/\partial Z = 0$ and the diagonal moves only through u. The diagonal residual $\nabla_d \ell$ acts through $(\partial d/\partial u)(\partial d/\partial u)^\top$ and the off-diagonal residual $\nabla_o \ell$ reaches the map only through the cross term $(\partial d/\partial u)(\partial o/\partial u)^\top$. In Conventional, $q = \operatorname{softmax}(u)$, $t = 1 + \|L\|_F^2$, and $d_i = (q_i + \|L_i\|_2^2)/t$. At fixed q, for $x_j = \operatorname{Re} L_j$ or $x_j = \operatorname{Im} L_j$,

$$\frac{\partial d_i}{\partial x_j} = \frac{2}{t}(\delta_{ij} - d_i)x_j,$$

where δ_{ij} is the Kronecker delta. Both residuals can therefore also move the map through $(\partial d/\partial F)(\partial d/\partial F)^\top$ and $(\partial d/\partial F)(\partial o/\partial F)^\top$. For the anchor, which depends only on d, the Conventional factor gradient is row-wise radial, as derived above in Proposition 1. A general off-diagonal objective need not give a radial factor gradient.

The coordinate map includes $Z_i = L_i/\sqrt{q_i}$ and a change in power variables, so equal parameter steps need not produce equal covariance steps. This agrees with the different development-selected rates in Figure 2, but does not order the paths. Adam uses a history-dependent direction h_t, giving $s_{t+1} = s_t - \eta A_\theta h_t + O(\eta^2)$ when $\theta_{t+1} = \theta_t - \eta h_t$. Our complex variant treats each complex entry as two real coordinates sharing one second moment. Coordinate choice therefore still affects the update. This explains the local update structure and PC's fixed-power diagonal invariance; spatial gains are established empirically in Section 5.3.

B Implementation and Statistical Details

Synthetic acoustic scene and data generation. Scan spacings are $\Delta x = 0.060\,\mathrm{m}$ and $\Delta y = 0.05625\,\mathrm{m}$, and each of the four Archimedean spiral arms sweeps 0.75 turns from $0.22R$ to R.

Training samples the six source families with equal probability. Point-center powers are uniform on $[0.5, 1]$. Gaussian rasterization uses $0.8\min(\Delta x, \Delta y) = 0.045\,\mathrm{m}$. Here *off-grid* means placing continuous supports before rasterization, not generating measurements with a separate continuous-source propagation model. The released generator specifies the other distributions.

Three development and eight paired evaluation seeds are fixed in the reproduction configurations.

Operator and observable inputs. If r_{mn} is the distance between a microphone and a scan point and $\bar{r}_n = M^{-1}\sum_m r_{mn}$, the internal steering response is

$$\widetilde{H}_{mn}(f) = \frac{\bar{r}_n}{r_{mn}} \exp\!\left(-\mathrm{i}\frac{2\pi f}{c} r_{mn}\right), \qquad H_{:n} = \widetilde{H}_{:n}/\|\widetilde{H}_{:n}\|_2. \tag{16}$$

The CNN receives independently sum-normalized BF and BF-DR maps plus a spatially constant frequency channel. Geometry enters the features through beamforming and enters the refiner explicitly through H. The CNN receives neither an operator tensor nor a source family label.

Network specification. The map trunk applies four stride-one, same-padded 3×3 convolutions of widths $3 \to 40 \to 40 \to 40 \to 1$ with GELU (Hendrycks & Gimpel, 2016) after the first three, then a $\mathrm{softplus}_{\beta=6}$ output and sum normalization to produce p. The factor head applies three 3×3 convolutions of widths $1 \to 32 \to 32 \to 2r$ to p, with GELU after the first two. At $r = 6$ its twelve output channels carry the real and imaginary factor components. The two stages hold 30,361 and 13,036 parameters, 43,397 in total, with no pooling, normalization or dropout. Training uses AdamW at learning rate 1.5×10^{-3}, weight decay 10^{-5} and global gradient-norm clipping at 1.0, over 2,400 batches of size eight, without microphone dropout or array-coordinate jitter. For each seed, both models train on the same online sequence of 19,200 examples.

Numerical objective and complex Adam. The supervised cross-entropy uses $\log \max(\tilde{p}_i, 10^{-8})$, as in the explicit formula preceding Eq. 6. For the anchor, normalized arguments are floored as $\bar{a}_i = \max(a_i, 10^{-8})$ and $\bar{b}_i = \max(b_i, 10^{-8})$ in both the ℓ_1 and KL expressions, without renormalization after flooring. Below the floor, the derivative is zero. This statement excludes the threshold itself.

During matched refinement, power logits start at $\log p$ and are mapped through softmax. In the complex-valued Adam variant used here, the first and second moments for a complex factor gradient g_t are

$$m_t = 0.9 m_{t-1} + 0.1 g_t, \qquad v_t = 0.999 v_{t-1} + 0.001 |g_t|^2, \tag{17}$$

followed by bias correction and the update $z_{t+1} = z_t - \eta \widehat{m}_t / (\sqrt{\widehat{v}_t} + 10^{-8})$. Each complex entry's real and imaginary components share a squared-magnitude second moment. Applying separate real/imaginary second moments is a different optimizer. Real power logit gradients use their ordinary square. The matched rank-six objective contains no factor-support penalty, sparsity, TV or gradient clipping. Optimization is performed in the selected parameterization.

Paired evaluation and statistical analysis. Each evaluation seed supplies 192 distinct CSMs: 16 observations per family and geometry, or 1,536 overall. Reuse across initializer origins and parameterizations does not increase the independent sample count. For each seed, we average each method's objective equally across these conditions, then compute $(J_{\mathrm{L}} - J_{\mathrm{PC}})/J_{\mathrm{L}}$. Student-$t$ intervals use the eight seed-level values. The two-sided sign-flip test enumerates all 2^8 assignments, with minimum $p = 2/2^8 = 0.0078125$. Ratios of mean objectives for individual conditions are labeled separately.

Budget and timing controls. Figure 6 fixes the CNN, initial covariance and anchor across budgets. Its Adam development grid is $\{0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2\}$. SGD independently selects 0.01 (Conventional) and 3.0 (PC), both interior to a common 13-rate grid from 10^{-4} to 100. For each paired batch of 16 observations, the equal-time target is the geometric mean of the two 80-update runtimes. We select the nearest recorded endpoints from trajectories bracketing the target, within a 2.5% tolerance. These average 81.7 and 78.4 updates, not an identical update count.

