DIFFUSION POSTERIOR SAMPLING WITH UNCERTAIN FORWARD MODELS

Anonymous authors
Paper under double-blind review

ABSTRACT

Diffusion posterior samplers reconstruct signals by combining a pretrained diffusion prior with a measurement likelihood specified by a forward model and a noise model. Calibration and reference data can characterize uncertainty in the forward model's parameters, such as blur kernels or detector responses. Fixing these parameters at estimated values leaves their uncertainty out of the likelihood, and parameter errors are thus absorbed into the reconstruction. We introduce an uncertainty-aware likelihood (UAL), a Gaussian approximation to the marginal likelihood obtained by integrating over the uncertain parameters. For each candidate signal, UAL propagates parameter uncertainty into a structured measurement covariance, reducing the quadratic penalty along residual directions that parameter variation can explain. Existing samplers can use UAL in their data updates while retaining their pretrained diffusion priors and sampling schedules. Under a fixed Gaussian nuisance model, we bound likelihood error from linearization and the resulting average divergence between the exact marginal and UAL posteriors. We compare samplers with and without UAL across six inverse problems spanning Gaussian and motion deblurring, CT, MRI, ultrasound, and fluorescence microscopy, and find improvements in reconstruction and ensemble prediction. UAL reduces geometric mean MSE by 9% to 20% across samplers in the main Gaussian-deblurring setting and by 51% to 92% in motion deblurring when non-blind samplers use the population-mean kernel. In motion deblurring, DAPS and PnP-DM with UAL also give 6.2 to 6.4 dB higher ensemble-mean PSNR than published BlindDPS and Fast Diffusion EM. On the ablation set, UAL outperforms trace-matched isotropic inflation for PnP-DM in both deblurring tasks; in motion deblurring, it also exceeds the best-tested fixed scalar data weighting by 3.3 dB for PnP-DM and 4.2 dB for DAPS in average PSNR.

1 INTRODUCTION

Diffusion posterior samplers reconstruct signals by combining a pretrained diffusion prior (Ho et al., 2020; Song et al., 2021) with a measurement likelihood specified by a forward model and an observation noise model (Chung et al., 2023b; Wu et al., 2024). Parameters in the forward model, such as a blur kernel or detector response, can remain uncertain after calibration. We treat these quantities as nuisance parameters: they affect the measurements but estimating them is not an explicit goal. Calibration and reference data provide nuisance parameter estimates and can also describe their uncertainty. Fixing these quantities at one estimate, however, leaves that uncertainty out of the reconstruction likelihood. A reconstruction process can then distort the signal to fit components of residuals caused by nuisance parameter error.

In motion deblurring, for example, a kernel bank describes variation in blur direction and extent across exposures. For a candidate image and a fixed kernel, the residual between the observed and predicted measurements combines image error, kernel error, and observation noise. A point likelihood, evaluated at one fixed kernel, penalizes this residual using only the observation noise model. For each candidate image, however, the kernel statistics and forward model predict correlated measurement changes caused by kernel variation. Residuals along these predicted directions provide weaker evidence of image error than residuals with the same observation-noise penalty that kernel variation cannot explain, so they should receive smaller quadratic penalties. The same effect arises in

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.

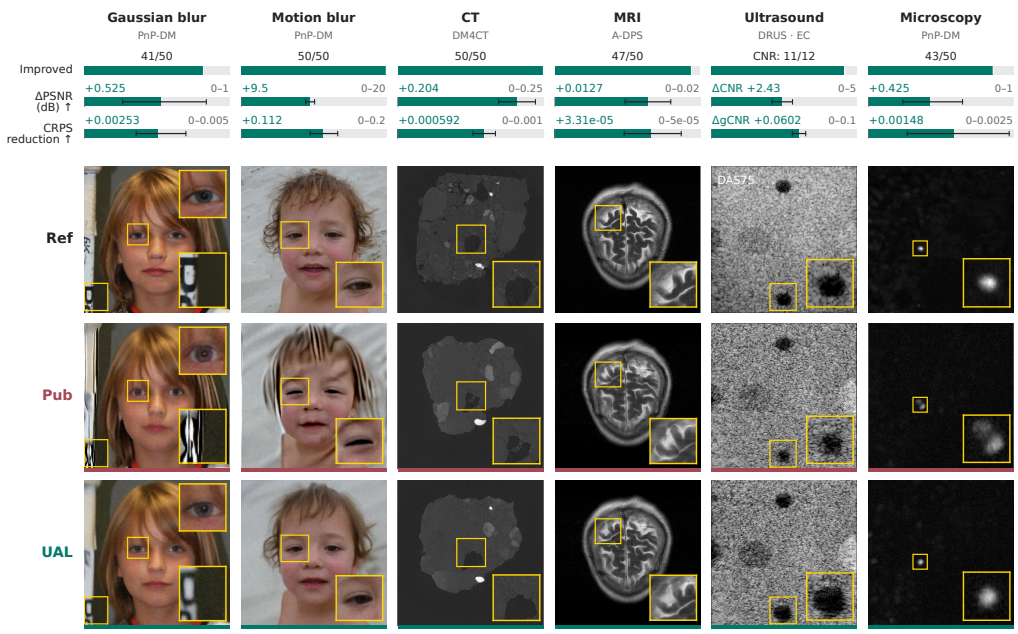

Figure 1: Reconstruction examples and quantitative effects of UAL across six inverse problems. Reconstruction rows show selected individual trajectories; bars summarize each evaluation set on task-specific scales. Ultrasound reports CNR and gCNR computed within each reconstruction on the contrast phantom, with one SE across its 12 paired trajectories; the delay-and-sum reference (DAS75) is shown. Other tasks report PSNR and CRPS with paired-bootstrap whiskers.

CT, where uncertainty in the flat and dark reference fields used for calibration changes the corrected projections even for fixed acquired data (Shi et al., 2026).

Non-blind diffusion posterior samplers condition on a fixed measurement model (Chung et al., 2023b; Kawar et al., 2022; Zhang et al., 2025), whereas blind methods sample or estimate the nuisance parameters together with the signal (Chung et al., 2023a; Murata et al., 2023; Wang et al., 2024). Some blind samplers, such as BlindDPS, train an additional diffusion prior for the nuisance parameters (Chung et al., 2023a).

We instead marginalize forward-model parameters using an uncertainty model specified by calibration and reference information. Let x denote the unknown signal, y the measurements, ψ the nuisance parameters, and q a fixed uncertainty distribution for ψ. The marginal likelihood is

$$p_{\text{marg}}(y \mid x) = \mathbb{E}_{\psi \sim q}[p(y \mid x, \psi)] \,.$$

We introduce an uncertainty-aware likelihood (UAL), a Gaussian approximation to the nuisance-marginal likelihood that allows this direction-dependent weighting to be incorporated into existing diffusion posterior samplers. At a candidate signal, UAL linearizes the forward model in the nuisance parameters and uses their mean and covariance to construct the measurement likelihood. UAL is exact for affine nuisance dependence with Gaussian nuisance parameters and independent Gaussian observation noise; for non-Gaussian q, UAL uses a Gaussian with the mean and covariance of the linearized measurement model (Section 3.1).

For a fixed Gaussian nuisance model, we bound likelihood error from linearization and the resulting posterior divergence averaged over measurements from the model (Corollary 1). In a fixed-sensitivity linear model with known signal covariance and matched nuisance moments, the linear estimator using UAL minimizes mean squared reconstruction error among linear estimators, averaged over signals, nuisance parameters, and observation noise. At a fixed nuisance realization, the point-likelihood estimator can instead have lower error. For a single nuisance parameter and distinct estimators, we give the crossover distance from the nuisance mean in closed form (Appendix B).

UAL incorporates calibration and reference statistics into existing samplers' data updates while retaining their pretrained priors and sampling schedules. We evaluate it with non-blind and blind diffusion samplers on synthetic and acquired measurements: Gaussian and motion deblurring, CT,

MRI, ultrasound, and fluorescence microscopy. A simpler alternative inflates the observation-noise variance, which in effect weakens the data term. For PnP-DM, the best tested scalar inflation gives mean per-trajectory PSNR of 27.03 dB in Gaussian deblurring and 23.37 dB in motion deblurring. UAL adds variance only along directions that parameter variation can explain and gives 27.48 and 26.64 dB, respectively (Section 4.3). UAL also improves ensemble prediction: it lowers the mean CRPS, energy, and local variogram scores for every sampler and application in Table 1. In Gaussian deblurring, UAL more closely matches the numerically marginalized likelihood and image score than the point and isotropic alternatives; covariance controls show the reconstruction benefit of structured, candidate-dependent weighting (Section 4.3). Applying UAL requires nuisance means and covariances and a forward model for propagating their variation; no additional diffusion-model training is needed. Covariance factorization reduces evaluation to forming and solving a system in nuisance coordinates. In the configurations of Table 22, UAL takes at most three times the published sampler's wall time per trajectory, except DAPS in motion deblurring at 7.1 times (Appendix D.8).

2 Background and Related Work

Diffusion priors for inverse problems. For fixed nuisance parameters ψ and a signal prior $p_0(x)$ that does not depend on ψ, Bayes' rule gives $p(x \mid y, \psi) \propto p(y \mid x, \psi)p_0(x)$, where $p(y \mid x, \psi)$ is the measurement likelihood. A pretrained diffusion model provides the prior through its denoisers or scores (Ho et al., 2020; Song et al., 2021). The likelihood enters diffusion reconstruction through three common data-update forms: gradient guidance, data-fidelity optimization, and sampling-based updates. The likelihood at a noisy diffusion state x_t, $p(y \mid x_t, \psi)$, averages over possible clean signals and is generally intractable. DPS-style gradient guidance approximates it using a denoised estimate $\hat{x}_0(x_t)$ and differentiates through that estimate (Chung et al., 2023b). ΠGDM propagates Gaussian approximations of clean-signal uncertainty through a fixed measurement operator to approximate the noisy-state likelihood (Song et al., 2023). Data-fidelity optimization solves a subproblem between denoising steps (Zhu et al., 2023). Sampling-based updates use the measurement in spectral coordinates (Kawar et al., 2022), in a split Gibbs sampler (Wu et al., 2024), or in annealed conditionals (Zhang et al., 2025). Diffusion priors with fixed-operator measurement models are also used in MRI (Aali et al., 2025), ultrasound (Zhang et al., 2024), and CT (Shi et al., 2026). In these fixed-operator formulations, the likelihood conditions on a single nuisance-parameter value. For plug-and-play unadjusted Langevin sampling, Renaud et al. (2024) bound posterior error under mismatched measurement models and denoisers. UAL instead accounts for uncertainty in the forward-model parameters and enters each sampler through its existing data update (Section 3.2).

Blind inverse problems. Blind reconstruction treats the nuisance parameters as additional inference variables. Joint samplers place a prior on them and sample them with the signal: BlindDPS couples an image diffusion model and an operator diffusion model through a shared measurement likelihood (Chung et al., 2023a); GibbsDDRM alternates approximate posterior updates under a generic operator prior (Murata et al., 2023); and PRISM pairs a measurement-conditioned kernel prior with an unconditional image prior in a split Gibbs sampler (Hu et al., 2026). Estimation methods fit them instead: Fast Diffusion EM updates the operator in a maximization step from image samples, with denoiser-based kernel regularization (Laroche et al., 2024); and DMPlug optimizes the input to a pretrained diffusion generator jointly with the operator parameters (Wang et al., 2024). Kernel Diffusion samples a conditional kernel distribution and uses a non-blind method to reconstruct the image (Sanghvi et al., 2025). Under the same joint model, exact joint inference and exact marginalization give the same signal posterior; the methods differ in how nuisance uncertainty reaches the signal update. UAL approximates the marginal measurement likelihood from a specified uncertainty model, without training an additional diffusion model for the nuisance parameters. Appendix C gives the image and operator updates for the evaluated samplers.

Calibration and reference information. We use calibration data and reference samples to specify the means and covariances of uncertain forward-model parameters. Repeated CT flat and dark frames characterize variation in calibration fields (Shi et al., 2026), and principal-component representations of flat fields support dynamic projection normalization (Nieuwenhove et al., 2015). A motion-kernel bank describes variation in blur across exposures. ESPIRiT derives coil sensitivity maps from calibration data and permits multiple maps when a single-map model is insufficient (Uecker et al., 2014). Differences between the sound speeds provided with CUBDL ac-

quisitions and their optimized values set the uncertainty scale for our PICMUS phantom reconstructions (Hyun et al., 2021; Liebgott et al., 2016). We represent the parameter variation by a distribution q and use its mean and covariance to construct UAL (Section 3.1).

Marginalization and UAL. Errors-in-variables models account for measurement error in the model's input variables (Fuller, 1987). Wiesel et al. (2008) derive maximum-likelihood estimators for linear regression with Gaussian uncertainty in the model matrix, and Beck & Eldar (2010) extend them to structured perturbations. Kumar et al. (2013) linearize sampling-location errors to obtain a Gaussian likelihood with signal-dependent variance, then update the variance weights in an iterative linear estimator. Using Bayesian approximation-error modeling (Kaipio & Somersalo, 2007; Kaipio & Kolehmainen, 2013; Nissinen et al., 2007), Mozumder et al. (2016) approximately marginalize uncertain optical coefficients with an enhanced Gaussian error model whose moments are estimated from prior samples and held fixed during reconstruction. For partially known random blurs, covariance weighting gives the classical LMMSE restoration filter (Galatsanos et al., 2000). UAL adapts this Gaussian treatment of forward-model uncertainty to diffusion posterior sampling through existing samplers' data updates, and we test it against numerical marginalization and scalar variance inflation.

3 METHOD

3.1 FORWARD MODEL AND UAL CONSTRUCTION

Let $x \in \mathbb{R}^n$ denote the unknown signal (an image in our applications), $y \in \mathbb{R}^m$ the measurement, and $\psi \in \mathbb{R}^P$ the nuisance parameters. We write the measurement model as

$$y = \mathcal{A}(x, \psi) + e, \qquad e \sim \mathcal{N}(0, \Sigma_e), \qquad \Sigma_e \succ 0, \tag{1}$$

where $\mathcal{A}$ is the forward model and e is zero-mean Gaussian observation noise with covariance $\Sigma_e \in \mathbb{R}^{m\times m}$. We represent complex measurements by stacking their real and imaginary parts.

Let q be the nuisance uncertainty model used for reconstruction, with mean $\mu = \mathbb{E}_q\psi$ and covariance $\Sigma_\psi = \mathrm{Cov}_q\psi \succeq 0$. This distribution can describe variation in a reference population or uncertainty in calibrated parameters; in controlled experiments, it can be specified directly. We hold the nuisance model fixed and model x and ψ as independent; the measurement noise is independent of both.

The marginal likelihood and its signal posterior are

$$p_{\mathrm{marg}}(y \mid x) = \mathbb{E}_{\psi\sim q}[\mathcal{N}(y; \mathcal{A}(x, \psi), \Sigma_e)], \qquad p_{\mathrm{marg}}(x \mid y) \propto p_0(x)\, p_{\mathrm{marg}}(y \mid x), \tag{2}$$

where p_0 is the signal prior. The expectation also covers a discrete reference population. By comparison, a point likelihood evaluates $\mathcal{N}(y; \mathcal{A}(x, \psi_0), \Sigma_e)$ at a single parameter value ψ_0. UAL approximates the expectation in (2), while keeping the signal prior fixed.

For each candidate signal x, we choose a reference ψ_0 and linearize the forward model in its nuisance parameters. With sensitivity $J_\psi(x)$ and remainder $r_x(\psi)$, we write

$$\mathcal{A}(x, \psi) = \mathcal{A}(x, \psi_0) + J_\psi(x)(\psi - \psi_0) + r_x(\psi), \qquad J_\psi(x) = D_\psi\mathcal{A}(x, \psi_0). \tag{3}$$

Dropping $r_x(\psi)$ and taking the mean and covariance over $\psi \sim q$ and the independent noise e gives

$$\bar{y}(x) = \mathcal{A}(x, \psi_0) + J_\psi(x)(\mu - \psi_0), \qquad S(x) = \Sigma_e + J_\psi(x)\Sigma_\psi J_\psi(x)^{\mathsf{T}}. \tag{4}$$

Nuisance uncertainty thus induces structured measurement uncertainty that may depend on x.

UAL is the Gaussian density $p_{\mathrm{UAL}}(y \mid x; \psi_0, q) = \mathcal{N}(y; \bar{y}(x), S(x))$ for this fixed local model. The mean correction $J_\psi(x)(\mu - \psi_0)$ accounts for the difference between the reference and the nuisance mean (Remark 3). It vanishes at $\psi_0 = \mu$.

Proposition 1 (Exact affine-Gaussian marginal)**.** *At fixed x, suppose $\mathcal{A}(x, \psi)$ is affine in ψ, $q = \mathcal{N}(\mu, \Sigma_\psi)$, and e is independent of ψ. Then $p_{\mathrm{marg}}(y \mid x) = \mathcal{N}(y; \bar{y}(x), S(x))$, with moments given by (4). For nonlinear dependence on ψ, the same moments are exact for the linearized forward model in (3).*

For non-Gaussian q, UAL approximates the measurement likelihood by a Gaussian with mean $\bar{y}(x)$ and covariance $S(x)$ from (4). Appendix A gives the proofs and bounds on moment errors from linearization and on covariance truncation.

3.2 UAL EVALUATION AND SAMPLER INTEGRATION

Energy and score. Define the measurement residual $d(x) = y - \bar{y}(x)$, the precision matrix $S(x)^{-1}$, and the precision-weighted residual $w(x) = S(x)^{-1}d(x)$. The normalized negative log likelihood is

$$\ell_{\mathrm{UAL}}(x; y) = \frac{1}{2}d(x)^{\mathsf{T}}S(x)^{-1}d(x) + \frac{1}{2}\log\det S(x) + \frac{m}{2}\log(2\pi). \tag{5}$$

Increasing measurement variance along a residual direction lowers its quadratic penalty, while the log determinant grows with the corresponding increase in covariance volume.

For the whitened residual $\Sigma_e^{-1/2}d$, the quadratic weight along a left singular direction of $\Sigma_e^{-1/2}J_\psi F$ is $1/(1+s_i^2)$, where $FF^{\mathsf{T}} = \Sigma_\psi$ and s_i is the corresponding singular value. Here s_i^2 is the added variance along that whitened direction. Directions orthogonal to the whitened columns keep unit weight. Residuals with the same observation-noise penalty can therefore receive different UAL quadratic penalties, according to their compatibility with the predicted nuisance variation. Scalar covariance inflation, by contrast, assigns equal weights in noise-whitened coordinates; with x and $\bar{y}(x)$ fixed, it has a positive minimum KL divergence from UAL unless the nuisance-induced covariance is proportional to the observation-noise covariance (Remark 1).

For differentiable $\bar{y}(x)$ and $S(x)$, with the nuisance model and reference fixed, the likelihood score is given componentwise by

$$\partial_{x_i}\log p_{\mathrm{UAL}}(y \mid x) = (\partial_{x_i}\bar{y})^{\mathsf{T}}w + \frac{1}{2}w^{\mathsf{T}}(\partial_{x_i}S)w - \frac{1}{2}\operatorname{tr}\big(S^{-1}\partial_{x_i}S\big). \tag{6}$$

The first term differentiates the predicted mean, and the remaining terms account for covariance weighting and normalization.

Efficient evaluation. Factoring the nuisance covariance reduces UAL evaluation to an $r \times r$ system, avoiding an explicit inverse of the $m \times m$ measurement covariance. Write $\Sigma_\psi = FF^{\mathsf{T}}$, where r is the number of columns of $F \in \mathbb{R}^{P\times r}$ and Σ_ψ may be singular, and set $U(x) = J_\psi(x)F \in \mathbb{R}^{m\times r}$. Define the $r \times r$ matrix $G(x) = I_r + U(x)^{\mathsf{T}}\Sigma_e^{-1}U(x)$. The reduced solve gives

$$Ga_{\mathrm{red}} = U^{\mathsf{T}}\Sigma_e^{-1}d, \qquad w = \Sigma_e^{-1}(d - Ua_{\mathrm{red}}).$$

Under the local Gaussian model at fixed x, Ua_{red} is the mean nuisance contribution conditional on the residual d. The remaining residual is then weighted by the observation-noise precision Σ_e^{-1}. The normalization follows from $\log\det S = \log\det\Sigma_e + \log\det G$. Appendix A.2 derives the factorized energy and gives Algorithm 1. The cost also includes evaluating the forward model and its parameter sensitivities and forming G. Keeping only leading covariance modes reduces the solve dimension r. With the predicted mean $\bar{y}(x)$ unchanged, Corollary 2 bounds the resulting likelihood error. Appendix D.8 reports recorded reconstruction times.

Use in diffusion samplers. Across applications, UAL treats forward-model uncertainty in the same way: it propagates parameter variation into a measurement covariance. We incorporate UAL through each sampler's published data update (Zhang et al., 2025; Wu et al., 2024), retaining its diffusion prior and sampling schedule. Appendix Table 4 compares the point likelihood and UAL. Appendix C specifies the parameter models, reference values, and sampler updates.

3.3 APPROXIMATION ERROR AND RECONSTRUCTION RISK

We bound the likelihood error introduced by nuisance linearization and the resulting average divergence between the exact marginal and UAL posteriors. For nonlinear nuisance dependence, likelihood error is controlled by the remainder $r_x(\psi)$ in (3), measured in observation-noise units.

Proposition 2 (Likelihood error from linearization)**.** *Fix x and ψ_0, let $q = \mathcal{N}(\mu, \Sigma_\psi)$, and assume $\mathbb{E}_q[r_x(\psi)^{\mathsf{T}}\Sigma_e^{-1}r_x(\psi)] < \infty$. Then*

$$D_{\mathrm{KL}}\big(p_{\mathrm{marg}}(\cdot \mid x) \,\big\|\, p_{\mathrm{UAL}}(\cdot \mid x; \psi_0, q)\big) \le \tfrac{1}{2}\mathbb{E}_q\big[r_x^{\mathsf{T}}\Sigma_e^{-1}r_x\big]. \tag{7}$$

Appendix A.3 gives the proof. Under the conditions of Corollary 1, the prior-averaged bound in (7) also bounds posterior divergence averaged over marginal-model measurements. These bounds cover the likelihood approximation and its induced posterior, excluding the sampler updates.

Reconstruction-risk analysis. In the linear model of Appendix B, with nuisance sensitivity held fixed as the signal varies, the linear estimator based on UAL minimizes mean squared reconstruction error among linear estimators for any nuisance distribution whose mean and covariance match the assumed ones (Proposition 4). For a fixed parameter realization near the mean, the point estimator can have lower error; Appendix B characterizes the crossover and the conditions for positive average gain when these moments are inaccurate.

4 EXPERIMENTS

4.1 EXPERIMENTAL PROTOCOL

We evaluate UAL at two levels: reconstruction and ensemble prediction in paired sampler comparisons (Section 4.2), and, at shared candidate images, likelihood and image-score accuracy against numerical nuisance marginalization (Section 4.3). Additional controls examine covariance structure, recomputation and differentiation, uncertainty level, and retained rank.

Published denotes the released sampler and UAL its uncertainty-aware version; each pair uses the same measurement data, diffusion checkpoint, sampling schedule, step sizes, and reconstruction seeds. Deblurring, CT, MRI, and microscopy use 50 cases, and ultrasound uses the acquired PICMUS contrast and resolution phantoms provided with DRUS; each case has 12 reconstruction seeds. The fixed deblurring ablation set contains 16 images, with 12 paired reconstruction seeds per image. Appendix C gives forward models, nuisance models, and sampler updates.

We report PSNR and SSIM both as the mean over trajectories and for the ensemble-mean reconstruction. CRPS, energy, and local variogram scores (lower is better) use the fair finite-ensemble estimators of Appendix C.7, which also gives the scoring coordinates and interval construction; ultrasound uses native contrast and resolution metrics.

4.2 RECONSTRUCTION ACROSS SAMPLERS AND APPLICATIONS

We compare the Published and UAL versions of DAPS, PnP-DM, DiffPIR, and BlindDPS for Gaussian and motion deblurring on FFHQ images (Karras et al., 2019) (Table 1). The Gaussian setting uses kernel widths with mean $\bar{s} = 3$ pixels and standard deviation $\tau = 0.5$ pixels. In motion deblurring, each non-blind pair uses the population-mean kernel, giving the two likelihoods the same mean response at a common candidate image. Motion deblurring also includes Fast Diffusion EM (Laroche et al., 2024), a blind sampler that estimates a kernel for each image. Its UAL image update uses the population-mean kernel instead, with covariance accounting for image and kernel uncertainty and their interaction. We estimate the kernel covariance and retain its leading 256 eigenmodes ($P = 256$), with this rank fixed before the covariance-rank sweep (Section 4.3).

In Table 1, UAL improves ensemble-mean PSNR and SSIM and lowers the mean CRPS, energy, and local variogram scores in every deblurring comparison. In motion deblurring, DAPS and PnP-DM with UAL exceed the published blind samplers BlindDPS and Fast Diffusion EM, which sample or estimate a kernel for each image, by 6.2 to 6.4 dB in ensemble-mean PSNR, with paired 95% intervals excluding zero (Table 9). For mean per-trajectory PSNR, UAL gives 28.84 dB for DAPS and 26.54 dB for PnP-DM, compared with 30.65 and 27.65 dB when the published samplers use the true kernel (Table 8).

UAL also improves reconstruction and ensemble prediction in CT, MRI, and fluorescence microscopy (Figure 1). Mean per-trajectory PSNR increases on all 50 CT detector rows, 47 of 50 MRI scans, and 43 of 50 microscopy sections. For CT, the gain persists when both versions use the same mean flat and dark reference fields (Appendix D). For MRI under the Gaussian coil-sensitivity model, UAL gives the exact marginal measurement likelihood. The MRI reconstruction gains are small but consistent: all seven reported metrics improve, with pointwise 95% paired-bootstrap intervals excluding zero. On the acquired PICMUS ultrasound contrast phantom, UAL increases CNR from 9.929 to 12.363 dB and gCNR from 0.9141 to 0.9743.

Table 1: Reconstruction and ensemble prediction. T/M: mean per-trajectory score / score of the ensemble-mean reconstruction. PSNR is in dB; CRPS, energy, and LVS are fair ensemble scores, with LVS scaled by 10^3. Bold marks better UAL values.

Sampler / task	Version	PSNR T/M ↑	SSIM T/M ↑	CRPS ↓	Energy ↓	LVS ↓
Gaussian blur						
DAPS	Published	29.49 / 30.60	.7416 / .8143	.02871	20.128	11.240
	UAL	**29.88 / 30.94**	**.7468 / .8171**	**.02796**	**19.449**	**11.021**
PnP-DM	Published	26.97 / 29.53	.5603 / .7616	.03209	20.874	15.286
	UAL	**27.50 / 30.35**	**.5721 / .7776**	**.02956**	**18.803**	**14.252**
DiffPIR	Published	29.13 / 30.28	.7687 / .8180	.02948	21.263	10.788
	UAL	**30.12 / 31.30**	**.7860 / .8331**	**.02583**	**18.325**	**9.353**
BlindDPS	Published	28.31 / 30.30	.7343 / .8002	.02698	19.322	9.739
	UAL	**29.21 / 31.08**	**.7708 / .8278**	**.02476**	**17.778**	**8.895**
Motion blur						
DAPS	Published	17.68 / 18.10	.3303 / .3779	.15043	96.013	57.076
	UAL	**28.84 / 29.58**	**.7103 / .7726**	**.03502**	**23.388**	**12.774**
PnP-DM	Published	17.04 / 17.90	.2766 / .3897	.14485	91.587	61.032
	UAL	**26.54 / 29.43**	**.5552 / .7628**	**.03236**	**20.943**	**14.959**
DiffPIR	Published	11.44 / 12.17	.0701 / .0877	.34406	172.270	179.800
	UAL	**14.55 / 15.19**	**.1440 / .1666**	**.24688**	**130.857**	**100.688**
BlindDPS	Published	22.12 / 23.21	.5559 / .6263	.06392	46.633	18.291
	UAL	21.95 / **23.33**	.5449 / **.6278**	**.06218**	**44.589**	**17.339**
Fast Diffusion EM	Published	22.76 / 23.27	.5236 / .5525	.07789	51.865	24.271
	UAL	**29.10 / 30.72**	**.7635 / .8108**	**.02708**	**18.332**	**9.633**
Imaging applications						
CT	Published	25.94 / 27.33	.4248 / .5195	.02194	23.464	5.361
	UAL	**26.15 / 27.54**	**.4321 / .5299**	**.02135**	**22.898**	**5.355**
MRI	Published	30.77 / 31.88	.8435 / .8179	.02007	11.562	5.817
	UAL	**30.78 / 31.89**	**.8438 / .8180**	**.02004**	**11.540**	**5.806**
Microscopy	Published	29.33 / 33.67	.5352 / .7912	.02265	8.383	10.621
	UAL	**29.76 / 34.22**	**.5505 / .7971**	**.02117**	**7.482**	**9.740**

4.3 Likelihood validation and controls

Likelihood approximation. For Gaussian deblurring, we evaluate the published PnP-DM point likelihood, UAL, and isotropic inflation at common final reconstructions produced by the published sampler. Numerical nuisance marginalization provides the reference. The isotropic control retains the UAL mean and scales the observation-noise covariance to $v_\star(x)\Sigma_e$, where $v_\star(x) = \mathrm{tr}(\Sigma_e^{-1}S(x))/m$ is the average measurement variance under UAL after noise whitening. The isotropic and UAL covariances therefore have the same whitened trace. Among Gaussian likelihoods with the same mean and covariance $v\Sigma_e$, this trace match minimizes the KL divergence from UAL (Remark 1). At final reconstructions, UAL's mean relative image-score error is below 1% at every tested uncertainty level, compared with 6% to 15% for the point likelihood. Its mean absolute negative-log-likelihood error is also the lowest of the three likelihoods (Figure 2a,b). During sampling, UAL also has the lowest mean relative image-score error at each of the five saved states and each uncertainty level. At later states, its mean error is at most 2%, compared with approximately 6% to 19% for the point likelihood (Figure 2c). Across these common candidates, UAL more closely approximates both the marginalized log likelihood and its image score than the two alternatives.

Covariance structure. We next test whether covariance structure improves reconstruction and ensemble prediction beyond increasing total variance, using the three likelihoods in deblurring with PnP-DM. UAL and the isotropic control both include Gaussian normalization and differentiate their signal-dependent covariance. UAL's structured covariance increases per-trajectory PSNR and reduces CRPS relative to trace-matched isotropic inflation in both tasks (Table 2). We also test fixed noise-variance multipliers and, for each task, report the one with the highest mean per-trajectory PSNR on the same ablation set used for comparison. The best tested values give 27.029 dB ($v = 3$) in Gaussian and 23.369 dB ($v = 48$) in motion deblurring, compared with 27.475 and 26.641 dB for UAL (Appendix D.4). For DAPS in motion deblurring, the best tested multiplier of its data-term variance ($v = 48$) gives 24.413 dB mean per-trajectory PSNR, compared with 28.631 dB for UAL

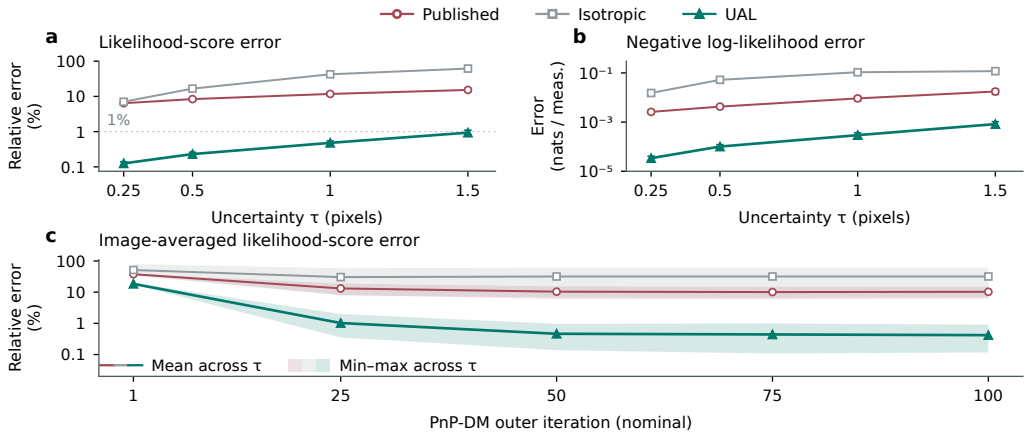

Figure 2: Likelihood accuracy in Gaussian deblurring against numerical marginalization. (a,b) Final-state relative image-score error and absolute negative-log-likelihood error per measurement: means ± one SE across the ablation set. (c) Score error at five saved states, using one trajectory per image. Errors are averaged over images within each uncertainty level. Lines and bands show the mean and range across $\tau = 0.25, 0.5, 1, 1.5$ pixels. Generating and assumed width uncertainty vary together. Appendix D.1 gives the stored indices corresponding to panel (c)'s nominal iterations.

on the ablation set. At the final UAL reconstructions in Gaussian deblurring with PnP-DM, nuisance variation adds about 20% to the average whitened measurement variance, carried by the two width-sensitivity responses (Appendix D.5). Trace-matched isotropic inflation distributes the same added variance uniformly in noise-whitened coordinates.

Table 2: Covariance structure at matched whitened trace on the ablation set. Appendix D.4 reports paired differences and standard errors.

	Gaussian blur			Motion blur		
Metric	Published	Isotropic	UAL	Published	Isotropic	UAL
PSNR T (dB) ↑	26.944	27.020	**27.475**	17.872	24.084	**26.641**
CRPS ($\times 10^3$) ↓	32.53	32.42	**29.81**	131.53	58.55	**31.49**

Covariance updates and derivatives. In Gaussian deblurring with PnP-DM, two further controls separate recomputing the covariance from differentiating it (Table 3). Fixing the covariance at the published sampler's ensemble mean removes nearly all of the PSNR gain and raises CRPS above the published sampler's. Recomputing it at each candidate keeps the gain whether or not its derivatives are propagated. By contrast, the motion-deblurring controls for Fast Diffusion EM keep the population-mean kernel fixed: differentiating the covariance raises mean per-trajectory PSNR from 23.80 to 28.49 dB, and including the log determinant gives 29.10 dB (Appendix D.4).

Table 3: Gaussian PnP-DM covariance-update controls on the ablation set. The two recomputed treatments agree within two standard errors in per-trajectory PSNR.

Metric	Published	Fixed covariance	Derivatives stopped	UAL
PSNR T (dB) ↑	26.944	26.964	27.506	27.475
CRPS ($\times 10^3$) ↓	32.53	34.84	29.76	29.81

Matched uncertainty. In Gaussian deblurring with PnP-DM, UAL's reconstruction advantage grows as the generating and modeled width uncertainty increase together (Figure 3). At $\tau = 1.5$ pixels, UAL exceeds Published by 1.75 dB and Isotropic by 0.47 dB in mean per-trajectory PSNR (Table 18). In a separate comparison of three combinations of mean width and width uncertainty, UAL gives higher mean per-trajectory PSNR and lower CRPS for all four samplers (Table 19).

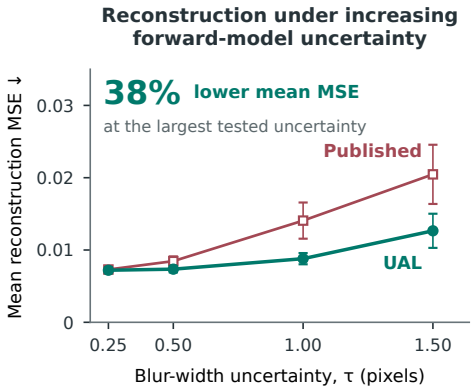

Figure 3: Reconstruction error under matched Gaussian-width uncertainty with PnP-DM at mean width 3 pixels. Generating and modeled width standard deviations increase together. Mean per-trajectory MSE ± one SE across the ablation set.

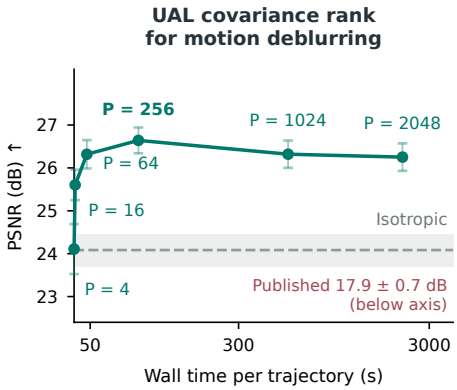

Figure 4: Covariance rank and runtime in motion deblurring with PnP-DM, labeled by the number P of retained kernel-covariance modes. Mean per-trajectory PSNR ± one SE across the ablation set; Isotropic is trace-matched at $P = 256$.

Covariance rank. The rank sweep with PnP-DM shows that a compact covariance representation preserves reconstruction quality in motion deblurring (Figure 4). The 256-mode representation retains 96% of the kernel variance. On the ablation set, it gives 26.641 dB mean per-trajectory PSNR, compared with 26.251 dB for 2048 modes. The corresponding mean wall times are 89.5 and 2171.3 seconds per trajectory, respectively.

Sensitivity to nuisance moments. With Gaussian-deblurring observations fixed, PnP-DM with UAL gives mean per-trajectory PSNR between 27.408 and 27.475 dB across the tested changes to either the parameter mean or covariance scale, up to approximately 20%. The published sampler with the correct mean width gives 26.944 dB, below every tested UAL setting (Figure 12). Underestimating the mean width increases CRPS more than any of the tested covariance rescalings. Motion deblurring is similarly insensitive to 20% changes in covariance scale (Table 17).

5 Conclusion and Limitations

UAL uses forward-model uncertainty to weight measurement-residual directions according to how well parameter variation can explain them. It enters existing samplers through their data updates without additional diffusion-model training, and it improves reconstruction in CT, MRI, and fluorescence microscopy and contrast in ultrasound. In controlled Gaussian and motion deblurring comparisons with PnP-DM, retaining covariance structure improves reconstruction beyond trace-matched isotropic inflation. Accounting for forward-model uncertainty improves ensemble prediction: UAL lowers the mean CRPS, energy, and local variogram scores in all 12 comparisons in Table 1. For PnP-DM in Gaussian deblurring, UAL has lower mean relative image-score error than the published and isotropic likelihoods at the final reconstructions and every evaluated sampling state, using numerical marginalization as the reference. UAL's accuracy and reconstruction benefit depend on the nuisance model and, for nonlinear nuisance dependence, on the local forward-model approximation. UAL adds computation that grows with the covariance rank and depends on the sampler (Appendix D.8). In deblurring, UAL keeps its gain under tested mean or covariance-scale errors of up to about 20% (Figure 12, Table 17); a different motion-kernel population was not tested.

AI use statement

With immense gratitude, we acknowledge the assistance of generative AI tools with writing and manuscript revision; retrieval and assessment of related literature; research ideation and execution, including feedback on methodology, experimental debugging, and interpretation of results; and mathematical proofs.

The authors directed the research and made the final decisions on methods, experiments, claims, and presentation. We manually checked the AI-assisted literature review, reviewed AI-debugged code and experiments, and verified reported numerical results against the underlying experiments. We also used multiple AI systems to cross-check mathematical arguments, notation, logical consistency, and manuscript clarity.

The AI-assisted draft underwent repeated human revision for technical accuracy, consistency, and readability. Suggestions that did not meet these standards were revised or rejected. AI was neither employed nor otherwise incorporated into the preparation, formulation, or drafting of this AI use statement. The authors take responsibility for the final content of this work.

Ethics statement

This work uses existing image datasets and acquired measurements to evaluate reconstruction and ensemble prediction. Appendix C identifies the data sources and includes dataset-specific license and acknowledgment information. Clinical use requires validation under the intended acquisition conditions, including assessment of reconstruction errors that could affect image interpretation.

Reproducibility statement

Appendices A and B give the assumptions and proofs. Appendix C specifies the task models, sampler updates, data, and metrics; Appendix D gives the numerical likelihood reference, ablations, and recorded timings. Code will be made available upon publication.

References

Asad Aali, Giannis Daras, Brett Levac, Sidharth Kumar, Alex Dimakis, and Jon Tamir. Ambient diffusion posterior sampling: Solving inverse problems with diffusion models trained on corrupted data. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=qeXcMutEZY.

Amir Beck and Yonina C. Eldar. Structured total maximum likelihood: An alternative to structured total least squares. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2623–2649, 2010. doi: 10.1137/090756338. URL https://doi.org/10.1137/090756338.

Hyungjin Chung, Jeongsol Kim, Sehui Kim, and Jong Chul Ye. Parallel diffusion models of operator and image for blind inverse problems. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6059–6069, June 2023a.

Hyungjin Chung, Jeongsol Kim, Michael Thompson Mccann, Marc Louis Klasky, and Jong Chul Ye. Diffusion posterior sampling for general noisy inverse problems. In *The Eleventh International Conference on Learning Representations*, 2023b. URL https://openreview.net/forum?id=OnD9zGAGT0k.

C. A. T. Ferro. Fair scores for ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 140(683):1917–1923, 2014. doi: https://doi.org/10.1002/qj.2270. URL https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/qj.2270.