Independent serial CPU timing. On a single-thread Intel Xeon W-2245, each configuration uses one warm-up and three retained repetitions across both arrays and six families, yielding 36 batches of 16 observations. Recomputed maps match the reference endpoints. CNN inference takes about 6.9 ms, deterministic initialization 394 ms, and covariance conversion 171–177 ms. PC's 80-update refinement is 7.4–10.6% slower than Conventional, including diagnostics but excluding observation generation, beamforming features and training. Full timings for each stage remain in the artifact. LowRank-CF processes one observation at a time, so its batch setting differs from that of the learned refiners.

Shape-L1 is $\|p - p^\star\|_1$ for unit-sum maps. Support metrics follow the definitions based on pixel counts in Section 5.1.

C Synthetic Optimization Controls

C.1 Initialization and Symmetric Learning Rate Controls

These controls reuse the eight trained CNN pairs, but generate new evaluation observations on both arrays. Three separate development streams select the rates before evaluation. The candidate

set is $\{0.0001, 0.0005, 0.002, 0.01, 0.02, 0.05, 0.1, 0.5, 1\}$. The no-anchor study assigns half the selection weight to deterministic initialization and one quarter to each CNN origin. A single rate per parameterization is then used for all three origins: 0.002 for Conventional and 0.05 for PC. The deterministic start floors the beamformed power at 10^{-4} of its maximum and uses six observation-derived eigenvectors with 30% coherent trace share. It uses no source labels. Paired conversion preserves the physical covariance, not the numerical values of the two parameter sets.

Table 4 reports all three budgets and spatial metrics. Table 5 reports direct CNN power separately because the Conventional-trained factor adds diagonal energy. Its direct map is not exactly the zero-update covariance diagonal. For PC-trained models, the two maps agree up to numerical precision.

Table 4: Initialization control, $\lambda = 0$, with one rate per parameterization. Conventional/PC means over eight seeds, two geometries and six families. Lower CSM, shape and FP-pix and higher recall and IoU are better.

Initializer	Updates	CSM	Shape	FP-pix	Recall	IoU
Deterministic	0	0.343/0.343	1.174/1.174	0.768/0.768	0.999/0.999	0.232/0.232
	20	0.061/0.055	1.144/1.017	0.782/0.740	0.999/0.998	0.218/0.260
	80	0.024/0.016	1.110/0.948	0.778/0.684	0.999/0.997	0.222/0.316
Conv.-trained CNN	0	0.410/0.410	0.627/0.627	0.348/0.348	0.899/0.899	0.594/0.594
	20	0.050/0.042	0.659/0.510	0.491/0.289	0.973/0.900	0.497/0.650
	80	0.018/0.011	0.677/0.443	0.527/0.239	0.984/0.905	0.465/0.698
PC-trained CNN	0	0.408/0.408	0.630/0.630	0.351/0.351	0.897/0.897	0.590/0.590
	20	0.052/0.043	0.659/0.512	0.490/0.289	0.971/0.897	0.497/0.649
	80	0.018/0.011	0.676/0.442	0.529/0.237	0.984/0.902	0.463/0.698

Table 5: Direct CNN outputs before covariance conversion or refinement. The same observations and networks are reused for both control studies.

Study	CNN	Shape	FP-pix	Recall	IoU
initialization	Conv.-trained CNN	0.628	0.349	0.900	0.594
initialization	PC-trained CNN	0.630	0.351	0.897	0.590
block rates	Conv.-trained CNN	0.628	0.349	0.900	0.594
block rates	PC-trained CNN	0.630	0.351	0.897	0.590

For $\lambda = 0.05$, each parameterization evaluates all pairs of the same nine rates for the power and factor blocks. The scalar policy evaluates the nine pairs with equal rates. Selection weights both CNN origins equally. The fixed power/factor rates are $(0.1, 0.002)$ for Conventional and $(0.02, 0.05)$ for PC. Scalar rates remain 0.002 and 0.05. All selections are inside the evaluated range. The 80-update reductions in Table 3 average the relative difference within each seed, not a ratio of pooled means. The saved results also include the 20-update comparisons, which favor Conventional with two rates.

With separate power and factor rates at $\lambda = 0$, PC has 26.5–30.0% lower shape error, 23.4–38.4% lower FP-pix and 11.3–14.6 percentage points higher IoU at 80 updates across the three initializations. CSM error is 4.9%, 5.3% and 16.1% higher for Conventional-trained, PC-trained and deterministic starts. Recall is 2.4–2.5 points lower for CNN starts and 0.4 points higher for the deterministic start. This paired follow-up searches all pairs on a 13-rate 1–2–5 grid from 10^{-4} to 1, using the same no-anchor development rule. Selected power/factor rates are $(0.1, 0.002)$ for Conventional and $(0.1, 0.02)$ for PC. It reuses eight seeds and both array streams, whose scalar-rate results were already known, so it is not an independent confirmation.

Figure 2 shows the full development grid. Separate rates reduce the development CSM score by 54.5% for Conventional and 10.3% for PC relative to each method's best shared rate. On the evaluation streams, the corresponding seedwise reductions are 43.2–45.6% and 2.4–6.9%. Thus, PC gains less from this extra tuning freedom under this grid and budget. In this grid, 4, 4 and 5 Conventional pairs and 1, 3 and 5 PC pairs lie within 5%, 10% and 20% of their own best development score, so the grid does not show a wider region of good rates for PC. All 169 pairs are finite for each method.

Table 6 separates the same paired results by source family. For both CNN origins, coherent lines and arcs have 39.8–43.5% lower shape error and 26.5–28.2 percentage points higher IoU with PC. These gains combine less spreading with better overlap, even though their CSM fits are worse. The

gains are not uniform across families: CNN-start line shape changes little, and point-source IoU is slightly lower. The deterministic start has larger shape gains on the incoherent point, line and arc families. This descriptive breakdown shows where the spatial gains occur, not an isolated causal effect of coherence.

Table 6: Source-family comparison with separate power and factor rates, $\lambda = 0$ and 80 updates. Positive values favor PC. Metrics are averaged across both arrays within seed before relative error reductions and percentage-point differences are calculated. Entries average eight seeds.