Wayne A. Fuller. *Measurement Error Models*. Wiley, June 1987. ISBN 9780470316665. doi: 10.1002/9780470316665. URL http://dx.doi.org/10.1002/9780470316665.

N.P. Galatsanos, V.Z. Mesarovic, R. Molina, and A.K. Katsaggelos. Hierarchical bayesian image restoration from partially known blurs. *IEEE Transactions on Image Processing*, 9(10):1784–1797, 2000. doi: 10.1109/83.869189.

Tilmann Gneiting and Adrian E Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: 10.1198/016214506000001437. URL https://doi.org/10.1198/016214506000001437.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 6840–6851. Curran Associates, Inc.,

2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/4c5bcfec8584af0d967f1ab10179ca4b-Paper.pdf.

Yuanyun Hu, Evan Bell, Guijin Wang, and Yu Sun. PRISM: Probabilistic and robust inverse solver with measurement-conditioned diffusion prior for blind inverse problems. In *ICASSP 2026 - 2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 11432–11436, 2026. doi: 10.1109/ICASSP55912.2026.11462668.

Dongwoon Hyun, Alycen Wiacek, Sobhan Goudarzi, Sven Rothlübbers, Amir Asif, Klaus Eickel, Yonina C. Eldar, Jiaqi Huang, Massimo Mischi, Hassan Rivaz, David Sinden, Ruud J. G. van Sloun, Hannah Strohm, and Muyinatu A. Lediju Bell. Deep learning for ultrasound image formation: CUBDL evaluation framework and open datasets. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 68(12):3466–3483, 2021. doi: 10.1109/TUFFC.2021.3094849.

Jari Kaipio and Ville Kolehmainen. 32 approximate marginalization over modelling errors and uncertainties in inverse problems. In Paul Damien, Petros Dellaportas, Nicholas G. Polson, and David A. Stephens (eds.), *Bayesian Theory and Applications*. Oxford University Press, 01 2013. ISBN 9780199695607. doi: 10.1093/acprof:oso/9780199695607.003.0032. URL https://doi.org/10.1093/acprof:oso/9780199695607.003.0032.

Jari Kaipio and Erkki Somersalo. Statistical inverse problems: Discretization, model reduction and inverse crimes. *Journal of Computational and Applied Mathematics*, 198(2):493–504, 2007. ISSN 0377-0427. doi: https://doi.org/10.1016/j.cam.2005.09.027. URL https://www.sciencedirect.com/science/article/pii/S0377042705007296. Special Issue: Applied Computational Inverse Problems.

Tero Karras, Samuli Laine, and Timo Aila. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 23593–23606. Curran Associates, Inc., 2022. doi: 10.52202/068431-1714. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/95504595b6169131b6ed6cd72eb05616-Paper-Conference.pdf.

Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J. Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzalv, Adriana Romero, Michael Rabbat, Pascal Vincent, James Pinkerton, Duo Wang, Nafissa Yakubova, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: A publicly available raw k-space and DICOM dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, January 2020. ISSN 2638-6100. doi: 10.1148/ryai.2020190007. URL http://dx.doi.org/10.1148/ryai.2020190007.

Sumeet Kumar, Vivek K Goyal, and Sanjay E. Sarma. Efficient parametric signal estimation from samples with location errors. *IEEE Transactions on Signal Processing*, 61(21):5285–5297, 2013. doi: 10.1109/TSP.2013.2274641.

Charles Laroche, Andrés Almansa, and Eva Coupeté. Fast diffusion EM: A diffusion model for blind inverse problems with application to deconvolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5271–5281, January 2024.

H. Liebgott, A. Rodriguez-Molares, F. Cervenansky, J.A. Jensen, and O. Bernard. Plane-wave imaging challenge in medical ultrasound. In *2016 IEEE International Ultrasonics Symposium (IUS)*, pp. 1–4, 2016. doi: 10.1109/ULTSYM.2016.7728908.

Qinxuan Luo, Ziwen Liu, and Ge Yang. Diffusion prior-based zero-shot denoising for real-world fluorescence microscopy images via encoder-decoder guided sampling. *Biomed. Opt. Express*, 17(1):145–159, Jan 2026. doi: 10.1364/BOE.578464. URL https://opg.optica.org/boe/abstract.cfm?URI=boe-17-1-145.

Meghdoot Mozumder, Tanja Tarvainen, Simon Arridge, Jari P. Kaipio, Cosimo D'Andrea, and Ville Kolehmainen. Approximate marginalization of absorption and scattering in fluorescence diffuse optical tomography. *Inverse Problems and Imaging*, 10(1):227–246, 2016. ISSN 1930-8337. doi: 10.3934/ipi.2016.10.227. URL https://www.aimsciences.org/article/id/47f527ac-1bc1-4891-b8d9-898621fe4a63.

Naoki Murata, Koichi Saito, Chieh-Hsin Lai, Yuhta Takida, Toshimitsu Uesaka, Yuki Mitsufuji, and Stefano Ermon. GibbsDDRM: A partially collapsed Gibbs sampler for solving blind inverse problems with denoising diffusion restoration. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 25501–25522. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/murata23a.html.

Vincent Van Nieuwenhove, Jan De Beenhouwer, Francesco De Carlo, Lucia Mancini, Federica Marone, and Jan Sijbers. Dynamic intensity normalization using eigen flat fields in X-ray imaging. *Opt. Express*, 23(21):27975–27989, Oct 2015. doi: 10.1364/OE.23.027975. URL https://opg.optica.org/oe/abstract.cfm?URI=oe-23-21-27975.

Antti Nissinen, Lasse M. Heikkinen, and Jari P. Kaipio. The Bayesian approximation error approach for electrical impedance tomography—experimental results. *Measurement Science and Technology*, 19:015501, 2007. URL https://api.semanticscholar.org/CorpusID:121803415.

Robert J. Noll. Zernike polynomials and atmospheric turbulence∗. *J. Opt. Soc. Am.*, 66(3):207–211, Mar 1976. doi: 10.1364/JOSA.66.000207. URL https://opg.optica.org/abstract.cfm?URI=josa-66-3-207.

Marien Renaud, Jiaming Liu, Valentin De Bortoli, Andres Almansa, and Ulugbek Kamilov. Plug-and-play posterior sampling under mismatched measurement and prior models. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 10139–10168, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/2a2874875861f6a6436b505dd77683d1-Paper-Conference.pdf.

Yash Sanghvi, Yiheng Chi, and Stanley H. Chan. Kernel diffusion: An alternate approach to blind deconvolution. In Aleš Leonardis, Elisa Ricci, Stefan Roth, Olga Russakovsky, Torsten Sattler, and Gül Varol (eds.), *Computer Vision – ECCV 2024*, pp. 1–20, Cham, 2025. Springer Nature Switzerland. ISBN 978-3-031-73202-7.

Michael Scheuerer and Thomas M. Hamill. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Monthly Weather Review*, 143(4):1321 – 1334, 2015. doi: 10.1175/MWR-D-14-00269.1. URL https://journals.ametsoc.org/view/journals/mwre/143/4/mwr-d-14-00269.1.xml.

Jiayang Shi, Daniel Pelt, and Joost Batenburg. DM4CT: Benchmarking diffusion models for computed tomography reconstruction. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=YE5scJekg5.

Jiaming Song, Arash Vahdat, Morteza Mardani, and Jan Kautz. Pseudoinverse-guided diffusion models for inverse problems. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=9_gsMA8MRKQ.

Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=PxTIG12RRHS.

Martin Uecker, Peng Lai, Mark J. Murphy, Patrick Virtue, Michael Elad, John M. Pauly, Shreyas S. Vasanawala, and Michael Lustig. ESPIRiT—an eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001, 2014. doi: https://doi.org/10.1002/mrm.24751. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.24751.

Hengkang Wang, Xu Zhang, Taihui Li, Yuxiang Wan, Tiancong Chen, and Ju Sun. DMPlug: A plug-in method for solving inverse problems with diffusion models. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 117881–117916. Curran Associates, Inc., 2024. doi: 10.52202/079017-3744. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/d5c04aa72b92c53bda5b525b60958295-Paper-Conference.pdf.

Martin Weigert, Uwe Schmidt, Tobias Boothe, Andreas Müller, Alexandr Dibrov, Akanksha Jain, Benjamin Wilhelm, Deborah Schmidt, Coleman Broaddus, Siân Culley, Mauricio Rocha-Martins, Fabián Segovia-Miranda, Caren Norden, Ricardo Henriques, Marino Zerial, Michele Solimena, Jochen Rink, Pavel Tomancak, Loic Royer, Florian Jug, and Eugene W. Myers. Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods*, 15(12): 1090–1097, 2018. doi: 10.1038/s41592-018-0216-7. URL https://doi.org/10.1038/s41592-018-0216-7.

Ami Wiesel, Yonina C. Eldar, and Arie Yeredor. Linear regression with Gaussian model uncertainty: Algorithms and bounds. *IEEE Transactions on Signal Processing*, 56(6):2194–2205, 2008. doi: 10.1109/TSP.2007.914323.

Zihui Wu, Yu Sun, Yifan Chen, Bingliang Zhang, Yisong Yue, and Katherine L. Bouman. Principled probabilistic imaging using diffusion models as plug-and-play priors. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, volume 37, pp. 118389–118427. Curran Associates, Inc., 2024. doi: 10.52202/079017-3759. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/d65c4ce22241138c1784ff753d4c746c-Paper-Conference.pdf.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An open dataset and benchmarks for accelerated MRI, 2019. URL https://arxiv.org/abs/1811.08839.

Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, and Yang Song. Improving diffusion inverse problem solving with decoupled noise annealing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 20895–20905, June 2025.

Yuxin Zhang, Clément Huneau, Jérôme Idier, and Diana Mateus. Ultrasound image reconstruction with denoising diffusion restoration models. In Anirban Mukhopadhyay, Ilkay Oksuz, Sandy Engelhardt, Dajiang Zhu, and Yixuan Yuan (eds.), *Deep Generative Models*, pp. 193–203, Cham, 2024. Springer Nature Switzerland. ISBN 978-3-031-53767-7.

Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 1219–1229, June 2023.

A UAL LIKELIHOOD: DERIVATIONS AND APPROXIMATION THEORY

UAL is exact for an affine nuisance response with Gaussian nuisance parameters and independent Gaussian observation noise. We derive the likelihood score and factorized evaluation. We then bound likelihood, averaged posterior, and moment errors from nuisance linearization, followed by the likelihood error from covariance truncation.

Table 4: Point likelihood and UAL at a fixed reference ψ_0. Here the residuals are $d_0 = y - \mathcal{A}(x, \psi_0)$ and $d = y - \bar{y}(x)$. Both energies omit $m \log(2\pi)/2$; $\log\det \Sigma_e$ is constant in x.

Quantity	Point Gaussian likelihood	UAL
Mean	$\mathcal{A}(x, \psi_0)$	$\bar{y}(x) = \mathcal{A}(x, \psi_0) + J_\psi(x)(\mu - \psi_0)$
Covariance	Σ_e	$S(x) = \Sigma_e + J_\psi(x)\Sigma_\psi J_\psi(x)^\mathsf{T}$
Negative log likelihood	$\frac{1}{2} d_0^\mathsf{T} \Sigma_e^{-1} d_0 + \frac{1}{2} \log\det \Sigma_e$	$\frac{1}{2} d^\mathsf{T} S(x)^{-1} d + \frac{1}{2} \log\det S(x)$

A.1 Affine Gaussian marginal and likelihood score

Proposition 1 (exact affine-Gaussian marginal, restated). At fixed x, suppose $\mathcal{A}(x, \psi)$ is affine in ψ, $q = \mathcal{N}(\mu, \Sigma_\psi)$, and $e \sim \mathcal{N}(0, \Sigma_e)$ is independent of ψ. Then $p_{\mathrm{marg}}(y \mid x) = \mathcal{N}(y; \bar{y}(x), S(x))$, with moments given by (4). For nonlinear dependence on ψ, (4) gives the exact mean and covariance of the linearized forward model obtained by dropping r_x in (3).

Proof. At fixed x, the measurement is an affine function of the independent Gaussian parameters and noise, hence Gaussian. Its mean and covariance are $\bar{y}(x)$ and $S(x)$ in (4), which gives the stated marginal.

For nonlinear dependence on ψ, dropping the remainder in (3) gives

$$y = \mathcal{A}(x, \psi_0) + J_\psi(x)(\psi - \psi_0) + e. \tag{8}$$

This response is affine in ψ, with offset $\mathcal{A}(x, \psi_0) - J_\psi(x)\psi_0$. The same moment calculation gives (4). At $\psi_0 = \mu$, Proposition 3 bounds the mean and covariance omitted by dropping the remainder. □

Likelihood score. Differentiating the normalized UAL energy in (5) and negating its derivative gives the likelihood score, with q and ψ_0 held fixed. Suppose that $\bar{y}(x)$ and $S(x)$ are differentiable. Write $\partial_i = \partial/\partial x_i$ and $w = S^{-1}d$, with $d = y - \bar{y}$. The required identities are

$$\begin{aligned} \partial_i d &= -\partial_i \bar{y}, \\ \partial_i S^{-1} &= -S^{-1}(\partial_i S)S^{-1}, \\ \partial_i \log\det S &= \mathrm{tr}(S^{-1}\partial_i S). \end{aligned} \tag{9}$$

Applying them to (5) gives

$$\partial_i \ell_{\mathrm{UAL}} = -(\partial_i \bar{y})^\mathsf{T} w - \frac{1}{2} w^\mathsf{T}(\partial_i S)w + \frac{1}{2}\,\mathrm{tr}(S^{-1}\partial_i S). \tag{10}$$

Negating this expression gives (6). When the covariance depends on the signal, the score accounts for changes in both residual weighting and Gaussian normalization. Holding the covariance fixed leaves the mean-response term.

A.2 Factorized evaluation of UAL

The UAL covariance is $S = \Sigma_e + UU^\mathsf{T}$, with U and G defined in Section 3.2. A single factorization of the $r \times r$ matrix G provides both $S^{-1}d$ and the determinant contribution $\log\det G$.

The Woodbury identity and matrix determinant lemma give

$$S^{-1} = \Sigma_e^{-1} - \Sigma_e^{-1} U G^{-1} U^\mathsf{T} \Sigma_e^{-1}, \tag{11}$$

$$\log\det S = \log\det \Sigma_e + \log\det G. \tag{12}$$

For $d = y - \bar{y}$, solve the reduced system

$$G a_{\mathrm{red}} = U^\mathsf{T}\Sigma_e^{-1} d, \qquad w = \Sigma_e^{-1}(d - U a_{\mathrm{red}}) = S^{-1}d. \tag{13}$$

In the equivalent local Gaussian model $d = Ua + e$, with $a \sim \mathcal{N}(0, I_r)$ independent of e, the solution a_{red} is the conditional mean of a given x and y. Thus $U a_{\mathrm{red}}$ is the conditional mean of the nuisance contribution to the residual.

With $\|v\|^2_{\Sigma_e^{-1}} = v^\mathsf{T}\Sigma_e^{-1}v$, the quadratic can be evaluated as a sum of nonnegative terms, avoiding cancellation between large quadratic terms:

$$\begin{aligned}\|d - Ua_{\mathrm{red}}\|^2_{\Sigma_e^{-1}} + \|a_{\mathrm{red}}\|^2 &= d^\mathsf{T}\Sigma_e^{-1}d - 2a_{\mathrm{red}}^\mathsf{T}U^\mathsf{T}\Sigma_e^{-1}d + a_{\mathrm{red}}^\mathsf{T}Ga_{\mathrm{red}} \\ &= d^\mathsf{T}\Sigma_e^{-1}d - a_{\mathrm{red}}^\mathsf{T}Ga_{\mathrm{red}} = d^\mathsf{T}S^{-1}d.\end{aligned} \tag{14}$$

The second equality uses (13); the last uses the Woodbury identity. Substitution into (5) gives

$$\begin{aligned}\ell_{\mathrm{UAL}} = \tfrac{1}{2}\|d - Ua_{\mathrm{red}}\|^2_{\Sigma_e^{-1}} + \tfrac{1}{2}\|a_{\mathrm{red}}\|^2 \\ + \tfrac{1}{2}\log\det\Sigma_e + \tfrac{1}{2}\log\det G + \tfrac{m}{2}\log(2\pi).\end{aligned} \tag{15}$$

For fixed x, a_{red} minimizes $\frac{1}{2}\|d - Ua\|^2_{\Sigma_e^{-1}} + \frac{1}{2}\|a\|^2$, the sum of the observation-noise penalty and the standard Gaussian prior penalty. Equation (15) therefore expresses the marginal energy as this penalized profile energy plus $\frac{1}{2}\log\det G$ and terms independent of x. The log determinant distinguishes marginalizing the local nuisance coordinates from optimizing them. Table 14 compares this penalized profile energy with UAL under the same sampler and nuisance model.

Computational implementation. Since $\Sigma_e \succ 0$, $G \succ 0$ even when Σ_ψ is singular. A Cholesky factorization $G = L_G L_G^\mathsf{T}$ evaluates (13) by triangular solves and gives $\log\det G = 2\sum_{i=1}^r \log(L_G)_{ii}$. For $r = 1$, the solve is scalar division. In our experiments, Σ_e is a scaled identity, so applying Σ_e^{-1} is a scalar division. For a general Σ_e, solve $\Sigma_e z = v$. For motion deblurring, DAPS and PnP-DM compute G and $S^{-1}d$ in Fourier coordinates without storing U. Explicit columnwise construction is a general alternative requiring r applications of J_ψ.

Algorithm 1 evaluates the same UAL Gaussian likelihood through its factorized energy. Covariance truncation introduces a separate approximation; Corollary 2 bounds its likelihood error when the mean is unchanged and the discarded covariance is positive semidefinite.

Algorithm 1 UAL data term

Require: signal argument x, measurement y, reference ψ_0, mean μ of q, factor F with $\Sigma_\psi = FF^\mathsf{T}$, noise covariance Σ_e, forward model $\mathcal{A}$ and its parameter Jacobian J_ψ at ψ_0

1: $\bar{y} \leftarrow \mathcal{A}(x, \psi_0) + J_\psi(x)(\mu - \psi_0)$; $\quad d \leftarrow y - \bar{y}$; $\quad U \leftarrow J_\psi(x)F$
2: Form $G \leftarrow I_r + U^\mathsf{T}\Sigma_e^{-1}U$ and compute its Cholesky factor $G = L_G L_G^\mathsf{T}$.
3: Solve for a_{red} in $L_G L_G^\mathsf{T} a_{\mathrm{red}} = U^\mathsf{T}\Sigma_e^{-1}d$ by forward and backward substitution.
4: $w \leftarrow \Sigma_e^{-1}(d - Ua_{\mathrm{red}})$ $\{w = S^{-1}d\}$
5: $\ell_{\mathrm{UAL}} \leftarrow \frac{1}{2}\|d - Ua_{\mathrm{red}}\|^2_{\Sigma_e^{-1}} + \frac{1}{2}\|a_{\mathrm{red}}\|^2 + \frac{1}{2}\log\det\Sigma_e + \frac{1}{2}\log\det G + \frac{m}{2}\log(2\pi)$

Ensure: energy ℓ_{UAL} and likelihood score by (6), differentiating the mean and covariance with μ, F, and ψ_0 fixed

Remark 1 (Scalar covariance inflation). Fix x and compare Gaussian likelihoods with the same mean $\bar{y}(x)$. For $v > 0$, the equal-mean Gaussian KL divergence is

$$\begin{aligned}K(v) &= D_{\mathrm{KL}}(\mathcal{N}(\bar{y}, S)\,\|\,\mathcal{N}(\bar{y}, v\Sigma_e)) \\ &= \tfrac{1}{2}\left[\frac{m + \operatorname{tr} G - r}{v} - m + m\log v - \log\det G\right],\end{aligned} \tag{16}$$

where $S = \Sigma_e + UU^\mathsf{T}$ and $G = I_r + U^\mathsf{T}\Sigma_e^{-1}U$. Differentiation gives the unique minimizer and its remaining divergence:

$$\begin{aligned}v_\star &= 1 + \frac{\operatorname{tr} G - r}{m}, \\ K(v_\star) &= \tfrac{1}{2}\left[m\log v_\star - \log\det G\right] \geq 0.\end{aligned} \tag{17}$$

The optimal scalar matches the average variance after whitening by the observation-noise covariance. At fixed whitened covariance trace and unchanged $\bar{y}(x)$, anisotropy in the whitened covariance leaves a positive minimum divergence; scalar inflation cannot reproduce its directional weighting. Concavity of the logarithm, applied to the eigenvalues of $I_m + \Sigma_e^{-1/2}UU^\mathsf{T}\Sigma_e^{-1/2}$, gives the inequality, with equality exactly when $UU^\mathsf{T} = \beta\Sigma_e$ for some $\beta \geq 0$. Under the affine-Gaussian assumptions of Proposition 1, UAL is the exact marginalized likelihood in this comparison.