Source family	CSM (%)	Shape (%)	FP (%)	Recall (pp)	IoU (pp)
Conventional-trained CNN					
Point	+24.5	+3.9	-5.4	+0.5	-0.4
Line	-0.3	-1.5	+3.1	-0.4	+0.0
Arc	+16.1	+5.0	+8.1	-0.8	+0.5
Coherent line	-38.8	+43.5	+47.4	-7.2	+26.9
Coherent arc	-22.9	+39.8	+41.7	-3.8	+28.2
Blade mixture	-11.7	+20.4	+65.0	-2.6	+14.7
All families	-4.9	+30.0	+38.4	-2.4	+11.7
PC-trained CNN					
Point	+22.4	+1.7	-7.0	+0.6	-0.5
Line	-1.2	-0.7	-2.1	-0.7	-0.6
Arc	+15.7	+2.7	+7.9	-0.8	+0.4
Coherent line	-38.2	+43.1	+46.9	-7.6	+26.5
Coherent arc	-21.9	+40.0	+41.3	-4.0	+28.1
Blade mixture	-10.3	+19.9	+63.9	-2.5	+13.7
All families	-5.3	+29.6	+37.7	-2.5	+11.3
Deterministic					
Point	-19.9	+35.2	+25.8	+1.1	+18.4
Line	-13.4	+30.1	+19.8	+1.0	+10.4
Arc	-29.8	+32.0	+26.6	-0.8	+15.8
Coherent line	-5.2	+19.6	+24.6	-0.6	+16.1
Coherent arc	+0.8	+21.7	+27.4	+1.6	+20.5
Blade mixture	-27.1	+8.9	+12.1	+0.0	+6.3
All families	-16.1	+26.5	+23.4	+0.4	+14.6

PC does not obtain these maps by simply suppressing low-rank power. A diagnostic replay exports the residual-diagonal and low-rank contributions to the unit-trace power map for every paired condition. All saved maps and metrics at 0, 20 and 80 updates are reproduced exactly. We measure power outside the reference map's -20 dB support, not a fraction of pixels. Table 7 gives the mean power fractions at 80 updates. For coherent lines and arcs from either CNN, PC's low-rank share is 49.6–50.4%, versus 44.1–46.0% for Conventional. Its low-rank power outside the support is only 6.1–7.6% of total power, versus 23.0–24.1%. Total power outside the support falls from 36.7–39.5% to 14.0–17.2%. The residual diagonal also contributes to the difference. This decomposition describes the observed optimization paths, not a causal intervention. PC separates power from factor updates but does not exclude diffuse maps or change the shared covariance family.

Table 7: Power distribution after 80 updates with two rates and no anchor. Entries are Conventional / PC shares of total reconstructed power (%), averaged over observations, both arrays and eight seeds. The last three columns measure power outside the reference support. Figure 4 plots the seed-wise reductions for all families and the composition for the Conventional-trained CNN.

Source family	Low-rank share	Outside: diagonal	Outside: low-rank	Outside: total
Conventional-trained CNN				
All families	20.6 / 22.8	6.8 / 4.8	9.9 / 2.5	16.7 / 7.3
Coherent line	46.0 / 49.6	13.7 / 7.9	23.0 / 6.1	36.7 / 14.0
Coherent arc	44.4 / 50.3	15.4 / 9.4	24.1 / 7.6	39.5 / 17.0
PC-trained CNN				
All families	20.3 / 22.8	6.8 / 4.8	9.7 / 2.6	16.5 / 7.4
Coherent line	46.0 / 49.6	13.7 / 8.0	23.1 / 6.2	36.8 / 14.3
Coherent arc	44.1 / 50.4	15.4 / 9.5	23.8 / 7.6	39.2 / 17.2
Deterministic				
All families	43.2 / 45.4	17.9 / 13.8	17.8 / 8.1	35.6 / 21.9
Coherent line	48.4 / 52.9	18.6 / 14.6	17.9 / 10.5	36.4 / 25.1
Coherent arc	46.9 / 54.1	17.5 / 16.1	19.4 / 10.8	36.9 / 26.9

The stronger LowRank-CF baseline keeps its pure-LL^* fitting implementation, random complex start and real/imaginary Adam convention. Only rank (six), budget (80) and the development-selected

rate (0.0005) change. It has no diagonal term, CNN or anchor. The learned refiners give better measurement fit and spatial overlap even with this larger baseline budget (Table 8). LowRank-CF has recall of one but high FP-pix, so coverage alone does not show accuracy.

Table 8: Strengthened LowRank-CF and learned refiners on the same new observations. All use rank six and 80 updates. Refiner objectives include the 0.05 map anchor. Entries are mean metrics, not percentage reductions.

Method	Initializer	Policy	CSM	Shape	FP-pix	Recall	IoU
LowRank-CF	Random		0.0369	1.5004	0.7887	1.0000	0.2113
Conv.	Conv.-trained CNN	One rate	0.0194	0.6318	0.4778	0.9767	0.5119
PC	Conv.-trained CNN	One rate	0.0133	0.5291	0.2770	0.9188	0.6731
Conv.	Conv.-trained CNN	Two rates	0.0139	0.5500	0.3876	0.9670	0.5954
PC	Conv.-trained CNN	Two rates	0.0160	0.5649	0.3114	0.9121	0.6368
Conv.	PC-trained CNN	One rate	0.0197	0.6331	0.4781	0.9751	0.5115
PC	PC-trained CNN	One rate	0.0132	0.5297	0.2759	0.9167	0.6730
Conv.	PC-trained CNN	Two rates	0.0136	0.5472	0.3827	0.9651	0.5997
PC	PC-trained CNN	Two rates	0.0160	0.5662	0.3118	0.9096	0.6353

C.2 Scalar Learning Rates and Optimization Budgets

Table 9: Same-$\mathbf{S}_0$ comparison after 80 Adam updates with one rate per parameterization and $\lambda = 0.05$. Entries are Conventional (L)/PC means over eight seeds and six equally weighted source families for each array and initializer origin. Lower is better.

Array	$\mathbf{S}_0$ origin	Objective L / PC	Shape L / PC	FP-pix L / PC	CSM L / PC
Circular	Conventional-trained	.0247 / .0154	.565 / .465	.480 / .240	.0174 / .0110
Circular	PC-trained	.0248 / .0158	.574 / .471	.482 / .244	.0177 / .0111
Spiral	Conventional-trained	.0266 / .0227	.700 / .599	.474 / .314	.0202 / .0161
Spiral	PC-trained	.0270 / .0225	.691 / .589	.469 / .306	.0207 / .0158

Under this single-rate anchored policy, PC improves all four metrics across arrays and initializer origins (Table 9). Origin denotes the CNN supplying the common initial covariance. Averaging terminal objectives equally over both origins and geometries within each seed gives a 25.8% reduction (paired 95% CI: 23.3–28.3%, 8/8 seeds, exact two-sided sign-flip $p = 0.0078$). With either initializer origin, PC's relative objective reduction is larger on the training circle than on the unseen spiral.

Figure 6 compares budgets and optimizers with frozen CNN weights. PC improves the objective in all eight rank-six seeds at every 20–320-update Adam and SGD budget. At 320 updates, reductions are 28.6% and 8.8%. With matched single-thread CPU time, PC lowers the objective by 24.6% (95% CI: 22.1–27.1%). Each paired batch of 16 observations uses the geometric mean of its two 80-update runtimes as the time target. We select the nearest recorded endpoints. The target averages 2.143 s, with at most 1.10% mismatch. This includes refinement diagnostics but excludes CNN training and initialization. Panel (d) shows why the parameterizations use separately tuned rates. Selection and timing details are in Appendix B.

After retraining and separate rate tuning at ranks 1, 2, 4, 6 and 12, PC performs better at 80 updates across eight evaluation seeds. At rank 12, Conventional performs better at 20 updates when both methods use their 80-update rates (-5.1%).

Effect of the anchor weight. With $\mathbf{S}_0$ fixed across $\lambda \in \{0, 0.0125, 0.05, 0.2, 0.8\}$ and independent development tuning, PC lowers the objective at every weight: 37.2% at $\lambda = 0$ (95% CI: 34.5–40.0%) and 18.4–34.7% at nonzero weights. The benefit does not require the anchor separation of Proposition 1.

Longer optimization budgets. A separate eight-seed, rank-six same-$\mathbf{S}_0$ control favors PC by 14.0% after 4096 Adam updates with separately tuned rates. Damped L-BFGS with Armijo backtracking encounters line-search failures. Symmetric, history-free restarts from Adam favor Conventional by

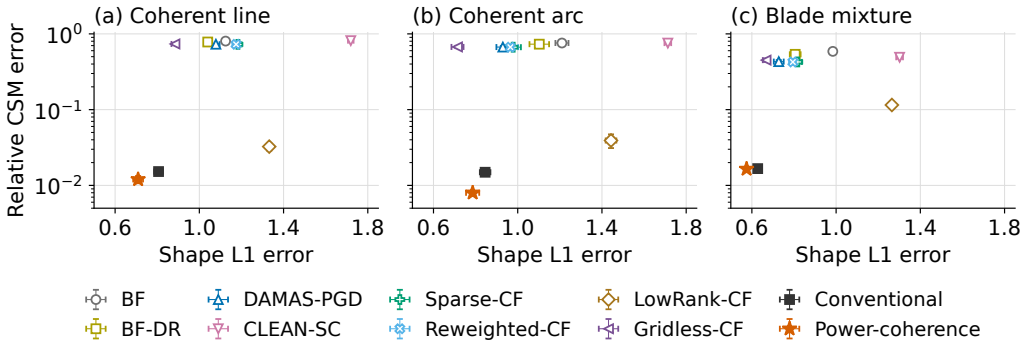


Figure 7: Complex source reconstruction: lower is better on both axes. The CSM axis is logarithmic. Eight seeds, two arrays and 16 observations per family, array and seed. Learned results average both initializer origins. Points are eight-seed means. Bars are marginal, exploratory 95% Student-t intervals, not paired significance tests.

about 6% under shared, untuned settings. These values describe performance at the chosen stopping points.

D Baseline Comparisons and Source Family Results

D.1 Baseline settings and reconstruction accuracy

Figure 7 and Table 10 compare learned refiners with eight non-learning baselines on identical observations. PC has the lowest mean errors for lines and blade mixtures. On arcs, Gridless-CF has lower shape error (0.714 versus 0.786), but PC has lower CSM error (0.008 versus 0.672). Several non-learning methods recover isolated incoherent points better. These methods differ in initialization, model and budget, so this comparison complements rather than replaces the matched parameterization controls.

Settings are fixed across observations and seeds. DAMAS-PGD applies 60 FISTA updates (Beck & Teboulle, 2009) to nonnegative least squares with point-spread matrix $P_{ij} = |h_i^* h_j|^2$, replacing the original DAMAS solver (Brooks & Humphreys, 2006). CLEAN-SC uses coherent-component subtraction (Sijtsma, 2007) for 20 iterations with loop gain 0.6.

The four covariance-fitting comparators are implementations for this study, not reproductions of MACS. Sparse-CF minimizes

$$\tfrac{1}{2}\|\mathcal{P}_{\text{off}}(H\,\text{diag}(x)H^* - C)\|_F^2 + 0.05\sum_i w_i x_i, \qquad x \geq 0,$$

where $\mathcal{P}_{\text{off}}$ removes the diagonal. It uses $w_i = 1$ and 80 FISTA updates. Reweighted-CF adapts iterative reweighted L1 minimization (Candès et al., 2008) to this loss with four warm-started solves of 40 updates. Starting with unit weights, it sets $u_i = (x_i/m + 0.02)^{-1}$ and $w_i = u_i/\,\text{mean}(u)$ between solves, where m is the median positive entry of x, floored at 10^{-12} (or 1 if none is positive). This scaling and weight normalization specify our variant, which performs no Bayesian evidence optimization.

LowRank-CF minimizes the best-scaled relative full CSM residual with $S = LL^*$, a random complex rank-4 factor and 40 Adam updates at 0.06. Gridless-CF minimizes the corresponding off-diagonal residual for six uncorrelated monopoles with continuous positions and nonnegative powers. It uses 60 Adam updates at 0.06, initialized at the six highest Sparse-CF grid entries. Sigmoid-bounded positions remain in the scan rectangle and powers use softplus. Gaussian rasterization, with width 0.85 times the minimum grid pitch, is used only for map metrics.