A.3 LIKELIHOOD AND POSTERIOR APPROXIMATION ERROR

The error from nuisance linearization is controlled by the response it omits, measured in observation-noise units.

Proof of Proposition 2. Fix x, and denote the affine response in (3) by $\widetilde{\mathcal{A}}_x(\psi)$. Let P_x and Q_x denote the joint distributions of (ψ, y) under the original and linearized forward models, respectively:

$$\begin{aligned} P_x(d\psi, dy) &= q(d\psi)\,\mathcal{N}(dy; \mathcal{A}(x,\psi), \Sigma_e), \\ Q_x(d\psi, dy) &= q(d\psi)\,\mathcal{N}(dy; \widetilde{\mathcal{A}}_x(\psi), \Sigma_e). \end{aligned} \tag{18}$$

They share the nuisance distribution q and the conditional noise covariance. The equal-covariance Gaussian KL formula gives

$$D_{\mathrm{KL}}(P_x \| Q_x) = \frac{1}{2}\mathbb{E}_q\left[r_x(\psi)^{\mathsf{T}}\Sigma_e^{-1} r_x(\psi)\right]. \tag{19}$$

Let p_{aff} be the measurement marginal of Q_x. Marginalizing ψ cannot increase divergence, so

$$D_{\mathrm{KL}}(p_{\mathrm{marg}}(\cdot \mid x) \| p_{\mathrm{aff}}(\cdot \mid x)) \le D_{\mathrm{KL}}(P_x \| Q_x). \tag{20}$$

For Gaussian q, Proposition 1 identifies p_{aff} with p_{UAL}. For non-Gaussian q, this argument bounds the divergence to the affine mixture, before Gaussian moment matching. □

A small expected squared remainder in observation-noise units therefore gives a small likelihood divergence.

Proposition 2 compares measurement likelihoods at a fixed signal. The next result connects that approximation to the inference target: with the same signal prior, the prior-averaged likelihood error also bounds the posterior error averaged over measurements.

Corollary 1 (Averaged posterior error)**.** *Use the same normalized signal prior p_0 in the exact marginal and UAL models. Fix the nuisance model $q = \mathcal{N}(\mu, \Sigma_\psi)$ and reference ψ_0 independently of the evaluated observation. The UAL posterior is $p_{\mathrm{UAL}}(x \mid y) \propto p_0(x) p_{\mathrm{UAL}}(y \mid x; \psi_0, q)$. If $\mathbb{E}_{x\sim p_0, \psi\sim q}[r_x(\psi)^{\mathsf{T}}\Sigma_e^{-1} r_x(\psi)] < \infty$, then*

$$\begin{aligned} &\mathbb{E}_{y\sim p_{\mathrm{marg}}(y)} D_{\mathrm{KL}}(p_{\mathrm{marg}}(\cdot \mid y) \,\|\, p_{\mathrm{UAL}}(\cdot \mid y)) \\ &\qquad \le \tfrac{1}{2}\mathbb{E}_{x\sim p_0, \psi\sim q}\left[r_x(\psi)^{\mathsf{T}}\Sigma_e^{-1} r_x(\psi)\right]. \end{aligned} \tag{21}$$

Here $p_{\mathrm{marg}}(y)$ is the measurement distribution obtained by drawing $x \sim p_0$ and then $y \sim p_{\mathrm{marg}}(\cdot \mid x)$.

Proof. Let $P(x,y) = p_0(x) p_{\mathrm{marg}}(y \mid x)$ and $Q(x,y) = p_0(x) p_{\mathrm{UAL}}(y \mid x)$. Because the prior is common, the KL chain rule first gives

$$D_{\mathrm{KL}}(P\|Q) = \mathbb{E}_{x\sim p_0} D_{\mathrm{KL}}(p_{\mathrm{marg}}(\cdot \mid x) \| p_{\mathrm{UAL}}(\cdot \mid x)). \tag{22}$$

Let $p_{\mathrm{UAL}}(y)$ be the measurement marginal of Q. Factoring the same joint laws in the other order gives

$$\begin{aligned} D_{\mathrm{KL}}(P\|Q) = {} & D_{\mathrm{KL}}(p_{\mathrm{marg}}(y) \| p_{\mathrm{UAL}}(y)) \\ & + \mathbb{E}_{y\sim p_{\mathrm{marg}}(y)} D_{\mathrm{KL}}(p_{\mathrm{marg}}(\cdot \mid y) \| p_{\mathrm{UAL}}(\cdot \mid y)). \end{aligned} \tag{23}$$

The first term is nonnegative. Thus average posterior divergence is at most joint divergence, which (22) and Proposition 2 bound by the right-hand side of (21).

□

A quadratic remainder envelope. Under the hypotheses of Proposition 2, suppose $L \ge 0$ and $\|r_x(\psi)\| \le L\|\psi - \psi_0\|^2/2$ for q-almost every ψ. This curvature bound controls the likelihood error through the fourth nuisance moment:

$$\frac{1}{2}\mathbb{E}_q[r_x^{\mathsf{T}}\Sigma_e^{-1} r_x] \le \frac{L^2}{8\lambda_{\min}(\Sigma_e)}\mathbb{E}_q\|\psi - \psi_0\|^4. \tag{24}$$

This follows from $r_x(\psi)^\mathsf{T}\Sigma_e^{-1}r_x(\psi) \le \|r_x(\psi)\|^2/\lambda_{\min}(\Sigma_e)$ and the remainder envelope. The fourth moment is finite for Gaussian q. At $\psi_0 = \mu$ with scalar $\Sigma_\psi = c$ and $\Sigma_e = \sigma^2 I$, $\mathbb{E}_q(\psi-\mu)^4 = 3c^2$ gives the bound $3L^2c^2/(8\sigma^2)$. At fixed curvature bound and observation-noise variance, this bound grows quadratically with the nuisance variance.

For the corresponding averaged posterior bound, allow the curvature bound to depend on the signal. Suppose additionally that $\|r_x(\psi)\| \le L(x)\|\psi-\psi_0\|^2/2$ for $p_0 \otimes q$-almost every (x,ψ), where $L(x) \ge 0$ is measurable and $\mathbb{E}_{x\sim p_0}[L(x)^2\mathbb{E}_q\|\psi-\psi_0\|^4] < \infty$. Then the right-hand side of (21) is at most

$$\frac{\mathbb{E}_{x\sim p_0}\left[L(x)^2\mathbb{E}_q\|\psi-\psi_0\|^4\right]}{8\lambda_{\min}(\Sigma_e)}. \tag{25}$$

Applying (24) inside the prior expectation gives (25).

A.4 Moment errors from nuisance linearization

Nuisance curvature can change both the marginal mean and covariance, even when the first-order mean correction vanishes.

Proposition 3 (Moment errors from linearization)**.** *At fixed x and $\psi_0 = \mu$, let $J = J_\psi(x)$ and suppose $L \ge 0$ and $\|r_x(\psi)\| \le L\|\psi-\mu\|^2/2$ on the support of q, and the moments below are finite. Writing $\delta = \psi - \mu$, the mean omitted by (4) is $b_r = \mathbb{E}r_x$ with $\|b_r\| \le L\,\mathrm{tr}\,\Sigma_\psi/2$. Let $\Delta S = \mathrm{Cov}(y \mid x) - S(x)$, where the conditional covariance retains the nonlinear remainder. Then*

$$\max\{|\,\mathrm{tr}\,\Delta S|, \|\Delta S\|_{\mathrm{op}}\} \le \frac{L^2}{4}\mathbb{E}\|\delta\|^4 + L\,\mathbb{E}\left[\|\delta\|^2\|J\delta\|\right]. \tag{26}$$

For scalar Gaussian δ with variance c, writing $J = j$, the right-hand side is $3L^2c^2/4 + \sqrt{8/\pi}\,L|j|c^{3/2}$.

Proof. Write $\delta = \psi - \mu$. The exact response is $y = \mathcal{A}(x,\mu) + J\delta + r_x(\psi) + e$. Its omitted mean is $b_r = \mathbb{E}r_x(\psi)$, so $\|b_r\| \le \mathbb{E}\|r_x(\psi)\| \le L\mathbb{E}\|\delta\|^2/2 = L\,\mathrm{tr}\,\Sigma_\psi/2$. Since $\mathbb{E}\delta = 0$ and the observation noise is independent,

$$\Delta S = \mathrm{Cov}(r_x(\psi)) + \mathbb{E}[J\delta r_x(\psi)^\mathsf{T}] + \mathbb{E}[r_x(\psi)\delta^\mathsf{T}J^\mathsf{T}]. \tag{27}$$

The covariance term is positive semidefinite, with $\|\mathrm{Cov}(r_x(\psi))\|_{\mathrm{op}} \le \mathrm{tr}\,\mathrm{Cov}(r_x(\psi)) \le \mathbb{E}\|r_x(\psi)\|^2$. Each cross term has both absolute trace and operator norm at most $\mathbb{E}[\|J\delta\|\|r_x(\psi)\|]$. Thus $|\,\mathrm{tr}\,\Delta S|$ and $\|\Delta S\|_{\mathrm{op}}$ are each bounded by $\mathbb{E}\|r_x(\psi)\|^2 + 2\mathbb{E}[\|J\delta\|\|r_x(\psi)\|]$. Substituting the remainder bound gives (26). For scalar Gaussian δ, $\mathbb{E}\delta^4 = 3c^2$ and $\mathbb{E}|\delta|^3 = \sqrt{8/\pi}\,c^{3/2}$ give the stated constants. □

For scalar Gaussian variance c, the omitted mean is $O(c)$ and the covariance-error bound is $O(c^{3/2} + c^2)$ at fixed curvature and sensitivity. The nonlinear remainder can increase or decrease variance along a measurement direction through its cross terms with the linear response.

Remark 2 (Phase responses: pulse-echo ultrasound). A phase response can have a nonlinear mean shift even when the first-order mean correction vanishes. For a complex sample s perturbed to $se^{i\theta}$ with $\theta \sim \mathcal{N}(0,\sigma_\theta^2)$,

$$\mathbb{E}[se^{i\theta}] = se^{-\sigma_\theta^2/2}. \tag{28}$$

Zero-mean Gaussian phase uncertainty attenuates the mean response by $e^{-\sigma_\theta^2/2}$; the first-order mean s omits this attenuation.

A.5 Likelihood error from covariance truncation

We next keep the UAL mean fixed and truncate only its covariance. For a positive-semidefinite discarded covariance, the following corollary bounds the divergence from UAL to the Gaussian likelihood with retained covariance S_K. The error is small when the discarded covariance is small after whitening by S_K.

Corollary 2 (Covariance-truncation divergence)**.** *Fix x and retain the UAL mean $\bar{y}(x)$. Let $C_{y,K} \in \mathbb{R}^{m\times m}$ be the retained nuisance-induced measurement covariance, satisfying $0 \preceq C_{y,K} \preceq J_\psi(x)\Sigma_\psi J_\psi(x)^\mathsf{T}$. The truncated likelihood is $p_{\mathrm{UAL},K} = \mathcal{N}(\bar{y}, S_K)$, with $S_K = \Sigma_e + C_{y,K}$. Its discarded covariance is $S - S_K$. Writing this discarded covariance in whitened coordinates gives $E_K = S_K^{-1/2}(S - S_K)S_K^{-1/2}$. Then*

$$D_{\mathrm{KL}}(p_{\mathrm{UAL}} \| p_{\mathrm{UAL},K}) = \frac{1}{2}[\operatorname{tr} E_K - \log\det(I + E_K)] \le \frac{1}{4}\|E_K\|_F^2. \tag{29}$$

This includes truncating the covariance factor, provided the discarded covariance is positive semidefinite and the mean is unchanged.

Proof. The equal-mean Gaussian KL formula gives $\frac{1}{2}[\operatorname{tr}(S_K^{-1}S) - m - \log\det(S_K^{-1}S)]$. The matrix $S_K^{-1/2}SS_K^{-1/2} = I + E_K$ has the same eigenvalues as $S_K^{-1}S$, which gives the identity. Let $\epsilon_i \ge 0$ be the eigenvalues of E_K. Applying $0 \le a - \log(1+a) \le a^2/2$ to each eigenvalue gives

$$\frac{1}{2}\sum_i [\epsilon_i - \log(1+\epsilon_i)] \le \frac{1}{4}\sum_i \epsilon_i^2 = \frac{1}{4}\|E_K\|_F^2. \tag{30}$$

□

For an unchanged mean and positive-semidefinite covariance discard, the retained covariance determines which discarded directions contribute most to likelihood divergence.

B Fixed-sensitivity reconstruction analysis

Under matched nuisance moments, the UAL linear estimator minimizes average squared error among linear estimators; at a fixed nuisance realization its error can exceed the point estimator's. We quantify this tradeoff and the reconstruction cost of covariance truncation.

B.1 Fixed-sensitivity model and estimator definitions

A fixed-sensitivity linear model lets us compute both filters' risks exactly and identify when covariance weighting helps. Holding the forward sensitivity and signal moments fixed isolates the reconstruction effect of changing the measurement-error covariance.

Assumption 1 (Fixed-sensitivity linear model)**.** *Let $A_0 \in \mathbb{R}^{m\times n}$ be a fixed reference operator and $J \in \mathbb{R}^{m\times P}$ a fixed nuisance sensitivity. Set $\psi_0 = \mu$ and consider*

$$y = A_0 x + J(\psi - \mu) + e.$$

The signal is centered, with $\mathbb{E}x = 0$ and $\mathrm{Cov}x = \Sigma_x$, and is independent of ψ and e. The uncertainty model q has mean μ and covariance Σ_ψ. The noise is centered, independent of ψ, and has covariance Σ_e. Thus the effective noise $\varepsilon = J(\psi - \mu) + e$ has mean zero and covariance $S = \Sigma_e + J\Sigma_\psi J^\mathsf{T}$. For a response linear in the signal, $\mathcal{A}(x,\psi) = A(\psi)x$, take $A_0 = A(\mu)$ and hold the local nuisance sensitivity J fixed as x varies.

The two estimators are $\widehat{x} = W_0 y$ and $\widehat{x} = W_S y$, with

$$M_0 = A_0\Sigma_x A_0^\mathsf{T} + \Sigma_e, \quad M_S = M_0 + J\Sigma_\psi J^\mathsf{T}, \qquad W_0 = \Sigma_x A_0^\mathsf{T} M_0^{-1}, \quad W_S = \Sigma_x A_0^\mathsf{T} M_S^{-1}. \tag{31}$$

The matrices M_0 and M_S are the measurement covariances under the point likelihood and UAL, respectively; both estimators use the same signal covariance Σ_x. At $\psi_0 = \mu$, the mean correction vanishes. Define $\mathcal{R}(W) = \mathbb{E}\|Wy - x\|^2$, averaging over the signal, nuisance parameters, and observation noise.

These moment assumptions suffice for the linear-risk comparisons. Gaussianity is needed only where the linear estimator is identified with the posterior mean. Conditional risk fixes the realized nuisance value $\psi^\star$ and averages over x and e.

Remark 3 (Nonzero prior mean or reference offset). Let $m_x = \mathbb{E}x$, $\delta_\mu = \mu - \psi_0$, and consider the fixed-matrix model $y = A_{\mathrm{ref}}x + J(\psi - \psi_0) + e$, where $A_{\mathrm{ref}} = A(\psi_0)$. The estimator Wy has mean error $a_W = (WA_{\mathrm{ref}} - I)m_x + WJ\delta_\mu$. Relative to the centered model with the same matrices, the average risk increases by $\|a_W\|^2$. Conditional on $\delta = \psi^\star - \mu$, the risk also includes the cross term $2a_W^\mathsf{T} WJ\delta$. The centered affine estimator

$$\widehat{x} = m_x + W[y - A_{\mathrm{ref}}m_x - J\delta_\mu]$$

removes these terms. The shift $J\delta_\mu$ is the mean correction in (4). The results below use fixed matrices with $m_x = 0$ and $\delta_\mu = 0$.

B.2 Average risk and the optimal linear estimator

Under the matched-moment model, the UAL filter uniquely minimizes average squared error among linear filters. The squared-norm term in (33) gives the excess risk of using a different filter.

Proposition 4 (Average reconstruction error (LMMSE optimality)). *Under Assumption 1, W_S minimizes $\mathcal{R}(W)$ among linear estimators, so $\mathcal{R}(W_S) \le \mathcal{R}(W_0)$. When x, ψ, and e are Gaussian, $W_S y$ is also the posterior mean.*

Proof. Under Assumption 1, write $Wy - x = (WA_0 - I)x + W\varepsilon$. Independence and zero means eliminate the cross term. Expanding the remaining traces gives, for any linear filter W,

$$\mathcal{R}(W) = \operatorname{tr}\Sigma_x - 2\operatorname{tr}(WA_0\Sigma_x) + \operatorname{tr}\left(WM_SW^\mathsf{T}\right). \tag{32}$$

Since $W_S M_S = \Sigma_x A_0^\mathsf{T}$, completing the square in (32) gives

$$\begin{aligned} \mathcal{R}(W) &= \mathcal{R}(W_S) + \|(W - W_S)M_S^{1/2}\|_F^2, \\ \mathcal{R}(W_S) &= \operatorname{tr}\Sigma_x - \operatorname{tr}(\Sigma_x A_0^\mathsf{T} M_S^{-1} A_0 \Sigma_x). \end{aligned} \tag{33}$$

The squared norm is the excess risk above W_S. Because $M_S \succ 0$, it vanishes only when $W = W_S$. The minimizer is therefore unique even if Σ_x is singular.

When x, ψ, and e are Gaussian, (x, y) is jointly Gaussian with $\mathrm{Cov}(x, y) = \Sigma_x A_0^\mathsf{T}$ and $\mathrm{Cov}(y) = M_S$. Hence

$$\mathbb{E}[x \mid y] = \Sigma_x A_0^\mathsf{T} M_S^{-1} y = W_S y.$$

□

The results below characterize the gain of W_S and the crossover at a fixed nuisance realization.

B.3 Average gain and its normalization

The gain is the point estimator's nuisance-error contribution multiplied by the fraction removed by covariance weighting.

Proposition 5 (Average gain and its normalization). *Under Assumption 1, use $\Sigma_\psi = FF^\mathsf{T}$ and $U = JF$ as in Section 3.2, with J now fixed. Define the dimensionless Gram matrix $\Lambda = U^\mathsf{T} M_0^{-1} U$. The matrix Λ uses the predictive covariance M_0; the likelihood-evaluation matrix G uses the observation-noise covariance Σ_e. Then*

$$\mathcal{R}(W_0) - \mathcal{R}(W_S) = \operatorname{tr}\left(W_0 U \Lambda (I + \Lambda)^{-1} U^\mathsf{T} W_0^\mathsf{T}\right) \ge 0. \tag{34}$$

Let $H_0 = \operatorname{tr}(W_0 J \Sigma_\psi J^\mathsf{T} W_0^\mathsf{T})$ be the point estimator's nuisance-error contribution. When $H_0 > 0$, define $\kappa_0 = [\mathcal{R}(W_0) - \mathcal{R}(W_S)]/H_0$. For one nuisance parameter, $\Sigma_\psi = c$ and $J = j$,

$$\kappa_0 = \frac{\lambda}{1 + \lambda}, \qquad \lambda = c\, j^\mathsf{T} M_0^{-1} j. \tag{35}$$

The net average gain recovers a small fraction of H_0 at small λ and nearly all of H_0 at large λ. For multiple parameters, κ_0 is a nonnegative weighted average of $\lambda_i/(1 + \lambda_i)$ over the eigenvalues of Λ.

Scaling up the nuisance covariance by a common positive factor increases the fraction of the point estimator's nuisance error removed by covariance weighting. The absolute gain is $H_0\kappa_0$.

If $H_0 = 0$, the estimators coincide and κ_0 is undefined.

Eigenvalue form. Let $\Lambda v_i = \lambda_i v_i$ with orthonormal v_i, and define $\omega_i = \|W_0 U v_i\|^2$. For $H_0 > 0$,

$$\kappa_0 = \frac{\sum_i \omega_i \lambda_i/(1+\lambda_i)}{\sum_i \omega_i}, \qquad \frac{\lambda_{\min}}{1+\lambda_{\min}} \le \kappa_0 \le \frac{\lambda_{\max}}{1+\lambda_{\max}}.$$

The weights measure how much each factor direction contributes to the point estimator's nuisance error.