All eight baselines use the same 1,536 observations as the learned refiners, without evaluation-data tuning (Table 10). Evaluation uses the full CSM for every method, including those fitted to off-diagonal entries: $\min_{\alpha\geq 0}\|C - \alpha\widehat{C}\|_F/\|C\|_F$. Here $\widehat{C}$ comes from a diagonal source covariance

for map methods, LL^* for LowRank-CF, and the continuous-source model for Gridless-CF. This common metric differs from the method-specific fitting objectives.

Baseline runtimes are single-thread CPU means per observation on one host. The learned refiners average about 130 ms per observation for 80 updates, including per-update diagnostic evaluations (Appendix B). These timings describe the implementations on that host. The learned results reuse the eight-seed means from Table 9 and average over both initializer origins.

Table 10: Non-learning and learned reconstruction on identical synthetic observations. Entries are eight-seed means. Learned results average both initializer origins from Table 9. Lower is better for every metric.

Method	Model / budget	Circular			Spiral			ms/obs
		Shape-L1	FP-pix	CSM	Shape-L1	FP-pix	CSM	
BF	incoherent map	1.295	0.787	0.6604	1.059	0.751	0.5110	0
BF-DR	incoherent map, diag. removed	1.110	0.737	0.5711	0.921	0.503	0.4641	0
DAMAS-PGD	incoherent deconv., 60 FISTA	0.719	0.250	0.3269	0.542	0.152	0.3098	17
CLEAN-SC	coherent removal, 20 it.	1.407	0.462	0.5471	1.445	0.272	0.4435	14
Sparse-CF	diag. cov. fit, 80 it.	0.990	0.206	0.3840	0.717	0.043	0.3407	68
Reweighted-CF	reweighted diag. fit, 4x40	0.993	0.206	0.3603	0.706	0.052	0.3218	220
LowRank-CF	rank-4 factor, 40 Adam	1.569	0.788	0.1627	1.409	0.787	0.2223	132
Gridless-CF	6 monopoles, 60 steps	0.682	0.222	0.3799	0.528	0.143	0.3237	303
Conventional, learned init., 80 Adam	rank-6 diag.+low-rank	0.569	0.481	0.0176	0.695	0.471	0.0205	–
Power–coherence, learned init., 80 Adam	rank-6 diag.+low-rank	0.468	0.242	0.0110	0.594	0.310	0.0160	–

E Comparisons across Measurement Models

E.1 External Learned Baseline and Finite-Snapshot Training

We use the authors' DAMAS-FISTA-Net code (Liang et al., 2024), commit `bd74ab0e` of `https://github.com/HauLiang/DAMAS-FISTA-Net`, and keep its five-layer forward computation. We resize the final mapping layer from 1,681 to 357 grid points and recompute the steering and point-spread matrices. With unit-norm steering, multiplying its steering input by the microphone count gives our beamforming convention. We keep squared-map loss and Adam weight decay 0.01. This is controlled retraining, not a reproduction of the original dataset scores. The network uses the incoherent-source DAMAS model as its architectural prior and predicts a power map. We retrain it on all six source families.

All models receive matched unit-trace CSMs and unit-sum targets for 2,400 updates, batch size eight. Analytic training uses exact simulated CSMs. Finite-snapshot training uses covariances from 256 independent proper complex Gaussian microphone snapshots, holding geometry, sources, architecture and loss fixed. Three development seeds select the training rate by mean validation shape-L1 across six source families. We extend the common grid $\{0.0003, 0.001, 0.0015, 0.003\}$ with 0.0001 and 0.01 for all models, adding 36 fits at the same budget. A candidate is excluded for a model and training condition if any development run has nonfinite gradients or outputs. DAMAS-FISTA-Net still selects 0.0003 in both conditions, with higher validation error at both new rates. Both CNNs still select 0.003. At 0.01, 11 of the 12 CNN fits fail, excluding this rate in both training conditions. The rate 0.003 remains the highest evaluated CNN rate that succeeds across all development seeds, not an established optimum. The selected models and evaluation maps are unchanged. Eight evaluation seeds provide 768 circular-array observations. Refinement uses existing Conventional/PC rates 0.002/0.05, $\lambda = 0.05$, rank six, and 80 updates, with a common initial covariance within each CNN origin.

In Table 11, PC gives lower shape error and higher overlap than direct CNN maps and Conventional. Conventional has higher recall but more false-positive pixels. DAMAS-FISTA-Net predicts maps quickly but is less accurate here. Direct-output CSM scores use diagonal source covariances. Refined outputs also include fitted coherence.

Table 11: Controlled learned-method comparison on circular-array observations. Entries average observation metrics within seed, then across eight seeds. Shape means shape-L1. Direct-output CSM scores use diagonal source covariances.

Training	Model / initializer	Output	Shape ↓	FP-pix ↓	Recall ↑	IoU ↑	CSM ↓
Analytic	DAMAS-FISTA-Net	Direct	0.613	0.276	0.755	0.567	0.445
Analytic	Conv.-trained CNN	Direct	0.432	0.288	0.902	0.658	0.395
Analytic	Conv.-trained CNN	Conv., 80	0.478	0.441	0.968	0.542	0.016
Analytic	Conv.-trained CNN	PC, 80	0.373	0.212	0.909	0.731	0.010
Analytic	PC-trained CNN	Direct	0.445	0.292	0.909	0.656	0.400
Analytic	PC-trained CNN	Conv., 80	0.490	0.445	0.972	0.541	0.016
Analytic	PC-trained CNN	PC, 80	0.382	0.216	0.917	0.731	0.010
Finite snapshots	DAMAS-FISTA-Net	Direct	0.615	0.278	0.757	0.566	0.446
Finite snapshots	Conv.-trained CNN	Direct	0.443	0.293	0.904	0.652	0.399
Finite snapshots	Conv.-trained CNN	Conv., 80	0.487	0.448	0.972	0.538	0.016
Finite snapshots	Conv.-trained CNN	PC, 80	0.379	0.215	0.911	0.729	0.010
Finite snapshots	PC-trained CNN	Direct	0.438	0.288	0.904	0.656	0.397
Finite snapshots	PC-trained CNN	Conv., 80	0.482	0.440	0.971	0.544	0.016
Finite snapshots	PC-trained CNN	PC, 80	0.376	0.216	0.912	0.729	0.010

Sequential RTX PRO 6000 Blackwell timing uses one warm-up and three repetitions for each of eight 16-observation batches. DAMAS-FISTA-Net takes 2.97–4.43 ms, direct CNN inference 3.41–4.06 ms, Conventional refinement 1.21–1.24 s and PC refinement 1.25–1.26 s. Incremental peak GPU allocation is 8.54, 5.99, 6.66 and 7.29 MiB, respectively. Timings include feature construction, inference, covariance initialization, refinement and final CSM assembly. They exclude training, data generation, host transfer, geometry-only precomputation and truth-based scoring.