Proof. Write $\mathcal{R}_0(W) = \mathbb{E}_{x,e}\|W(A_0 x + e) - x\|^2$ for the risk when the nuisance displacement is zero. For the estimator W_0, the average risk separates into this zero-displacement risk and its nuisance contribution:

$$\mathcal{R}_0(W_0) = \operatorname{tr}\Sigma_x - \operatorname{tr}(\Sigma_x A_0^\mathsf{T} M_0^{-1} A_0 \Sigma_x), \qquad \mathcal{R}(W_0) = \mathcal{R}_0(W_0) + H_0,$$

where $H_0 = \|W_0 U\|_F^2$. The Woodbury identity, expressed through F and valid for singular Σ_ψ, gives

$$M_S^{-1} = M_0^{-1} - M_0^{-1} U (I+\Lambda)^{-1} U^\mathsf{T} M_0^{-1}.$$

Substitution into (33) yields

$$\mathcal{R}(W_S) = \mathcal{R}_0(W_0) + \operatorname{tr}\big(W_0 U (I+\Lambda)^{-1} U^\mathsf{T} W_0^\mathsf{T}\big).$$

Subtracting this expression from $\mathcal{R}(W_0)$ and using $I - (I+\Lambda)^{-1} = \Lambda(I+\Lambda)^{-1} \succeq 0$ proves (34).

For one parameter, divide by $H_0 = c\|W_0 j\|^2 > 0$ to obtain (35). For multiple parameters, express both traces in the eigenbasis of Λ: their terms are $\omega_i\lambda_i/(1+\lambda_i)$ and ω_i. If $H_0 = 0$, then $W_0 U = 0$ and the Woodbury formula gives $W_S = W_0$.

□

B.4 Conditional reconstruction error and the crossover

At a fixed nuisance displacement, covariance weighting changes the nuisance-error contribution and incurs a filter-change cost.

Proposition 6 (Multivariate conditional reconstruction error)**.** *Under Assumption 1, fix $\psi^\star = \mu + \delta$ and write $b = J\delta$ for the realized nuisance term. Let $\mathcal{R}_\delta(W)$ be the mean squared error averaged over x and e with δ fixed. For any linear W,*

$$\begin{aligned} \mathcal{R}_\delta(W) &= \operatorname{tr}\big((WA_0 - I)\Sigma_x (WA_0 - I)^\mathsf{T}\big) + \operatorname{tr}(W\Sigma_e W^\mathsf{T}) + \|Wb\|^2, \\ \mathcal{R}_\delta(W_0) - \mathcal{R}_\delta(W_S) &= \underbrace{\|W_0 J\delta\|^2 - \|W_S J\delta\|^2}_{\textit{difference in nuisance-error contributions}} - \underbrace{\tau_0(\Sigma_\psi)}_{\textit{filter-change cost}}, \\ \tau_0(\Sigma_\psi) &= \operatorname{tr}\big((W_S - W_0) M_0 (W_S - W_0)^\mathsf{T}\big) \ge 0. \end{aligned} \tag{36}$$

Moreover, $\tau_0(\Sigma_\psi) > 0$ if and only if $W_0 J \Sigma_\psi \neq 0$.

Proof. For a fixed $b = J\delta$, write

$$Wy - x = (WA_0 - I)x + We + Wb.$$

The first two terms are centered and independent, while Wb is deterministic. Their squared norms add in expectation, giving the first line of (36). Completing the square in the part independent of b gives

$$\mathcal{R}_\delta(W) = \mathcal{R}_0(W_0) + \|(W - W_0)M_0^{1/2}\|_F^2 + \|Wb\|^2.$$

Subtract the expressions for W_0 and W_S to obtain the conditional gain. The term $\tau_0(\Sigma_\psi)$ is the cost of changing the filter when the nuisance contribution is absent. Since $M_0 \succ 0$, it vanishes exactly when $W_S = W_0$.

Using $J\Sigma_\psi J^\mathsf{T} = UU^\mathsf{T}$, the resolvent identity gives

$$W_S - W_0 = -W_0 U U^\mathsf{T} M_S^{-1}.$$

Thus $W_S = W_0$ exactly when $W_0 U U^\mathsf{T} = 0$, equivalently $W_0 U = 0$, because $W_0 U U^\mathsf{T} W_0^\mathsf{T} = (W_0 U)(W_0 U)^\mathsf{T}$. The condition $W_0 U = 0$ implies $W_0 J \Sigma_\psi = W_0 U F^\mathsf{T} = 0$; conversely, $W_0 J \Sigma_\psi = 0$ implies $W_0 U U^\mathsf{T} = 0$. This proves the equality condition even for singular Σ_ψ.

□

For one nuisance parameter, write $\Sigma_\psi = c$ and $J = j$, with λ as defined in (35).

Corollary 3 (Error for a fixed nuisance parameter value)**.** *Under Assumption 1 with one nuisance parameter, fix $\psi^\star = \mu + \delta$. Let $\mathcal{R}_\delta(W)$ be the mean squared error averaged over x and e with δ held fixed. The scalar identities $\|W_S j\|^2 = \|W_0 j\|^2/(1+\lambda)^2$ and $\tau_0(c) = c\lambda\|W_0 j\|^2/(1+\lambda)^2$ give the conditional difference*

$$\mathcal{R}_\delta(W_0) - \mathcal{R}_\delta(W_S) = \frac{\lambda\|W_0 j\|^2}{(1+\lambda)^2}\big[(\lambda+2)\delta^2 - c\big]. \tag{37}$$

For distinct estimators the prefactor is positive. Their conditional errors agree at the threshold

$$|\delta^\star|^2 = \frac{c}{\lambda+2}. \tag{38}$$

Proof. For one parameter, Sherman–Morrison gives

$$W_S - W_0 = -\frac{c}{1+\lambda} W_0 j\, j^\mathsf{T} M_0^{-1}, \qquad W_S j = \frac{W_0 j}{1+\lambda}.$$

Consequently,

$$\tau_0(c) = \frac{c\lambda}{(1+\lambda)^2}\|W_0 j\|^2, \qquad \|W_0 j\delta\|^2 - \|W_S j\delta\|^2 = \frac{\lambda(\lambda+2)\delta^2}{(1+\lambda)^2}\|W_0 j\|^2.$$

Subtracting these terms proves (37). For distinct estimators the common prefactor is positive; setting the gain to zero gives $|\delta^\star|^2 = c/(\lambda+2)$. □

For one nuisance parameter and distinct estimators, the point-likelihood estimator has lower conditional error when $|\delta| < |\delta^\star|$, and UAL when $|\delta| > |\delta^\star|$. Relative to the nuisance standard deviation $\sqrt{c}$, the crossover distance is $(\lambda+2)^{-1/2}$, which decreases as λ increases.

Remark 4 (Multivariate conditional gain). For a fixed direction v, let $a_v = \|W_0 Jv\|^2 - \|W_S Jv\|^2$. Along $\delta = sv$, the gain is

$$\mathcal{R}_{sv}(W_0) - \mathcal{R}_{sv}(W_S) = s^2 a_v - \tau_0(\Sigma_\psi).$$

If $a_v > 0$, the crossover occurs at $|s| = \sqrt{\tau_0(\Sigma_\psi)/a_v}$. If $a_v \le 0$, increasing $|s|$ does not produce a positive gain. The direction of the nuisance error therefore determines whether increasing its magnitude can favor UAL.

Mismatch between assumed and generating moments. The scalar comparison also quantifies how inaccurate nuisance moments affect average reconstruction error. Keep $A_0, J, \Sigma_x, \Sigma_e$ fixed, and draw $\psi^\star$ independently of x and e from a distribution that differs from q. For one nuisance parameter, averaging (37) replaces δ^2 by its second moment about the mean μ,

$$\mathbb{E}[(\psi^\star - \mu)^2] = \mathrm{Var}(\psi^\star) + (\mathbb{E}\psi^\star - \mu)^2.$$

For assumed variance $c > 0$ and distinct estimators, average gain is positive exactly when

$$\mathrm{Var}(\psi^\star) + (\mathbb{E}\psi^\star - \mu)^2 > \frac{c}{\lambda+2}. \tag{39}$$

Below this threshold, the filter-change cost exceeds the reduction in nuisance-induced error, so the point estimator has lower average risk. The risks agree at the threshold. A sufficient condition for positive average gain is $0 < c \le 2\mathbb{E}[(\psi^\star - \mu)^2]$: distinct filters imply $\lambda > 0$, so $c/(\lambda+2) < c/2$. Thus the assumed variance c can be as large as twice the generating second moment about the mean μ while preserving the average advantage in this scalar fixed-sensitivity model.

Define $\lambda^\star = \mathbb{E}[(\psi^\star - \mu)^2] j^\mathsf{T} M_0^{-1} j$. Compared with λ, $\lambda^\star$ uses the generating second moment about μ. The average gain is negative exactly when

$$\lambda^\star < 1 \quad \text{and} \quad c > \frac{2\mathbb{E}[(\psi^\star - \mu)^2]}{1 - \lambda^\star}. \tag{40}$$

For one nuisance parameter and distinct filters, using too large a variance c can reverse the average gain only when $\lambda^\star < 1$. For $\lambda^\star \ge 1$, the gain remains positive for every $c > 0$.

B.5 Reconstruction error from covariance truncation

Corollary 2 bounds Gaussian-likelihood divergence after covariance truncation. The next result bounds its reconstruction cost in the fixed-sensitivity linear model, using the discarded measurement-covariance modes.

Proposition 7 (Covariance truncation)**.** *Under Assumption 1, let $\Sigma_e = \sigma^2 I$ and choose a retained covariance $\sigma^2 I \preceq S_K \preceq S$. Write the discarded covariance as $D = S - S_K = \sum_i \nu_i u_i u_i^\mathsf{T}$, with orthonormal u_i and $\nu_i \geq 0$, and set $W_{S_K} = \Sigma_x A_0^\mathsf{T}(A_0 \Sigma_x A_0^\mathsf{T} + S_K)^{-1}$. Evaluate both filters under the full covariance S. Then*

$$0 \leq \mathcal{R}(W_{S_K}) - \mathcal{R}(W_S) \leq \|W_{S_K}\|_{\mathrm{op}}^2 \sum_i \frac{\nu_i^2}{\sigma^2 + \nu_i} \leq \frac{\|\Sigma_x A_0^\mathsf{T}\|_{\mathrm{op}}^2}{\sigma^4} \sum_i \nu_i. \tag{41}$$

When $S_K = S$, both sides are zero.

In the upper bound, a discarded variance contributes approximately ν_i^2/σ^2 when $\nu_i \ll \sigma^2$ and approximately ν_i when $\nu_i \gg \sigma^2$. The factor $\|W_{S_K}\|_{\mathrm{op}}^2$ accounts for amplification by the reconstruction filter. The bound also applies when the nuisance covariance is truncated: $S_K = \sigma^2 I + J \Sigma_{\psi,K} J^\mathsf{T}$ with $0 \preceq \Sigma_{\psi,K} \preceq \Sigma_\psi$.

Proof. Write $M_{S_K} = A_0 \Sigma_x A_0^\mathsf{T} + S_K$, so $M_S = M_{S_K} + D$. The completed square (33) gives

$$\mathcal{R}(W_{S_K}) - \mathcal{R}(W_S) = \|(W_{S_K} - W_S) M_S^{1/2}\|_F^2 \geq 0.$$

By the resolvent identity,

$$W_{S_K} - W_S = \Sigma_x A_0^\mathsf{T}(M_{S_K}^{-1} - M_S^{-1}) = W_{S_K} D M_S^{-1}.$$

Hence the excess risk is

$$\mathrm{tr}\big(W_{S_K} D M_S^{-1} D W_{S_K}^\mathsf{T}\big).$$

Since $M_{S_K} \succeq \sigma^2 I$, we have $M_S \succeq \sigma^2 I + D$. Inversion reverses the ordering and congruence by D preserves it:

$$D M_S^{-1} D \preceq D(\sigma^2 I + D)^{-1} D = \sum_i \frac{\nu_i^2}{\sigma^2 + \nu_i} u_i u_i^\mathsf{T} \preceq D.$$

For $X \succeq 0$, $\mathrm{tr}(WXW^\mathsf{T}) \leq \|W\|_{\mathrm{op}}^2 \,\mathrm{tr}\, X$. Applying this inequality gives the middle bound in (41). Finally,

$$\|W_{S_K}\|_{\mathrm{op}} \leq \frac{\|\Sigma_x A_0^\mathsf{T}\|_{\mathrm{op}}}{\sigma^2}, \qquad \frac{\nu_i^2}{\sigma^2 + \nu_i} \leq \nu_i,$$

which gives the last bound. □

C Task models and UAL implementations

We use the notation of Section 3.1: $\mathcal{A}$ is the forward model, Σ_e is the observation-noise covariance, and μ, Σ_ψ are the mean and covariance of the specified nuisance model q. For each application, we construct a measurement likelihood from the nuisance statistics and incorporate its data term into the published sampler update. Each UAL version retains the published data-fitting scales and step sizes. So that these step sizes apply unchanged, BlindDPS rescales the UAL image guidance to the published image-guidance norm before the denoiser Jacobian, and CT rescales the precision-weighted residual to the unweighted residual norm before backprojection. DiffPIR evaluates the covariance at the current prediction and holds it fixed within the data step, which keeps that step a quadratic subproblem.

For fixed $\Sigma_e \succ 0$, let $d = y - \bar{y}$ and define

$$Q(x) = d(x)^\mathsf{T} S(x)^{-1} d(x) + \log\det S(x) - \log\det \Sigma_e. \tag{42}$$

For measurement dimension m, $Q = 2\ell_{\mathrm{UAL}} - \log\det\Sigma_e - m\log(2\pi)$. Removing the noise-only constant from $2\ell_{\mathrm{UAL}}$ leaves its gradient unchanged. The square-root fidelity used below puts this energy on the scale of the sampler's residual-norm data term. Since $S \succeq \Sigma_e$, $Q \geq 0$, with the square-root derivative defined for $Q > 0$. The determinant identity also gives $Q = d^\mathsf{T} S^{-1} d + \log\det G$, evaluated with the reduced factorization in Section 3.2.

DPS-style guidance. The UAL score acts on a candidate clean signal. DPS-style samplers update the noisy diffusion state, so the denoiser Jacobian maps that score into their guidance direction (Chung et al., 2023b; Aali et al., 2025):

$$g_t = [D_{x_t}\widehat{x}_0(x_t)]^{\mathsf{T}}\, \nabla_x \log p_{\mathrm{UAL}}(y \mid x)|_{x=\widehat{x}_0(x_t)}\,. \tag{43}$$

Here $\widehat{x}_0(x_t)$ is the denoised estimate from state x_t.

Algorithms pair the published operation in burgundy with its UAL replacement in teal; black steps are shared. Only one of the two alternatives is used. The listings retain the published notation and label code transcriptions separately. Table 5 summarizes the nuisance statistics and their use in each task.

Table 5: Nuisance statistics and UAL constructions by task. The rows summarize the specifications below; the final column states how each task forms the UAL mean and covariance.

Task	Nuisance coordinates	Nuisance statistics	UAL construction
Gaussian blur	Kernel widths (s_x, s_y)	Width mean and covariance μ and $\Sigma_\psi = \tau^2 I_2$	Width-sensitivity covariance (Sections C.1–C.2).
Motion blur	Coefficients in the retained kernel basis	Mean and covariance from 131,072 kernels	Gaussian moment matching in retained coordinates (Section C.1).
CT	Flat and dark reference fields	Mean and covariance across 20 flat and 20 dark frames	Linearized flat–dark calibration map (Section C.3).
MRI	Real coil-map coefficients	32 calibration replicates; calibrated maps for the mean prediction	Exact Gaussian marginal for a forward model affine in Gaussian coil-map coefficients (Section C.4).
Ultrasound	Sound speed c_s	Gaussian scale from CUBDL reference spread; truncated quadrature	Per-mode variances from nonlinear sound-speed perturbations (Section C.5).
Microscopy	Five pupil-phase coefficients (Noll modes 5, 6, 7, 8, 11)	Prescribed Gaussian phase law, also used to generate the synthetic observations	Derivatives of the normalized PSF; energy with candidate-dependent covariance (Section C.6).

C.1 Deblurring forward models and nuisance statistics

Gaussian deblurring uses two uncertain kernel widths; motion deblurring uses coefficients in a retained kernel basis. The specified width moments or kernel-population moments determine the uncertainty propagated into each sampler's data term.

Measurements. We use 256×256 FFHQ images (Karras et al., 2019). The 50 candidate images are drawn randomly from the FFHQ validation partition, IDs 60000–69999. The ablation set is a fixed 16-image subset of these 50 images. The same image identifiers are retained across the deblurring samplers. For deblurring, (1) specializes to

$$y = \mathcal{A}(x, \psi) + e, \qquad \mathcal{A}(x, \psi) = A(\psi)x, \qquad e \sim \mathcal{N}(0, \Sigma_e), \qquad \Sigma_e = \sigma_y^2 I. \tag{44}$$

Here $A(\psi)$ applies the kernel $k(\psi)$ to each image channel. We write A_k for this same linear action when specifying the kernel directly, so $A(\psi) = A_{k(\psi)}$. Images use $[-1, 1]$ intensity coordinates; observations use the same scale, with $\sigma_y = 0.02$.

Each paired comparison uses the same observation. The kernel origin and boundary convention are specified below. FFHQ's dataset metadata and accompanying materials are distributed under CC BY-NC-SA 4.0; individual images retain the licenses and author information recorded in their metadata.[1]

[1] https://github.com/NVlabs/ffhq-dataset#licenses

Gaussian blur. The nuisance vector $\psi = (s_x, s_y)$ contains the horizontal and vertical kernel widths in pixels. On the 64×64 kernel grid, indexed by $u, v \in \{-32, \dots, 31\}$, we use

$$k_\psi(u,v) = \frac{1}{Z_\psi} \exp\left(-\frac{u^2}{2s_y^2} - \frac{v^2}{2s_x^2}\right) \mathbf{1}_{\{|u|,|v| \le 31\}}, \tag{45}$$

where Z_ψ normalizes the kernel to unit sum. The width mean and covariance are $\mu = (\bar{s}, \bar{s})$ and $\Sigma_\psi = \tau^2 I_2$. Table 19 gives the three reconstruction settings. The width generator samples the two coordinates independently from normal distributions truncated to $(0, \infty)$, with $\bar{s}$ and τ specifying the post-truncation mean and standard deviation. The kernel sensitivities include the derivative of the unit-sum normalization. Let $\langle f \rangle_{k_\psi} = \sum_{u,v} k_\psi(u,v) f(u,v)$. Then

$$\kappa_x(u,v) = k_\psi(u,v) \frac{v^2 - \langle v^2 \rangle_{k_\psi}}{s_x^3}, \qquad \kappa_y(u,v) = k_\psi(u,v) \frac{u^2 - \langle u^2 \rangle_{k_\psi}}{s_y^3}.$$

Each derivative kernel sums to zero. Evaluating them at the reference widths gives

$$J_\psi(x) = [A_{\kappa_x} x, A_{\kappa_y} x], \qquad U(x) = \tau J_\psi(x).$$

These columns propagate the two width variances into the measurement covariance in (4). We write $A_\mu = A_{k_\mu}$ for the kernel rendered at the mean widths.

BlindDPS estimates a kernel instead of two widths (Chung et al., 2023a). We read the widths directly from the normalized fitted kernel $\widehat{k}$:

$$\begin{aligned} \widehat{s}_x &= \frac{1}{|\omega|} \sqrt{-2 \log \left| \sum_{u,v} \widehat{k}(u,v) e^{-\mathrm{i}\omega v} \right|}, \\ \widehat{s}_y &= \frac{1}{|\omega|} \sqrt{-2 \log \left| \sum_{u,v} \widehat{k}(u,v) e^{-\mathrm{i}\omega u} \right|}, \qquad \omega = \frac{4\pi}{64}. \end{aligned} \tag{46}$$

The Fourier magnitude of a continuous Gaussian of width s is $\exp(-s^2\omega^2/2)$; solving this relation gives the readout. We apply it to the fitted discrete kernel when the Fourier magnitudes lie strictly between zero and one.

This Gaussian frequency-matching rule estimates the current reference widths. It leaves the population mean and covariance unchanged. The non-blind implementations use their reference widths directly.

Motion blur. We generate 131,072 reference kernels per population with the `motionblur` generator included in the PnP-DM release.[2] The kernel size is 64×64, the generator intensity is 0.5, and the reference seeds are $0, \dots, 131071$. Observation kernels use seeds $133120, \dots, 133169$, disjoint from the reference seeds.

For motion blur, the nuisance coordinates describe variation in the generated kernels themselves. We estimate a kernel mean and covariance from each reference population and retain its leading covariance modes.

Identify each unit-sum kernel with its vectorized form $k_j \in \mathbb{R}^{4096}$, and let $M = 131{,}072$. We compute

$$\bar{k} = \frac{1}{M} \sum_{j=1}^{M} k_j, \qquad C_k = \frac{1}{M-1} \sum_{j=1}^{M} (k_j - \bar{k})(k_j - \bar{k})^\mathsf{T}. \tag{47}$$

The full covariance $C_k \in \mathbb{R}^{4096 \times 4096}$ is estimated before dimensionality reduction. Write its eigendecomposition as $C_k = B \operatorname{diag}(\lambda_1, \dots, \lambda_{4096}) B^\mathsf{T}$, with eigenvalues in decreasing order. We retain the leading $P = 256$ covariance eigenvectors in $B_P \in \mathbb{R}^{4096 \times P}$ and their variances in $\Lambda_P \in \mathbb{R}^{P \times P}$, giving $C_k \approx B_P \Lambda_P B_P^\mathsf{T}$. We fixed $P = 256$ before running the covariance-rank sweep, which

[2] `https://github.com/LeviBorodenko/motionblur`

varies P. The choice balances running time against covariance truncation error: the retained modes carry 96% of the kernel variance. Equivalently, let $X = [k_1 - \bar{k}, \ldots, k_M - \bar{k}]/\sqrt{M-1}$. Since $C_k = XX^\mathsf{T}$, its eigenvectors are the left singular vectors of X, and its eigenvalues are the squared singular values. We accumulate and eigendecompose C_k directly. Unit-sum kernels have zero-sum centered deviations, so the covariance has at most 4095 nonzero modes. We remove numerical components in the all-ones direction. In the notation of Section 3.1, the retained nuisance coordinates and specified statistics are

$$k(\psi) = \bar{k} + B_P\psi, \qquad \mu = 0, \qquad \Sigma_\psi = \Lambda_P, \qquad \widehat{\psi} = B_P^\mathsf{T}(\widehat{k} - \bar{k}). \tag{48}$$

Here C_k is the full kernel covariance; Σ_ψ is the covariance in the retained nuisance coordinates. The kernel still has 4096 entries; its uncertain displacement is represented by P coefficients. The projection $\widehat{\psi} = B_P^\mathsf{T}(\widehat{k} - \bar{k})$ extracts the retained component of a fitted kernel. Components orthogonal to B_P are outside this representation.

Let $b_j = \mathrm{unvec}_{64\times 64}((B_P)_{:,j})$ be the kernel-shaped jth covariance eigenvector. Each b_j is a signed, zero-sum perturbation direction, whereas $\bar{k}$ is the unit-sum mean kernel. By linearity of convolution in the kernel,

$$A_{k(\psi)}x = A_{\bar{k}}x + \sum_{j=1}^{P} \psi_j A_{b_j}x.$$

Thus the derivative with respect to ψ_j is $A_{b_j}x$.

For the retained model, the UAL mean and covariance are

$$\bar{y}(x) = \mathcal{A}(x,\mu) = A_{\bar{k}}x, \qquad S(x) = \Sigma_e + \sum_{j=1}^{P} \lambda_j (A_{b_j}x)(A_{b_j}x)^\mathsf{T}. \tag{49}$$

Projecting each centered reference kernel gives $\psi^{(j)} = B_P^\mathsf{T}(k_j - \bar{k})$. These coefficients have sample mean zero and sample covariance

$$\frac{1}{M-1}\sum_{j=1}^{M} \psi^{(j)}(\psi^{(j)})^\mathsf{T} = B_P^\mathsf{T} C_k B_P = \Lambda_P.$$

UAL uses the Gaussian coefficient model $\psi \sim \mathcal{N}(0, \Lambda_P)$. With $F = \Lambda_P^{1/2}$ and $\psi = Fa$, where $a \sim \mathcal{N}(0, I_P)$, the observation model is

$$y = A_{\bar{k}}x + U(x)a + e, \qquad U_{:\ell}(x) = \sqrt{\lambda_\ell} A_{b_\ell}x.$$

Integrating over a gives (49). Convolution is affine in the retained coefficients; the approximations are the Gaussian coefficient law and the truncation of the kernel covariance.

Forward conventions. In Gaussian deblurring, DAPS, PnP-DM, and BlindDPS use zero-padded cross-correlation, with padding of 31 pixels before and 32 after each image dimension. DiffPIR's released closed-form Fourier deblurring step requires circular convolution, so we retain that convention and generate its observations with it. Both members of each sampler pair receive the same observation. Motion observations use circular convolution. DAPS, PnP-DM, and DiffPIR use circular mean and mode operators.

The DiffPIR likelihood step uses $x_+ = (x_- + 1)/2$ and $y_+ = (y_- + 1)/2$, converting the stored $[-1, 1]$ coordinates to $[0, 1]$. For unit-sum circular convolution, $A\mathbf{1} = \mathbf{1}$, so $y_+ = Ax_+ + e_-/2$. Consequently $\sigma_+ = 0.01$ and $\Sigma_{e,+} = \Sigma_{e,-}/4$. Its data-step output is converted back by $x_- = 2x_+ - 1$ before the diffusion update.

C.2 UAL IN DEBLURRING SAMPLERS

Each sampler incorporates UAL's measurement mean and covariance through its existing data update while retaining its pretrained diffusion prior and sampling schedule. Two PnP-DM controls in Gaussian deblurring separate covariance recomputation from differentiation. One recomputes covariance at each candidate while stopping its derivatives; the other fixes covariance at the ensemble

mean of the published sampler's 12 reconstructions. The latter reference is shared across that image's reconstruction seeds; all controls use the same observations and evaluation images. A third control multiplies the point likelihood's noise variance by $v \in \{1, 1.5, 2, 3, 5\}$ on the ablation set. In motion deblurring, the noise-variance multipliers are $v \in \{1, 3, 6, 12, 24, 48, 96, 192, 384\}$ on the ablation set. We compare the multipliers by mean per-trajectory PSNR, keeping the Langevin step size unchanged. Appendix D.4 reports the scalar-control results.

C.2.1 DAPS

DAPS alternates reverse diffusion, conditional sampling, and re-noising (Zhang et al., 2025). We retain this outer loop and differentiate the UAL energy in (5) in the inner measurement update for both Gaussian and motion deblurring. At noise level σ_t, reverse diffusion provides $\widehat{x}_0(\mathbf{x}_t)$ in the published notation. Approximating the inner prior by $\mathcal{N}(\widehat{x}_0(\mathbf{x}_t), \sigma_t^2 I)$ gives the drift $(\widehat{x}_0 - z)/\sigma_t^2$ at candidate z.

Let τ_{D} be DAPS's data-fitting scale. For Gaussian and motion deblurring, $\tau_{\mathrm{D}} = 0.01$. We use $A_\mu = A_{k_\mu}$ for Gaussian blur and $A_\mu = A_{\bar{k}}$ for motion blur. With $d(z) = y - A_\mu z$, the data drift is the negative gradient of the scaled UAL energy

$$E_{\mathrm{D}}(z) = \frac{\sigma_y^2}{\tau_{\mathrm{D}}^2}\left[d(z)^{\mathsf{T}} S(z)^{-1} d(z) + \log\det S(z) - \log\det \Sigma_e\right].$$

Thus

$$g_{\mathrm{D}}(z) = -\nabla_z E_{\mathrm{D}}(z) = -\frac{2\sigma_y^2}{\tau_{\mathrm{D}}^2}\nabla_z \ell_{\mathrm{UAL}}(z; y). \tag{50}$$

Since $\Sigma_e = \sigma_y^2 I$, setting $\Sigma_\psi = 0$ gives $E_{\mathrm{D}}(z) = \|y - A_\mu z\|^2/\tau_{\mathrm{D}}^2$, recovering the published DAPS squared-residual data term. The coefficient in (50) therefore preserves the data-fitting scale τ_{D}, distinct from the observation-noise scale σ_y.

Gaussian blur uses two width-sensitivity columns; motion blur uses the retained kernel-mode responses. We evaluate the energy in (15). Algorithm 2 shows the evaluated branch of the published DAPS algorithm. The paired configurations share their inner step sizes, prior, and re-noising schedule. Our comparison uses the Langevin branch. DAPS uses N_A outer steps and N Langevin steps of size η_t; sampling starts at $T = t_{N_A}$.

Algorithm 2 Decoupled Annealing Posterior Sampling (DAPS): the evaluated branch of published Algorithm 1 (Zhang et al., 2025), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Require: Score model $\mathbf{s}_\theta$, measurement $\mathbf{y}$, noise schedule σ_t, $(t_i)_{i\in\{0,\dots,N_A\}}$.

1: Sample $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \sigma_T^2 I)$.
2: **for** $i = N_A, N_A - 1, \dots, 1$ **do**
3: Compute $\widehat{\mathbf{x}}_0^{(0)} = \widehat{\mathbf{x}}_0(\mathbf{x}_{t_i})$ by solving the probability flow ODE in Eq. (48) of the cited supplement with $\mathbf{s}_\theta$
4: **for** $j = 0, \dots, N-1$ **do**
5: $\widehat{\mathbf{x}}_0^{(j+1)} \leftarrow \widehat{\mathbf{x}}_0^{(j)} + \eta_t\big(\nabla_{\widehat{\mathbf{x}}_0} \log p(\widehat{\mathbf{x}}_0^{(j)} \mid \mathbf{x}_{t_i}) + \nabla_{\widehat{\mathbf{x}}_0} \log p(\mathbf{y} \mid \widehat{\mathbf{x}}_0^{(j)})\big) + \sqrt{2\eta_t}\,\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(\mathbf{0}, I).$
UAL $\widehat{\mathbf{x}}_0^{(j+1)} \leftarrow \widehat{\mathbf{x}}_0^{(j)} + \eta_t\big(\nabla_{\widehat{\mathbf{x}}_0} \log p(\widehat{\mathbf{x}}_0^{(j)} \mid \mathbf{x}_{t_i}) + g_{\mathrm{D}}(\widehat{\mathbf{x}}_0^{(j)})\big) + \sqrt{2\eta_t}\,\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(\mathbf{0}, I).$ (UAL drift; 50)
6: **end for**
7: Sample $\mathbf{x}_{t_{i-1}} \sim \mathcal{N}(\widehat{\mathbf{x}}_0^{(N)}, \sigma_{t_{i-1}}^2 I)$.
8: **end for**
9: **return** $\mathbf{x}_0$

C.2.2 PnP-DM

PnP-DM alternates a likelihood conditional with a diffusion-prior conditional (Wu et al., 2024). At outer iteration k, the likelihood step starts from $\mathbf{x}^{(k)}$ and targets the conditional distribution of an

auxiliary image $\mathbf{u}$, with potential $f(\mathbf{u};\mathbf{y})+\|\mathbf{u}-\mathbf{x}^{(k)}\|^2/(2\rho_k^2)$. Here f is the negative log likelihood potential in the published notation. For UAL, we set $f(\mathbf{u};\mathbf{y}) = \ell_{\mathrm{UAL}}(\mathbf{u};\mathbf{y})$, giving

$$\pi_k(\mathbf{u}) \propto \exp\left[-\ell_{\mathrm{UAL}}(\mathbf{u};\mathbf{y}) - \frac{\|\mathbf{u}-\mathbf{x}^{(k)}\|^2}{2\rho_k^2}\right]. \tag{51}$$

For Gaussian and motion deblurring, both versions use the published general Langevin likelihood routine with 100 inner iterations and step size 10^{-4}. The coupling scale starts at 10, decreases by a factor of 0.9, and is bounded below by 0.3.

Each column of $U(x)$ is linear in x: $u_\ell(x) = L_\ell x$. For Gaussian blur, $L_1 = \tau A_{\kappa_x}$ and $L_2 = \tau A_{\kappa_y}$; for motion blur, $L_\ell = \sqrt{\lambda_\ell} A_{b_\ell}$. With $d = y - A_\mu x$ and $w = S(x)^{-1}d$, the likelihood step uses

$$\nabla_x \ell_{\mathrm{UAL}} = -A_\mu^\mathsf{T} w + \sum_\ell L_\ell^\mathsf{T}\left[S(x)^{-1}u_\ell - w(u_\ell^\mathsf{T} w)\right].$$

The first term differentiates the predicted measurement mean. The two terms in the sum differentiate the log determinant and the covariance-weighted residual penalty. PnP-DM evaluates this gradient at each inner Langevin candidate.

Algorithms 3 and 4 show the published outer sampler and its likelihood subroutine. The prior update is unchanged from the released method. D_θ is the denoiser, and $p(\cdot;\sigma)$ is the Gaussian-smoothed prior at noise level σ.

Algorithm 3 Plug-and-Play Diffusion Models (PnP-DM): published Algorithm 1 (Wu et al., 2024), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Input: initialization $\mathbf{x}_0 \in \mathbb{R}^n$, total number of iterations $K > 0$, coupling strength schedule $\{\rho_k > 0\}_{k=0}^{K-1}$, likelihood potential $f(\,\cdot\,;\mathbf{y})$ with measurements $\mathbf{y} \in \mathbb{R}^m$, pretrained model $D_\theta(\,\cdot\,;\,\cdot\,)$ that approximates $\nabla \log p(\mathbf{x};\sigma)$ with $(D_\theta(\mathbf{x};\sigma)-\mathbf{x})/\sigma^2$.

1: **for** $k = 0,\dots,K-1$ **do**
2: $\quad \mathbf{z}^{(k)} \leftarrow \texttt{LikelihoodStep}(\mathbf{x}^{(k)}, \rho_k, f(\,\cdot\,;\mathbf{y}))$
UAL $\quad \mathbf{z}^{(k)} \leftarrow \texttt{LikelihoodStep}(\mathbf{x}^{(k)}, \rho_k, \ell_{\mathrm{UAL}}(\,\cdot\,;\mathbf{y}))$ (5)
3: $\quad \mathbf{x}^{(k+1)} \leftarrow \texttt{PriorStep}(\mathbf{z}^{(k)}, \rho_k, D_\theta(\,\cdot\,;\,\cdot\,))$
4: **end for**
5: **return** $\mathbf{x}^{(k+1)}$

C.2.3 DIFFPIR

DiffPIR combines a diffusion prediction with a data-fidelity subproblem (Zhu et al., 2023). Let v_t be its prediction in the likelihood coordinates and let $\rho_t^{\mathrm{pub}} > 0$ be the released quadratic coupling coefficient. Since the UAL precision already includes the noise variance, we set $\widetilde{\rho}_t = \rho_t^{\mathrm{pub}}/\sigma_+^2$. The data step uses image and measurement values in $[0,1]$. Define $d(z) = y - A_\mu z$. At each diffusion step, the UAL update evaluates the covariance at the prediction, $S_t = S(v_t)$, and solves

$$z_t = \arg\min_z \left\{\frac{1}{2}(y-A_\mu z)^\mathsf{T} S_t^{-1}(y-A_\mu z) + \frac{\widetilde{\rho}_t}{2}\|z-v_t\|^2\right\}. \tag{52}$$

The normal equation is

$$(A_\mu^\mathsf{T} S_t^{-1} A_\mu + \widetilde{\rho}_t I) z_t = A_\mu^\mathsf{T} S_t^{-1} y + \widetilde{\rho}_t v_t.$$

Multiplying the objective by $2\sigma_+^2$ explains the coupling $\widetilde{\rho}_t = \rho_t^{\mathrm{pub}}/\sigma_+^2$ and recovers the published squared-residual units at zero nuisance covariance. The inverse identity gives the solution

$$z_t = v_t + \widetilde{\rho}_t^{-1} A_\mu^\mathsf{T}(S_t + \widetilde{\rho}_t^{-1} A_\mu A_\mu^\mathsf{T})^{-1}(y - A_\mu v_t). \tag{53}$$

We form $U_t = U(v_t)$ at the diffusion prediction. Set

$$D_t = \sigma_+^2 I + \widetilde{\rho}_t^{-1} A_\mu A_\mu^\mathsf{T}, \qquad H_t = I + U_t^\mathsf{T} D_t^{-1} U_t.$$

We evaluate the inverse in (53) with the Woodbury identity:

$$(D_t + U_t U_t^\mathsf{T})^{-1} = D_t^{-1} - D_t^{-1} U_t H_t^{-1} U_t^\mathsf{T} D_t^{-1}.$$

Algorithm 4 PnP-DM likelihood step: published Algorithm 2 (Wu et al., 2024), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Input: state $\mathbf{x}$, coupling strength $\rho > 0$, likelihood potential $f(\,\cdot\,;\mathbf{y})$ with measurements $\mathbf{y}$
Hyperparameter: step size $\gamma > 0$, number of iterations $J > 0$

$\mathbf{u}_0 \leftarrow \mathbf{x}$
for $j = 0, \dots, J-1$ **do**
$\quad \mathbf{u}_{j+1} \leftarrow \mathbf{u}_j - \gamma \nabla f(\mathbf{u}_j;\mathbf{y}) - \frac{\gamma}{\rho^2}(\mathbf{u}_j - \mathbf{x}) + \sqrt{2\gamma}\,\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(\mathbf{0}, I)$
UAL $\quad \mathbf{u}_{j+1} \leftarrow \mathbf{u}_j - \gamma \nabla \ell_{\mathrm{UAL}}(\mathbf{u}_j;\mathbf{y}) - \frac{\gamma}{\rho^2}(\mathbf{u}_j - \mathbf{x}) + \sqrt{2\gamma}\,\epsilon_j, \quad \epsilon_j \sim \mathcal{N}(\mathbf{0}, I)$ (51)
end for
return $\mathbf{u}_J$

Circular convolution diagonalizes D_t in Fourier coordinates, with diagonal $\sigma_+^2 + |\widehat{k}_\mu|^2/\widetilde{\rho}_t$, where $\widehat{k}_\mu$ is the Fourier transform of the mean kernel. The remaining solve has the dimension of the retained nuisance representation.

Algorithm 5 retains the published reverse-diffusion loop and shows the original data-proximal line and its UAL replacement. In Algorithm 5, $\mathcal{H} = A_\mu$, $\sigma_n = \sigma_+$, and $\rho_t = \rho_t^{\mathrm{pub}}$. We follow the published DiffPIR update, using regularization weight $\lambda = 7$, noise mixing $\zeta = 0.3$, and DDIM noise parameter $\eta = 0$, with 100 denoiser evaluations on the quadratic subsequence of a 1000-step grid. Here $\bar{\alpha}_t$ is the cumulative signal coefficient and $\bar{\sigma}_t = \sqrt{(1-\bar{\alpha}_t)/\bar{\alpha}_t}$ is the corresponding noise level; $t-1$ denotes the next selected time.

Algorithm 5 DiffPIR: published Algorithm 1 (Zhu et al., 2023), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Require: Score model $\mathbf{s}_\theta$, $T, \mathbf{y}, \sigma_n, \{\bar{\sigma}_t\}_{t=1}^T, \zeta, \lambda$
Initialize $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, I)$, pre-calculate $\rho_t \triangleq \lambda \sigma_n^2/\bar{\sigma}_t^2$.
for $t = T$ to 1 **do**
$\quad \mathbf{x}_0^{(t)} = \frac{1}{\sqrt{\bar{\alpha}_t}}\big(\mathbf{x}_t + (1-\bar{\alpha}_t)\mathbf{s}_\theta(\mathbf{x}_t, t)\big)$
$\quad \widehat{\mathbf{x}}_0^{(t)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathcal{H}(\mathbf{x})\|^2 + \rho_t \|\mathbf{x} - \mathbf{x}_0^{(t)}\|^2$
UAL $\quad \widehat{\mathbf{x}}_0^{(t)} = 2\left[\arg\min_{\mathbf{z}} \left\{\sigma_n^2 d(\mathbf{z})^\top S_t^{-1} d(\mathbf{z}) + \rho_t \|\mathbf{z} - v_t\|^2\right\}\right] - \mathbf{1}, \quad v_t = (\mathbf{x}_0^{(t)} + \mathbf{1})/2, \quad S_t = S(v_t).$ (52)
$\quad \widehat{\epsilon} = \frac{1}{\sqrt{1-\bar{\alpha}_t}}\big(\mathbf{x}_t - \sqrt{\bar{\alpha}_t}\widehat{\mathbf{x}}_0^{(t)}\big)$
$\quad \epsilon_t \sim \mathcal{N}(\mathbf{0}, I)$
$\quad \mathbf{x}_{t-1} = \sqrt{\bar{\alpha}_{t-1}}\widehat{\mathbf{x}}_0^{(t)} + \sqrt{1-\bar{\alpha}_{t-1}}\big(\sqrt{1-\zeta}\,\widehat{\epsilon} + \sqrt{\zeta}\,\epsilon_t\big)$
end for
return $\mathbf{x}_0$

C.2.4 BLINDDPS

BlindDPS evolves image and kernel diffusion states and couples them through measurement guidance (Chung et al., 2023a). At each guidance query, the image branch differentiates the UAL energy in (5). Published BlindDPS measures data fit by the residual norm $\|y - \widehat{k} * \widehat{x}_0\|$. With $\Sigma_e = \sigma_y^2 I$, this norm is σ_y times the square root of the point-likelihood energy $\|y - \widehat{k} * \widehat{x}_0\|^2/\sigma_y^2$. We keep this form and replace the point-likelihood energy by Q from (42):

$$D_{\mathrm{UAL}} = \sigma_y \sqrt{Q}, \qquad \nabla D_{\mathrm{UAL}} = \frac{\sigma_y}{\sqrt{Q}} \nabla \ell_{\mathrm{UAL}}. \tag{54}$$

The gradient is a positive multiple of $\nabla \ell_{\mathrm{UAL}}$, so the image step follows the UAL score. The gradient requires $Q > 0$, which holds unless $d = 0$ and $U = 0$. Because D_{UAL} has the units of the residual norm, the kernel weight λ is used unchanged.