E.2 Cross-Simulator Evaluation with Acoular

Acoular 26.04 (Sarradj & Herold, 2017; Sarradj et al., 2026) generates the observations by time-domain propagation with fractional delays, spherical spreading, microphone noise, and finite-snapshot Welch CSM estimation. Each observation is 10 s at 12.8 kHz. The Welch block length is 1024 and the uncorrelated microphone noise fraction is 0.05. Inputs are the CSM, array geometry, and frequency. The matched parameterization comparison uses eight observations per source class over eight paired seeds.

Known single-source checks of microphone/grid ordering, FFT bins, conjugation and beamforming peak locations all pass. The original eight-seed comparison gives objective reductions of 20.3% for independent points and 25.5% for the coherent-pair class. Recomputed metrics give initial IoU of 0.036–0.041, Conventional IoU of 0.019–0.024 and PC IoU of 0.039–0.049. PC improves the relative objective, but absolute spatial overlap remains low.

Table 12: Cross-simulator comparison on fresh Acoular observations. Entries are eight-seed means, using the same metrics and direct-output CSM convention as Table 11.

Training	Model / initializer	Output	Shape ↓	FP-pix ↓	Recall ↑	IoU ↑	CSM ↓
Analytic	DAMAS-FISTA-Net	Direct	1.735	0.956	0.969	0.044	0.538
Analytic	Conv.-trained CNN	Direct	1.696	0.960	0.987	0.040	0.527
Analytic	Conv.-trained CNN	Conv., 80	1.746	0.979	1.000	0.021	0.032
Analytic	Conv.-trained CNN	PC, 80	1.607	0.955	0.997	0.045	0.028
Analytic	PC-trained CNN	Direct	1.685	0.960	0.982	0.040	0.526
Analytic	PC-trained CNN	Conv., 80	1.740	0.978	0.997	0.022	0.032
Analytic	PC-trained CNN	PC, 80	1.591	0.954	0.992	0.046	0.028
Finite snapshots	DAMAS-FISTA-Net	Direct	1.736	0.956	0.971	0.044	0.539
Finite snapshots	Conv.-trained CNN	Direct	1.694	0.961	0.987	0.039	0.526
Finite snapshots	Conv.-trained CNN	Conv., 80	1.744	0.978	0.997	0.022	0.032
Finite snapshots	Conv.-trained CNN	PC, 80	1.609	0.956	0.997	0.044	0.028
Finite snapshots	PC-trained CNN	Direct	1.684	0.959	0.982	0.041	0.525
Finite snapshots	PC-trained CNN	Conv., 80	1.739	0.978	1.000	0.022	0.032
Finite snapshots	PC-trained CNN	PC, 80	1.596	0.953	0.997	0.047	0.028

The controlled retraining comparison in Appendix E.1 uses 128 new Acoular observations, with no Acoular-specific training or tuning. PC's gains in shape and CSM error over Conventional hold for both CNN origins and training distributions (Table 12). Finite-snapshot training changes PC's IoU by at most 0.0011. Broad false-positive regions explain the high recall and low overlap under both training conditions.

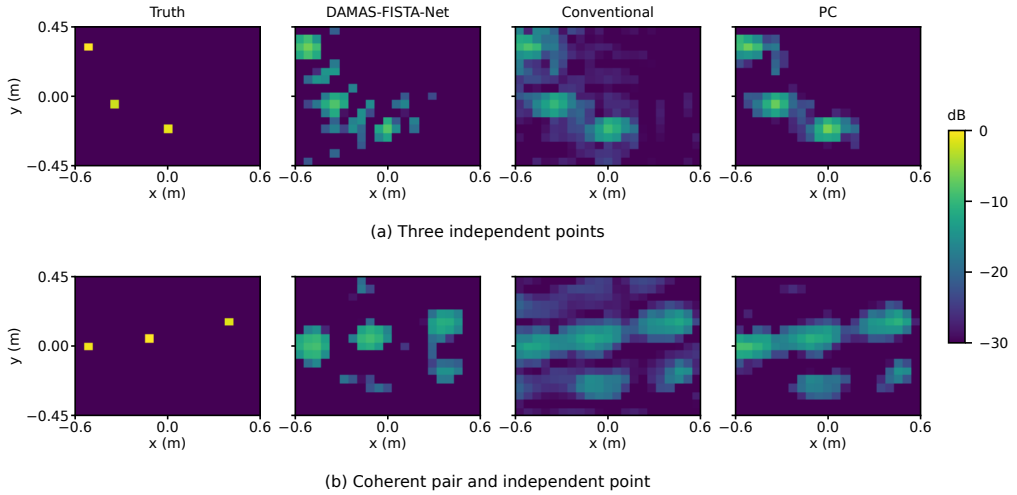


Figure 8: Representative Acoular reconstructions. Conventional and PC share the analytic-trained CNN initialization. Unit-sum maps use the same true-peak reference over $[-30, 0]$ dB within each row.

Figure 8 shows how PC reduces spatial spreading. For each source family, we use the observation closest to the median PC shape-L1 in the first evaluation seed, with the analytic-trained Conventional initializer. False energy still connects the coherent pair, while DAMAS-FISTA-Net gives shifted or broad peaks. These patterns explain the overlap and false-positive metrics in Table 12.

E.3 Gaussian and Fourier Measurements

We examine whether PC's optimization gains depend on the specific structure of acoustic steering operators. Complex Gaussian and Fourier-sampling operators (32×357) replace acoustic H, retaining HSH^* and all six source families. Refiners share a deterministic beamforming/eigenfactor start, rank six and 80 Adam updates, with separately developed rates. PC lowers the objective in all eight seeds by 76.2–78.3% without an anchor and 44.7–46.1% with $\lambda = 0.05$ (Table 13). With source families fixed, these results support fitting gains beyond acoustics, not learned-model transfer. Absolute shape errors are 0.89–1.49.