With $\Sigma_\psi = 0$, $D_{\mathrm{UAL}} = \|d\|$, the residual norm at the UAL mean. The denoiser Jacobian maps the full image derivative to the diffusion state as in (43). For Gaussian and motion blur, we scale

the UAL image guidance to have the same magnitude as the published image guidance before applying the denoiser Jacobian. Here $\widehat{k}_i$ is the projected kernel estimate at iteration i, denoted $\widehat{\mathbf{k}}_0$ in Algorithm 6. At the current denoised image, let

$$r_{\mathrm{pub},i} = \nabla_z \|y - A_{\widehat{k}_i} z\|,$$
$$r_{\mathrm{UAL},i} = \nabla_z D_{\mathrm{UAL}}(z; y), \qquad c_i = \frac{\|r_{\mathrm{pub},i}\|}{\|r_{\mathrm{UAL},i}\|}.$$

Then $\|c_i r_{\mathrm{UAL},i}\| = \|r_{\mathrm{pub},i}\|$. The image update applies $-\alpha c_i [D_{x_i} \widehat{x}_0]^{\mathsf{T}} r_{\mathrm{UAL},i}$, where α is the guidance step size. We retain the published guidance step size α. For motion blur, $r_{\mathrm{UAL},i}$ is the gradient of the motion fidelity defined below.

Gaussian blur. The fitted kernel gives the current reference widths $\widehat{s}_i$ through (46). The population moments remain $\mu = (\bar{s}, \bar{s})$ and $\Sigma_\psi = \tau^2 I_2$. Define

$$J_i(z) = D_s(A_{k(s)} z)\big|_{s=\widehat{s}_i}.$$

In the model linearized at these reference widths,

$$y \mid z, a \sim \mathcal{N}\big(\mathcal{A}(z, \widehat{s}_i) + J_i(z)(\mu - \widehat{s}_i) + \tau J_i(z) a, \sigma_y^2 I\big),$$
$$a \sim \mathcal{N}(0, I_2).$$

Integrating over a gives the measurement mean

$$\bar{y}(z) = \mathcal{A}(z, \widehat{s}_i) + J_i(z)(\mu - \widehat{s}_i), \qquad \psi_0 = \widehat{s}_i, \tag{55}$$

and covariance $S_i(z) = \sigma_y^2 I + \tau^2 J_i(z) J_i(z)^{\mathsf{T}}$. The correction maps $\mu - \widehat{s}_i$ into measurement space. The image derivative includes the dependence of this mean correction and covariance on z, and the guidance uses the square-root fidelity in (54). In Gaussian blur, the fidelity depends on the fitted kernel through the reference widths $\widehat{s}_i$. The kernel diffusion and projection steps are retained.

Motion blur. For motion blur, let $\widehat{k}_i$ be the current fitted kernel and set $J_i(z) = [A_{b_1} z, \ldots, A_{b_P} z]$. The retained coordinates are $\widehat{\psi}_i = B_P^{\mathsf{T}}(\widehat{k}_i - \bar{k})$, as in (48). The residual and covariance are

$$d_i(z) = y - A_{\widehat{k}_i} z + J_i(z)\widehat{\psi}_i,$$
$$S_i(z) = \Sigma_e + J_i(z) \Lambda_P J_i(z)^{\mathsf{T}}. \tag{56}$$

Both remain differentiable functions of the candidate image. The mean correction gives the effective kernel $k_{i,\mathrm{eff}} = \bar{k} + (I - B_P B_P^{\mathsf{T}})(\widehat{k}_i - \bar{k})$, so $D_z d_i = -A_{k_{i,\mathrm{eff}}}$. It corrects the retained kernel deviation toward the population mean and preserves the deviation outside the retained subspace. The kernel update retains the released guidance. Algorithm 6 retains the published sampler and identifies the two guidance substitutions. In this listing, $\widehat{\mathbf{s}}^i$ and $\widehat{\mathbf{s}}^k$ denote negative image and kernel noise predictions, respectively. The kernel projection P_C enforces nonnegativity and unit sum. We retain the kernel regularizer $R_{\mathbf{k}}$ and weight λ used by the published BlindDPS method.

C.2.5 Fast Diffusion EM

Fast Diffusion EM alternates diffusion posterior sampling of the image with a maximization step for the kernel (Laroche et al., 2024); the released configuration we use guides the image with ΠGDM (Song et al., 2023). Let r_t be the clean-image uncertainty scale of that guidance. The released schedule sets $r_t^2 = v_t/(1 + v_t)$, with $v_t = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}(1 - \frac{\bar{\alpha}_t}{\bar{\alpha}_{t-1}})$, where $\bar{\alpha}_t$ is the cumulative diffusion coefficient and $t-1$ denotes the next scheduled reverse step. At a denoised candidate z, UAL uses the mean response $A_{\bar{k}} z$ and combines clean-image and kernel-coefficient uncertainty in the measurement covariance:

$$\bar{y}_t(z) = A_{\bar{k}} z,$$
$$S_t(z) = \Sigma_e + r_t^2 A_{\bar{k}} A_{\bar{k}}^{\mathsf{T}} + J_\psi(z) \Lambda_P J_\psi(z)^{\mathsf{T}} + r_t^2 \sum_{j=1}^{P} \lambda_j A_{b_j} A_{b_j}^{\mathsf{T}}. \tag{57}$$

Algorithm 6 BlindDPS for blind deblurring: published Algorithm 1 (Chung et al., 2023a), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Require: $N, \mathbf{y}, \alpha, \{\widetilde{\sigma}_i\}_{i=1}^{N}, \lambda, R_{\mathbf{k}}(\cdot)$

$\mathbf{x}_N, \mathbf{k}_N \sim \mathcal{N}(\mathbf{0}, I)$
for $i = N-1$ to 0 **do**
  $\widehat{\mathbf{s}}^{\,i} \leftarrow \mathbf{s}^{i}_{\theta^*}(\mathbf{x}_i, i)$
  $\widehat{\mathbf{s}}^{\,k} \leftarrow \mathbf{s}^{k}_{\theta^*}(\mathbf{k}_i, i)$
  $\widehat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\mathbf{x}_i + \sqrt{1-\bar{\alpha}_i}\,\widehat{\mathbf{s}}^{\,i})$
  $\widehat{\mathbf{k}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_i}}(\mathbf{k}_i + \sqrt{1-\bar{\alpha}_i}\,\widehat{\mathbf{s}}^{\,k})$
  $\widehat{\mathbf{k}}_0 \leftarrow P_C(\widehat{\mathbf{k}}_0)$
  $\mathbf{z}_i, \mathbf{z}_k \sim \mathcal{N}(\mathbf{0}, I)$
  $\mathbf{x}'_{i-1} \leftarrow \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\mathbf{x}_i + \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\widehat{\mathbf{x}}_0 + \widetilde{\sigma}_i\mathbf{z}_i$
  $\mathbf{k}'_{i-1} \leftarrow \frac{\sqrt{\alpha_i}(1-\bar{\alpha}_{i-1})}{1-\bar{\alpha}_i}\mathbf{k}_i + \frac{\sqrt{\bar{\alpha}_{i-1}}\beta_i}{1-\bar{\alpha}_i}\widehat{\mathbf{k}}_0 + \widetilde{\sigma}_i\mathbf{z}_k$
  $\mathbf{x}_{i-1} \leftarrow \mathbf{x}'_{i-1} - \alpha\nabla_{\mathbf{x}_i}\|\mathbf{y} - \widehat{\mathbf{k}}_0 * \widehat{\mathbf{x}}_0\|_2$
UAL $\mathbf{x}_{i-1} \leftarrow \mathbf{x}'_{i-1} - \alpha c_i \nabla_{\mathbf{x}_i} D_{\mathrm{UAL}}(\widehat{\mathbf{x}}_0; \mathbf{y})$ (UAL gradient; 54)
  $L_{\mathbf{k}} \leftarrow \|\mathbf{y} - \widehat{\mathbf{k}}_0 * \widehat{\mathbf{x}}_0\|_2 + \lambda R_{\mathbf{k}}(\widehat{\mathbf{k}}_0)$
UAL $L_{\mathbf{k}} \leftarrow D_{\mathrm{UAL}}(\widehat{\mathbf{x}}_0; \mathbf{y}) + \lambda R_{\mathbf{k}}(\widehat{\mathbf{k}}_0)$ (Gaussian blur; for motion blur, use the original line.) (54)
  $\mathbf{k}_{i-1} \leftarrow \mathbf{k}'_{i-1} - \alpha\nabla_{\mathbf{k}_i} L_{\mathbf{k}}$
end for
return $\mathbf{x}_0, \mathbf{k}_0$

The image update uses the marginal mean $A_{\bar{k}}z$, obtained by averaging over the kernel-coefficient model in (48), in place of the fitted-kernel response; the kernel-maximization step is unchanged. The final covariance term accounts for the interaction between image and kernel uncertainty. These quantities use Fast Diffusion EM's guidance coordinates, with the $[0, 1]$ noise-scale conversion described for DiffPIR. The negative gradient of the Gaussian energy with these moments, including the log determinant and the covariance derivative, replaces the published guidance vector at the denoised candidate; the mapping to the diffusion state and the step size are unchanged. Algorithm 7 gives the published loop in its source notation, with H the kernel operator, and marks the replaced guidance line.

Algorithm 7 Fast Diffusion EM with ΠGDM guidance: published Algorithm 2 (Laroche et al., 2024), with the UAL replacement line.

Original operation | UAL replacement | Shared steps

Require: y, σ, H_T, T

$\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, I)$
for $t = T$ to 1 **do**
  $\widehat{\boldsymbol{\epsilon}} \leftarrow \epsilon(\mathbf{x}_t, t)$
  $\widehat{\mathbf{x}}_0 \leftarrow \frac{1}{\sqrt{\bar{\alpha}_t}}(\mathbf{x}_t - \sqrt{1-\bar{\alpha}_t}\,\widehat{\boldsymbol{\epsilon}})$
  $H_{t-1} \leftarrow$ M-step$(y, \widehat{\mathbf{x}}_0, \sigma)$
  $\mathbf{g} \leftarrow \nabla_{\mathbf{x}_t} \log p(y \mid \mathbf{x}_t, H_{t-1})$ (ΠGDM)
UAL $\mathbf{g} \leftarrow \nabla_{\mathbf{x}_t} \log \mathcal{N}(y; A_{\bar{k}}\widehat{\mathbf{x}}_0, S_t(\widehat{\mathbf{x}}_0))$ (57)
  $\mathbf{s} \leftarrow \zeta_t\mathbf{g} - \frac{1}{\sqrt{1-\bar{\alpha}_t}}\widehat{\boldsymbol{\epsilon}}$
  $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, I)$
  $\mathbf{x}_{t-1} \leftarrow \frac{1}{\sqrt{\alpha_t}}(\mathbf{x}_t + \beta_t\mathbf{s}) + \widetilde{\sigma}_t\mathbf{z}$
end for
return $\mathbf{x}_0, H_0$

C.3 CT WITH FLAT AND DARK REFERENCE UNCERTAINTY

Data and calibration. We use the DM4CT synchrotron rock data (Shi et al., 2026), distributed through Zenodo.[3] It contains 1200 projections, 20 flat frames, and 20 dark frames.

Our application uses 60 views and the released DM4CT synchrotron diffusion prior with DAPS (Zhang et al., 2025). We follow the published DM4CT preprocessing and reference reconstructions. Published uses one flat and one dark frame; UAL uses the mean reference fields. Appendix D also compares both versions using those mean fields. Each CT case is a two-dimensional reconstruction from the same detector row across the 60 projection views.

Here ψ collects the flat and dark reference fields. For raw projection p, flat field f, and dark field d_{dark}, the calibration map acts elementwise as

$$z(p; f, d_{\mathrm{dark}}) = -a_{\mathrm{CT}} \log \max\left\{ \frac{p - d_{\mathrm{dark}}}{f - d_{\mathrm{dark}}}, t_{\min} \right\}. \tag{58}$$

Here a_{CT} is the log-sinogram scale and $t_{\min}$ is the transmission floor. The nuisance vector is $\psi = (f^{\mathsf{T}}, d_{\mathrm{dark}}^{\mathsf{T}})^{\mathsf{T}}$. From each set of $N = 20$ reference frames, we compute

$$\bar{f} = \frac{1}{N}\sum_{j=1}^{N} f_j, \qquad \bar{d}_{\mathrm{dark}} = \frac{1}{N}\sum_{j=1}^{N} d_{\mathrm{dark},j},$$

$$F_f = \frac{[f_1 - \bar{f}, \ldots, f_N - \bar{f}]}{\sqrt{N-1}},$$
$$F_d = \frac{[d_{\mathrm{dark},1} - \bar{d}_{\mathrm{dark}}, \ldots, d_{\mathrm{dark},N} - \bar{d}_{\mathrm{dark}}]}{\sqrt{N-1}}.$$

The Gaussian reference-field model has

$$\mu = \begin{bmatrix} \bar{f} \\ \bar{d}_{\mathrm{dark}} \end{bmatrix}, \qquad \Sigma_\psi = \begin{bmatrix} F_f F_f^{\mathsf{T}} & 0 \\ 0 & F_d F_d^{\mathsf{T}} \end{bmatrix}.$$

These statistics are estimated once and shared across the reconstructions. We use frame-to-frame variability to model the uncertain flat and dark fields, with independent Gaussian fluctuations.

Calibration statistics and propagated covariance. The calibrated measurement is $y = z(p; \bar{f}, \bar{d}_{\mathrm{dark}})$, and A_{CT} is the selected parallel-beam projection operator. Conditioning on the raw projection p, the calibration-error response is

$$\mathcal{A}_p(x, \psi) = A_{\mathrm{CT}} x + z(p; \mu) - z(p; \psi).$$

At unclipped entries, the calibration sensitivities are

$$\frac{\partial z}{\partial f} = \frac{a_{\mathrm{CT}}}{f - d_{\mathrm{dark}}}, \qquad \frac{\partial z}{\partial d_{\mathrm{dark}}} = a_{\mathrm{CT}} \left(\frac{1}{p - d_{\mathrm{dark}}} - \frac{1}{f - d_{\mathrm{dark}}} \right). \tag{59}$$

Clipped entries have zero derivative. Let R broadcast a detector field over the selected angles, and let D_f and D_d be diagonal matrices of these sensitivities evaluated at $(\bar{f}, \bar{d}_{\mathrm{dark}})$. Linearization gives

$$\bar{y}(x) = A_{\mathrm{CT}} x, \qquad J_\psi = -[D_f R, D_d R], \qquad U_{\mathrm{CT}} = -[D_f R F_f, D_d R F_d].$$

A reference-field fluctuation affects all selected angles through the same field, so the propagated covariance retains the resulting measurement correlations. In CT, the observation-noise scale is $\sigma_{\mathrm{CT}} = 1.5$, and the DAPS data-fitting scale is $\tau_{\mathrm{D,CT}} = 1.5$. The covariance is

$$S_{\mathrm{CT}} = \sigma_{\mathrm{CT}}^2 I + U_{\mathrm{CT}} U_{\mathrm{CT}}^{\mathsf{T}}. \tag{60}$$

[3] https://zenodo.org/records/15420527

Reconstruction update. The propagated calibration model gives

$$p_{\mathrm{UAL}}(y \mid x) = \mathcal{N}(y; A_{\mathrm{CT}}x, S_{\mathrm{CT}}), \qquad \nabla_x \log p_{\mathrm{UAL}} = A_{\mathrm{CT}}^{\mathsf{T}} S_{\mathrm{CT}}^{-1}(y - A_{\mathrm{CT}}x).$$

With $d = y - A_{\mathrm{CT}}x$ and $w = S_{\mathrm{CT}}^{-1}d$, the DAPS data drift is

$$g_{\mathrm{CT}}(x) = \frac{2}{\tau_{\mathrm{D,CT}}^2} A_{\mathrm{CT}}^{\mathsf{T}} w \, \frac{\|d\|}{\|w\|}. \tag{61}$$

When $U_{\mathrm{CT}} = 0$, $w = d/\sigma_{\mathrm{CT}}^2$ and $w\|d\|/\|w\| = d$, recovering the published point-likelihood drift $2A_{\mathrm{CT}}^{\mathsf{T}} d/\tau_{\mathrm{D,CT}}^2$. The drift is a positive multiple of the UAL score. The norm ratio preserves the unweighted residual norm before backprojection, so the published step sizes and data-fitting scale are used unchanged while covariance weighting changes the residual direction.

CT uses Algorithm 2 with the calibrated measurement specified above. Its Langevin likelihood-score term is replaced by $g_{\mathrm{CT}}(\widehat{\mathbf{x}}_0^{(j)})$ from (61).

Dataset terms. The DM4CT dataset is distributed under CC BY 4.0 through Zenodo record 15420527; we cite the accompanying DM4CT paper. The pretrained synchrotron diffusion checkpoint is distributed separately under the MIT license.[4]

C.4 MRI WITH COIL-MAP UNCERTAINTY

We estimate coil-map variation by perturbing the calibration measurements and recalibrating the sensitivity maps. The resulting covariance determines the Gaussian coefficient model used in the Ambient DPS likelihood.

Data and forward model. We use multi-coil brain measurements from the NYU fastMRI Initiative (Knoll et al., 2020; Zbontar et al., 2019). We take the middle slice of each volume, apply coil-noise prewhitening and intensity normalization, and use the resulting k-space measurements, sampling mask, and sensitivity maps. Both versions use the released Ambient Diffusion checkpoint trained at acceleration 6, the released acceleration-8 masks, and the same coil-map calibration width and noise convention.

Reference images are reconstructed from full k-space data using the released preparation procedure. We evaluate 50 T2-weighted brain volumes selected at random without replacement from the fastMRI validation set, excluding six development volumes.

Here x denotes the complex image. For sensitivity maps $s = (s_c)_c$, the response of coil c is

$$\mathcal{A}_c(x, s) = M\mathcal{F}(s_c \odot x).$$

Here $\mathcal{F}$ is the released orthonormal Fourier transform and M is the sampling mask (Aali et al., 2025).

Nuisance statistics. We perturb the acquired calibration data at the measured noise level. Let Y_{ACS} contain the eight central calibration lines, and let $\mathcal{E}$ denote ESPIRiT calibration (Uecker et al., 2014). The reference sensitivity maps are $s_0 = \mathcal{E}(Y_{\mathrm{ACS}})$. For each scan, $\sigma_{\mathbb{C}}^2$ is the empirical complex variance of the pooled 30×30 upper-left corner patches of the prewhitened, normalized coil images. We generate $B = 32$ calibration replicates,

$$\begin{aligned} Y_{\mathrm{ACS}}^{(j)} &= Y_{\mathrm{ACS}} + \eta_j, \\ \eta_j &= \frac{\sigma_{\mathbb{C}}}{\sqrt{2}}(\xi_j + \mathrm{i}\zeta_j), \qquad \xi_j, \zeta_j \overset{\mathrm{iid}}{\sim} \mathcal{N}(0, I), \end{aligned}$$

with noise added within the calibration lines. We recalibrate each replicate, $\widetilde{s}_j = \mathcal{E}(Y_{\mathrm{ACS}}^{(j)})$, and align its common coil phase at each pixel to s_0:

$$\rho_j(r) = \sum_c \overline{\widetilde{s}_{j,c}(r)} s_{c,0}(r), \qquad s_{j,c}(r) = \widetilde{s}_{j,c}(r) \frac{\rho_j(r)}{|\rho_j(r)|}.$$

[4] https://huggingface.co/jiayangshi/synchrotron_pixel_diffusion

The phase factor is set to one where the overlap is negligible. Let $\mathcal{R}$ stack real and imaginary parts over all coils and pixels, and define

$$\delta_j = \mathcal{R}(s_j - s_0), \qquad \bar{\delta} = \frac{1}{B}\sum_{j=1}^{B}\delta_j, \qquad D = \frac{[\delta_1 - \bar{\delta}, \ldots, \delta_B - \bar{\delta}]}{\sqrt{B-1}}.$$

An SVD $D = V\,\mathrm{diag}(\varsigma_\ell)W^{\mathsf{T}}$ gives map modes $\mathcal{R}(b_\ell) = V_{:\ell}$ and variances $\lambda_\ell = \varsigma_\ell^2$ that factor the sample covariance DD^{T} of the aligned map perturbations. We retain all 32 returned columns, setting $P = 32$; centering the replicates leaves at most 31 nonzero covariance modes. The likelihood mean uses the reference maps s_0; the centered replicates estimate coil-map variability under the calibration-noise model above.