Table 13: Gaussian and Fourier operator comparison over six families and eight seeds. J is the final combined CSM and anchor objective, with lower values better. Reduction averages seed-wise $100(J_{\mathrm{Conv}} - J_{\mathrm{PC}})/J_{\mathrm{Conv}}$ with a 95% Student-t interval.

Operator	λ	Conv. J	PC J	Reduction (%) [95% CI]	Seeds
Gaussian	0	0.1299	0.0282	78.3 [76.9, 79.6]	8/8
Gaussian	0.05	0.1352	0.0728	46.1 [44.8, 47.4]	8/8
Fourier	0	0.1436	0.0345	76.2 [71.0, 81.5]	8/8
Fourier	0.05	0.1505	0.0829	44.7 [37.1, 52.4]	8/8

F Measured Proxy CSM Construction and Refinement

The 528 pair files are in `windturbine_experiment/labview_frequency_domain_pairwise_spectra/` of the repository accompanying Xue & Yang (2024), at commit `d65bf632` (Xue, 2026). This snapshot matches the files used here. The exports contain neither raw time-domain recordings nor an independently exported Welch CSM.

Pairwise assembly and phase consistency. The export supplies 32 autopowers and 496 relative phases in radians. Setting $|C_{ij}| = \sqrt{C_{ii}C_{jj}}$ forces unit coherence. The matrix is PSD, with rank one, only when its phases are transitive modulo 2π. The observed phases do not satisfy this condition.

After correction, these matrices are *proxy CSMs*: their non-unit coherence magnitudes come from correction, not direct measurement.

PSD correction. Eigenvalue clipping follows 10^{-8} mean-diagonal loading and uses the complex Hermitian analogue of the Frobenius nearest-PSD projection for real matrices (Higham, 1988). Fixed-diagonal correction uses a complex Hermitian analogue of the real symmetric Dykstra-corrected alternating projection (Higham, 2002). It projects onto the PSD/unit-diagonal intersection in normalized correlation space, then restores autopowers. The distance is minimized in correlation space, not raw CSM space. Both satisfy a 5×10^{-9} relative PSD tolerance and are used throughout the reconstruction comparison.

Geometry and frequency split. A 32-microphone nested-elliptical ground array observes a rotor of radius 44 m and hub height 75 m. Reconstruction uses a 51×51 rotor-plane grid at 72 m range and 2 m pitch. The two ground ellipses have 20 outer and 12 inner channels, with cross-stream/streamwise semi-axes of $2/4$ m and $1/2$ m. The grid spans $[-50, 50]$ m horizontally and $[25.5, 125.5]$ m in height. The column-normalized moving-source Green operator uses $c = 344$ m/s, 17 rpm and 7 m/s mean inflow. Of 81 rows, 59 determine adaptation and tuning, 16 stride-5 rows from 720–2220 Hz evaluate residuals, and six disjoint rows at 320, 800, 1000, 1240, 1600 and 2000 Hz evaluate injections. These subsets do not overlap. All subsets come from one previously analyzed export and are not independent recordings.

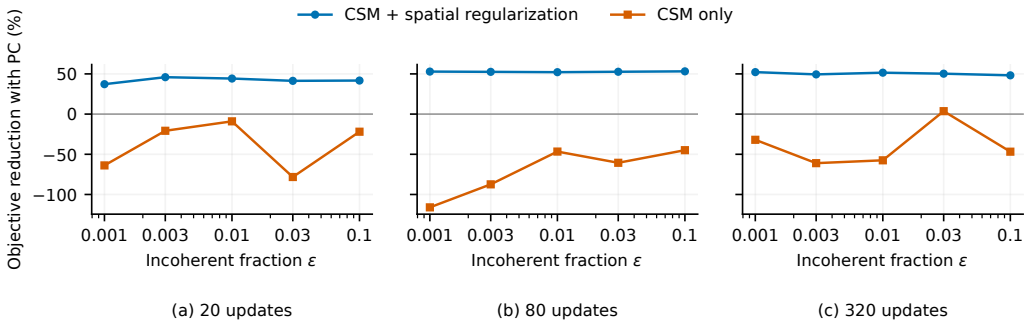


Figure 9: Sensitivity to initial coherence in wind turbine proxy CSM refinement ($r = 16$). Positive reductions favor PC. Full and CSM-only objectives are tuned separately. Larger ϵ gives a less coherent start at unchanged power. Both parameterizations share each initial covariance. Panels use common vertical limits.

F.1 Expanded Learning Rate Comparison

Factors use deterministic BF-DR or learned measurement adaptation with three seeds. Checkpoints adapted under clipping are reused under fixed-diagonal correction. After rotor support, both origins give $\mathbf{S}_{\text{pure}} = LL^*/\operatorname{tr}(LL^*)$. We initialize with

$$\mathbf{S}_0(\epsilon) = (1-\epsilon)\mathbf{S}_{\text{pure}} + \epsilon \operatorname{diag}\big(\operatorname{diag}(\mathbf{S}_{\text{pure}})\big), \qquad \epsilon \in \{0.001, 0.003, 0.01, 0.03, 0.1\}.$$

The incoherent power fraction $\epsilon > 0$ reduces coherence at fixed power and trace, keeping PC parameters finite. The Full objective combines off-diagonal CSM error, square-root sparsity (weight 0.03) and TV (weight 0.02).

Within each ϵ, correction, initializer, seed and frequency, both parameterizations share the initial covariance and use 16-column factors, hard rotor support, no map anchor and complex Adam (Appendix B). CSM-only independently optimizes the off-diagonal loss without Full's sparsity and TV terms. Excluding diagonal entries avoids direct diagonal-noise contributions, but not correlated noise or proxy-construction errors. No noise covariance is estimated or subtracted.

Full extends the ten-rate grid $\{0.0005, 0.001, 0.002, 0.005, 0.01, 0.02, 0.05, 0.1, 0.2, 0.5\}$ with $\{0.75, 1, 2, 5\}$. CSM-only keeps the ten-rate search. The 59 training rows select a separate rate for each objective, parameterization and ϵ by its weighted mean at 80 updates. We average frequencies first, then learned seeds within each origin, and give both origins and both corrections equal

weight. The selected rates are reused at 20, 80 and 320 updates on the same 16 previously evaluated residual rows. This follow-up is not an independent confirmation. We exclude a candidate rate if any of its development runs fails. This excludes rate 5 for PC at every ϵ. All evaluation runs with the selected settings remain finite. Reductions are $100(\bar{J}_{\mathrm{Conv}} - \bar{J}_{\mathrm{PC}})/\bar{J}_{\mathrm{Conv}}$, using these weighted means, not averages of individual percentages. All runs use one previously analyzed export, so we do not assign IID confidence intervals or p-values.

Table 14: Measured proxy CSM refinement with expanded learning rate selection. Entries are Conventional/PC weighted mean objectives. Lower is better. η denotes the selected rate, reused at all three budgets.

Objective	ϵ	η_{Conv}	η_{PC}	20 updates	80 updates	320 updates
Full	0.001	0.05	0.5	0.7540/0.4735	0.5134/0.2419	0.3905/0.1866
	0.003	0.05	0.5	0.8041/0.4349	0.5110/0.2424	0.3888/0.1968
	0.01	0.02	0.5	0.7670/0.4278	0.5103/0.2442	0.3930/0.1905
	0.03	0.01	0.5	0.7645/0.4483	0.5201/0.2462	0.3943/0.1961
	0.1	0.01	0.5	0.8679/0.5058	0.5319/0.2490	0.3894/0.2016
CSM only	0.001	0.01	0.2	0.1324/0.2170	0.0115/0.0248	0.0103/0.0136
	0.003	0.005	0.2	0.1446/0.1747	0.0112/0.0211	0.0092/0.0148
	0.01	0.002	0.1	0.1788/0.1949	0.0130/0.0190	0.0076/0.0119
	0.03	0.002	0.05	0.1235/0.2202	0.0120/0.0192	0.0101/0.0097
	0.1	0.001	0.05	0.1281/0.1562	0.0117/0.0169	0.0082/0.0120

Table 14 shows regularized fitting gains at every reported budget and initial coherence level. The reductions across the five settings are 37.2–45.9%, 52.1–53.2% and 48.2–52.2% after 20, 80 and 320 updates. These ranges describe settings, not uncertainty intervals. Short-budget reversals from the narrower search disappear. PC still selects 0.5 after adding the four larger rates. All selected Full and CSM-only rates lie inside their evaluated ranges. Because the selected rates are unchanged, we verify and reuse the matched evaluation trajectories rather than treat them as new data. Without spatial penalties, Conventional gives lower CSM error at 20 and 80 updates for all five settings and at 320 for four. At $\epsilon = 0.03$ and 320 updates, PC is 3.6% lower.

Table 15 shows where the regularized gain comes from. At $\epsilon = 0.01$ and 80 updates, the lower weighted sparsity penalty more than offsets the higher CSM residual. The weighted TV term contributes little to the objective. At 320 updates, both the sparsity penalty and CSM residual are lower. The Full/CSM-only comparison changes both the objective and the selected rate, so it does not isolate the effect of regularization alone.

Table 15: Objective components at $\epsilon = 0.01$. Entries are Conventional/PC weighted means. Sparsity and TV include their objective weights. Positive total objective reductions favor PC.

Objective	Updates	CSM	Weighted sparsity	Weighted TV	Total	Reduction (%)
Full	20	0.09428/0.19582	0.67271/0.23195	0.00001/0.00003	0.76701/0.42779	+44.2
Full	80	0.05536/0.07686	0.45488/0.16729	0.00002/0.00003	0.51026/0.24417	+52.1
Full	320	0.06669/0.05723	0.32625/0.13327	0.00002/0.00003	0.39296/0.19053	+51.5
CSM only	20	0.17882/0.19495	0.00000/0.00000	0.00000/0.00000	0.17882/0.19495	-9.0
CSM only	80	0.01299/0.01905	0.00000/0.00000	0.00000/0.00000	0.01299/0.01905	-46.6
CSM only	320	0.00757/0.01192	0.00000/0.00000	0.00000/0.00000	0.00757/0.01192	-57.5

F.2 Localization of Synthetic Sources in Proxy Backgrounds

At each of nine continuous positions and six frequencies, we add $\ell a_{\mathrm{bg}} h h^*$ with unit-norm steering h, $a_{\mathrm{bg}} = \mathrm{tr}(C_{\mathrm{raw}})/32$ and $\ell \in \{1, 0.25, 0.0625\}$. The nominal 0, -6 and -12 dB levels describe injection trace gain relative to mean background autopower, not array-wide signal-to-background ratios. Both corrections share this reference. Besides matched propagation, we evaluate 1% higher sound speed and zero-mean microphone offsets of 5 mm planar RMS, retaining nominal reconstruction geometry.

Rank-six refiners share a deterministic observation-derived covariance for each background or injected observation. They use 80 Adam updates, $\lambda = 0$ or 0.05, and no sparsity, TV or rotor penalties. Development-selected rates are 0.001/0.05 for Conventional/PC. We scale each reconstructed co-

variance by its nonnegative off-diagonal least-squares amplitude and subtract the background map from the injected map. A case succeeds if the peak of the signed difference map falls within $1.5\sqrt{2}(2\,\mathrm{m}) = 4.24\,\mathrm{m}$ of the injected position. Each condition has $6 \times 9 = 54$ cases.

Table 16 reports every condition. Both methods localize most strong matched injections at $\lambda = 0.05$. But with clipping at -12 dB, PC succeeds in 9.3% of cases, compared with Conventional's 63.0%. These injections do not provide independent measurements of turbine source locations.

Table 16: Injection localization success (%). Each entry gives eigenvalue-clipping/fixed-diagonal results over 54 cases. Levels are the nominal trace gain settings defined above. Both parameterizations use amplitude-aligned map differencing.

Map differencing	λ	Matched: clip/fixed			Perturbed: clip/fixed		
		0 dB	-6 dB	-12 dB	0 dB	-6 dB	-12 dB
PC	0.05	85.2/90.7	70.4/68.5	9.3/55.6	75.9/87.0	64.8/68.5	7.4/48.1
PC	0	37.0/83.3	33.3/55.6	33.3/51.9	42.6/79.6	31.5/51.9	31.5/46.3
Conventional	0.05	83.3/90.7	74.1/72.2	63.0/55.6	75.9/90.7	70.4/72.2	53.7/50.0
Conventional	0	81.5/88.9	74.1/66.7	72.2/66.7	75.9/87.0	68.5/61.1	64.8/64.8