The coefficients are modeled as $\psi \sim \mathcal{N}(0, \mathrm{diag}(\lambda_1, \ldots, \lambda_P))$, with complex map modes and real coefficients:

$$s_c(\psi) = s_{c,0} + \sum_{\ell=1}^{P} b_{\ell,c}\psi_\ell, \qquad \mathcal{A}_c(x,\psi) = M\mathcal{F}(s_c(\psi)\odot x). \tag{62}$$

We also apply $\mathcal{R}$ to the measurements and write y for the resulting real vector. The predicted mean and covariance factor are

$$\bar{y}(x) = \mathcal{R}(\{M\mathcal{F}(s_{c,0}\odot x)\}_c)\,,$$
$$U_{:\ell}(x) = \sqrt{\lambda_\ell}\,\mathcal{R}(\{M\mathcal{F}(b_{\ell,c}\odot x)\}_c)\,.$$

For proper complex observation noise, the variance per real component is $\sigma_{\mathrm{R}}^2 = \sigma_{\mathbb{C}}^2/2$, giving $\Sigma_e = \sigma_{\mathrm{R}}^2 I$. Integrating the Gaussian coefficient model gives

$$p_{\mathrm{UAL}}(y \mid x) = \mathcal{N}(y; \bar{y}(x), S(x)), \qquad S(x) = \sigma_{\mathrm{R}}^2 I + U(x)U(x)^{\mathsf{T}}.$$

At fixed x, the response is affine in ψ, so this Gaussian marginal requires no nuisance linearization.

Ambient DPS update. Ambient DPS evaluates its data term at a denoised image estimate (Aali et al., 2025). Let v_b be the candidate used by the released loop for batch member b, and let $d_b = y_b - \bar{y}(v_b)$. With Q from (42), UAL substitutes

$$E_b = E_{\mathrm{MRI}}(v_b; y_b) = \sigma_{\mathrm{R}}^2 Q(v_b; y_b) \tag{63}$$

for the squared-residual energy. The factor σ_{R}^2 puts the energy in the units used by the published update; when coil-map covariance is zero, $E_b = \|d_b\|^2$. Both the energy derivative and the released guidance normalization use this substituted energy:

$$g = \nabla_{x_{\mathrm{cur}}} \sum_b E_b, \qquad x_{\mathrm{next},b} = \widehat{x}_b + h\,d_{\mathrm{cur},b} - \frac{l_{\mathrm{ss}}}{\sqrt{E_b}} g_b.$$

Here l_{ss} is the guidance scale and d_{cur} the diffusion drift. The derivative includes the dependence of $\bar{y}(v_b)$ and $S(v_b)$ on the candidate image. Algorithm 8 defines the diffusion quantities and evaluation points. Since $g_b/\sqrt{E_b} = 2\nabla\sqrt{E_b}$, the guidance descends the square-root fidelity $\sqrt{E_b} = \sigma_{\mathrm{R}}\sqrt{Q}$, as in (54).

The pair uses 500 diffusion steps and zero churn.

Sampling loop. Algorithm 8 gives the Ambient DPS sampling loop with UAL. The diffusion scaling and noise schedules are $s(t)$ and $\sigma(t)$. The data term is evaluated at $v_b = E_{\mathrm{xstart,b}}$, using the DPS or ALD definition in the algorithm. The training and inference operators are $\mathcal{A}_{\mathrm{train}}$ and $\mathcal{A}_{\mathrm{inf}}$. The returned reconstruction is x_{cur}.

Dataset acknowledgment. Data used in the preparation of this article were obtained from the NYU fastMRI Initiative database (fastmri.med.nyu.edu) (Knoll et al., 2020; Zbontar et al., 2019). As such, NYU fastMRI investigators provided data but did not participate in analysis or writing of this report. A listing of NYU fastMRI investigators, subject to updates, can be found at fastmri.med.nyu.edu. The primary goal of fastMRI is to test whether machine learning can aid in the reconstruction of medical images.

Algorithm 8 Ambient DPS MRI sampling loop from the official release (Aali et al., 2025), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Require: Released time grid $\{t_i\}_{i=0}^{N}$, s, σ, network D_θ, $\mathcal{A}_{\text{train}}$, $\mathcal{A}_{\text{inf}}$, corruption mask, measurement y, Gaussian latents, and the released churn and guidance settings

$x_{\text{next}} \leftarrow \texttt{latents}\,\sigma(t_0)s(t_0)$
for $i = 0, \dots, N-1$ **do**
  $t_{\text{cur}} \leftarrow t_i$, $t_{\text{next}} \leftarrow t_{i+1}$, $x_{\text{cur}} \leftarrow x_{\text{next}}$
  $\gamma \leftarrow \min(S_{\text{churn}}/N, \sqrt{2}-1)$ if $S_{\text{min}} \le \sigma(t_{\text{cur}}) \le S_{\text{max}}$, else 0
  $\widehat{t} \leftarrow \sigma^{-1}\big(\texttt{round_sigma}((1+\gamma)\sigma(t_{\text{cur}}))\big)$
  Draw $\epsilon \sim \mathcal{N}(0, I)$
  $\widehat{x} \leftarrow \dfrac{s(\widehat{t})}{s(t_{\text{cur}})} x_{\text{cur}} + s(\widehat{t}) S_{\text{noise}} \sqrt{[\sigma(\widehat{t})^2 - \sigma(t_{\text{cur}})^2]_+}\,\epsilon$
  $h \leftarrow t_{\text{next}} - \widehat{t}$
  if the network was trained on fully sampled data **then**
    $\texttt{denoised} \leftarrow D_\theta(\widehat{x}/s(\widehat{t}), \sigma(\widehat{t}))$
  else
    $\texttt{masked} \leftarrow \mathcal{A}^*_{\text{train}}\mathcal{A}_{\text{train}}(\widehat{x})$
    Stack its real and imaginary channels and concatenate the corruption mask to form $\texttt{net_input}$
    $\texttt{denoised} \leftarrow D_\theta(\texttt{net_input}/s(\widehat{t}), \sigma(\widehat{t}))$, retaining the released image-output channels
  end if
  $d_{\text{cur}} \leftarrow \left(\dfrac{\dot\sigma(\widehat{t})}{\sigma(\widehat{t})} + \dfrac{\dot s(\widehat{t})}{s(\widehat{t})}\right)\widehat{x} - \dfrac{\dot\sigma(\widehat{t})s(\widehat{t})}{\sigma(\widehat{t})}\texttt{denoised}$
  if $\texttt{l_type} = \texttt{DPS}$ **then**
    $E_{\text{xstart}} \leftarrow \dfrac{x_{\text{cur}} + s(t_{\text{cur}})^2(\texttt{denoised} - x_{\text{cur}})}{s(t_{\text{cur}})}$
  else
    $E_{\text{xstart}} \leftarrow \texttt{denoised}$
  end if
  $E_b \leftarrow \|y_b - \mathcal{A}_{\text{inf}}(E_{\text{xstart},b})\|^2$ for each batch member b
UAL $E_b \leftarrow E_{\text{MRI}}(E_{\text{xstart},b}; y_b) = \sigma_{\text{R}}^2 Q(E_{\text{xstart},b}; y_b)$ for each batch member b (63)
  $g \leftarrow \nabla_{x_{\text{cur}}} \sum_b E_b$
  $x_{\text{next},b} \leftarrow \widehat{x}_b + h\, d_{\text{cur},b} - (l_{\text{ss}}/\sqrt{E_b})g_b$
  Detach x_{next} before the next iteration
end for
return x_{cur}

C.5 Ultrasound with sound-speed uncertainty

Differences between submitted and optimized sound speeds in the CUBDL acquisitions set the uncertainty scale. We propagate nonlinear sound-speed variation at a fixed pilot reconstruction to obtain the per-mode variances used by DRUS.

Data and forward model. We use the acquired PICMUS contrast and resolution phantoms provided with the DRUS release and its linked supplement, with 12 paired trajectories per phantom (Liebgott et al., 2016; Zhang et al., 2024). Channel measurements y_{raw} and beamformed measurements y_d are related by

$$y_{\text{raw}} = H(c_s)x + n, \qquad y_d = B_0 y_{\text{raw}} = B_0 H(c_s)x + B_0 n, \qquad B_0 = B(c_{s,0}). \tag{64}$$

UAL uses the beamformed measurement $y = y_d$ in (1). The signal is reflectivity, the nuisance is $\psi = c_s$, and $\mathcal{A}(x, c_s) = B_0 H(c_s)x$. The reference sound speed is $c_{s,0} = 1540\,\text{m/s}$. When varying c_s, we keep the beamformer B_0, channel-data layout, and measurement window fixed at their reference definitions. Only $H(c_s)$ changes.

The reference beamformed operator is $H_d = B_0 H(c_{s,0})$, with singular value decomposition $H_d = U_d S_d V_d^{\mathsf{T}}$. The published DRUS spectral step uses a scalar noise model; the UAL replacement uses per-mode variances. The sound-speed uncertainty scale comes from the nine CUBDL 040GSE reference acquisitions (Hyun et al., 2021), MYO001–MYO006 and UFL003–UFL005 in

the released workbook.[5] For their submitted-minus-optimized speed offsets $\delta_j = 1540 - c_{s,j}^{\mathrm{opt}}$, we compute

$$\bar{\delta} = \frac{1}{9}\sum_{j=1}^{9}\delta_j, \qquad \sigma_c^2 = \frac{1}{8}\sum_{j=1}^{9}(\delta_j - \bar{\delta})^2.$$

This gives $\sigma_c = 26.9\,\mathrm{m/s}$. The offsets determine the spread; the Gaussian sound-speed model remains centered at the released value $c_{s,0} = 1540\,\mathrm{m/s}$.

Nuisance propagation. For each acquired case, let (t_k, w_k) be the nodes and weights of a 32-node Gauss–Hermite rule for a standard Gaussian. We retain the eight nodes with $|t_k| \le 2$ and use

$$c_{s,k} = c_{s,0} + \sigma_c t_k, \qquad \pi_k = \frac{w_k}{\sum_{j:\,|t_j|\le 2} w_j}.$$

Thus σ_c is the Gaussian scale before truncation. Sound speed changes the propagation delay $(d_{\mathrm{tx},p} + d_{\mathrm{rx},pe})/c_s$, where $d_{\mathrm{tx},p}$ and $d_{\mathrm{rx},pe}$ are the transmit and receive path lengths for pixel p and receive element e. The pulse model converts these delays into $H(c_s)$; the beamformer, channel layout, and measurement window remain fixed. At each retained node, we evaluate the nonlinear perturbation

$$\Delta_k = B_0[H(c_{s,k}) - H(c_{s,0})]x_{\mathrm{ref}}, \qquad \bar{\Delta} = \sum_k \pi_k \Delta_k. \tag{65}$$

Here x_{ref} is a pilot reconstruction from the same beamformed measurement, produced by the released DRUS sampler and averaged over its three channels. The per-mode variances are computed once before sampling and use no ground truth. The weighted covariance is

$$C_\Delta = \sum_k \pi_k(\Delta_k - \bar{\Delta})(\Delta_k - \bar{\Delta})^{\mathsf{T}}.$$

These are quadrature moments, so the covariance uses the normalized weights directly. For column u_i of U_d, the per-mode variance is

$$\gamma_i^2 = \gamma^2 + u_i^{\mathsf{T}} C_\Delta u_i = \gamma^2 + \sum_k \pi_k [u_i^{\mathsf{T}}(\Delta_k - \bar{\Delta})]^2. \tag{66}$$

Here γ is the released scalar noise standard deviation, including its $\sqrt{3}$ factor for the three-channel representation. The same spectral variances are used in each channel. The perturbations retain the nonlinear dependence on sound speed.

DRUS update. For each image channel, write $\widetilde{y} = U_d^{\mathsf{T}} y_d$ and $z = V_d^{\mathsf{T}} x$, and let s_i be the singular values of H_d. The spectral Gaussian data term is

$$\ell_{\mathrm{US}}(x; y_d) = \frac{1}{2}\sum_i \left[\frac{(\widetilde{y}_i - s_i z_i)^2}{\gamma_i^2} + \log \gamma_i^2\right] + \mathrm{const.}$$

DRUS retains the nominal mean response $H_d x$ and replaces its scalar noise level γ with γ_i in initialization, mode-selection thresholds, residual weights, and sampling variances (Algorithm 9). The diffusion prior and remaining sampling operations are unchanged.

Spectral sampling loop. Algorithm 9 gives the DRUS spectral updates (Zhang et al., 2024). It uses $H_d = U_d S_d V_d^{\mathsf{T}}$, with singular values s_l. Its scalar noise level σ_0 corresponds to γ. Here η_C controls noise mixing in the default spectral update, η_A controls the residual/noise mixture on I_A, and η_B controls the measurement blend and sampling variance on I_B. Unmeasured coordinates retain the published update, and independent Gaussian draws retain their published ordering.

Dataset terms. The PICMUS organizers permit use of the data and code and request acknowledgment of the challenge and citation of its proceedings paper (Liebgott et al., 2016).[6]

[5] https://gitlab.com/dongwoon.hyun/cubdl/-/blob/master/datasets/CUBDL_Phantom_Details.xlsx

[6] https://www.creatis.insa-lyon.fr/Challenge/IEEE_IUS_2016/node/48

Algorithm 9 DRUS spectral sampling loop from the official release (Zhang et al., 2024), with UAL replacement lines.

Original operation | UAL replacement | Shared steps

Require: Gaussian input ξ, released time sequence `seq`, noise predictor, $\bar{\alpha}_t$, U_d, S_d, V_d, measurement y_d, $\sigma_0, \eta_A, \eta_B, \eta_C$

$\widetilde{y} \leftarrow U_d^{\mathsf{T}} y_d$, $\sigma_T \leftarrow \sqrt{(1-\bar{\alpha}_{\texttt{seq[-1]}})/\bar{\alpha}_{\texttt{seq[-1]}}}$

Set $a_l \leftarrow \widetilde{y}_l/s_l$ and $b_l \leftarrow \sigma_0/s_l$ where $s_l\sigma_T > \sigma_0$; set both to zero elsewhere

UAL Set $a_l \leftarrow \widetilde{y}_l/s_l$ and $b_l \leftarrow \gamma_l/s_l$ where $s_l\sigma_T > \gamma_l$; set both to zero elsewhere (66)

$x \leftarrow V_d\big(a + \sqrt{[\sigma_T^2 - b^2]_+} \odot \xi\big)/\sigma_T$

`xs` $\leftarrow [x]$; `x0_preds` $\leftarrow []$; `seq_next` $\leftarrow [-1] +$ `seq[:-1]`

for (i, j) in `zip(reversed(seq), reversed(seq_next))` **do**

  $\epsilon \leftarrow$ released noise predictor at (`xs[-1]`, i), retaining its image-noise channels

  $\widehat{x}_0 \leftarrow (\texttt{xs[-1]} - \sqrt{1-\bar{\alpha}_i}\,\epsilon)/\sqrt{\bar{\alpha}_i}$

  $\sigma_{\text{next}} \leftarrow \sqrt{(1-\bar{\alpha}_j)/\bar{\alpha}_j}$, with $\bar{\alpha}_{-1} = 1$

  $v \leftarrow V_d^{\mathsf{T}} \widehat{x}_0$

  $I_A \leftarrow \{l : s_l\sigma_{\text{next}} < \sigma_0\}$, $I_B \leftarrow \{l : s_l\sigma_{\text{next}} > \sigma_0\}$

UAL $I_A \leftarrow \{l : s_l\sigma_{\text{next}} < \gamma_l\}$, $I_B \leftarrow \{l : s_l\sigma_{\text{next}} > \gamma_l\}$ (66)

  Draw a full signal-coordinate Gaussian vector ξ_C

  $\overline{x}_{\text{next}} \leftarrow v + \sigma_{\text{next}}\sqrt{1-\eta_C^2}\, V_d^{\mathsf{T}}\epsilon + \sigma_{\text{next}}\eta_C\xi_C$

  Draw ξ_A on I_A and overwrite, for $l \in I_A$,

$$\overline{x}_{\text{next},l} \leftarrow v_l + \sigma_{\text{next}}\sqrt{1-\eta_A^2}\,\frac{\widetilde{y}_l - s_l v_l}{\sigma_0} + \sigma_{\text{next}}\eta_A\xi_{A,l}$$

UAL
$$\overline{x}_{\text{next},l} \leftarrow v_l + \sigma_{\text{next}}\sqrt{1-\eta_A^2}\,\frac{\widetilde{y}_l - s_l v_l}{\gamma_l} + \sigma_{\text{next}}\eta_A\xi_{A,l} \quad (66)$$

  Draw a full measurement-coordinate Gaussian vector ξ_B and overwrite, for $l \in I_B$,

$$\overline{x}_{\text{next},l} \leftarrow \eta_B\frac{\widetilde{y}_l}{s_l} + (1-\eta_B)v_l + \sqrt{\sigma_{\text{next}}^2 - \eta_B^2\sigma_0^2/s_l^2}\,\xi_{B,l}$$

UAL
$$\overline{x}_{\text{next},l} \leftarrow \eta_B\frac{\widetilde{y}_l}{s_l} + (1-\eta_B)v_l + \sqrt{\sigma_{\text{next}}^2 - \eta_B^2\gamma_l^2/s_l^2}\,\xi_{B,l} \quad (66)$$

  $x_{\text{next}} \leftarrow \sqrt{\bar{\alpha}_j}\, V_d\overline{x}_{\text{next}}$

  Append $\widehat{x}_0$ to `x0_preds` and x_{next} to `xs`

end for

return `xs`, `x0_preds`; the released reconstruction is `xs[-1]`

C.6 Fluorescence microscopy with pupil-phase uncertainty

We reconstruct a thin two-dimensional fluorescent object from a blurred measurement. Five uncertain pupil-phase coefficients determine the point-spread function. Their prescribed Gaussian distribution defines the synthetic aberrations and the uncertainty propagated by UAL.

Reference images and measurements. We use four stacks from the CARE Planaria test data (Weigert et al., 2018). From the 192 available axial sections, we select 13, 13, 12, and 12 uniformly spaced sections from the four stacks in filename order. We take a central 256×256 crop from each section. Selection uses no reconstruction scores. The crop is mapped to $[-1, 1]$ using the minimum and maximum of its full source section, then used as the reference signal x in

$$y = A_{k(\psi^\star)}x + e, \qquad e \sim \mathcal{N}(0, \sigma_e^2 I), \qquad \sigma_e = 0.02, \tag{67}$$

where A_k denotes circular convolution with kernel k, and $\psi^\star$ contains the realized pupil-phase coefficients. The reference is an acquired CARE image section; the aberrated measurement is generated by this forward model.

Optical model. The scalar pupil model uses wavelength $\lambda_{\text{em}} = 0.6\,\mu\text{m}$, numerical aperture $\text{NA} = 1.2$, refractive index $n_{\text{med}} = 1.33$, and a fixed focal offset $z_f = 1\,\mu\text{m}$. For normalized pupil coordinates $\xi = (u, v)$ in the unit disk, the phase is

$$\phi(\xi;\psi) = \sum_{j=1}^{5}\psi_j Z_{\iota_j}(\xi) - 2\pi z_f\left[\sqrt{\left(\frac{n_{\text{med}}}{\lambda_{\text{em}}}\right)^2 - \left(\frac{\text{NA}}{\lambda_{\text{em}}}\right)^2\|\xi\|^2} - \frac{n_{\text{med}}}{\lambda_{\text{em}}}\right]. \tag{68}$$

Here $\iota = (5, 6, 7, 8, 11)$ selects unit-RMS Zernike modes in Noll ordering (Noll, 1976), and ψ_j is measured in phase radians. Writing $r^2 = u^2 + v^2$, the implemented modes are $\sqrt{6}(2uv)$, $\sqrt{6}(u^2 -$

$v^2)$, $\sqrt{8}\,v(3r^2-2)$, $\sqrt{8}\,u(3r^2-2)$, and $\sqrt{5}(6r^4-6r^2+1)$. The focal offset is fixed; it is not one of the nuisance coordinates.

Let Ω be the points of a 64×64 midpoint grid inside the unit disk. A positive-exponent matrix Fourier transform gives the field at detector positions $\mathbf{r}$:

$$h_\psi(\mathbf{r}) = \sum_{\xi\in\Omega} \exp\big(i\phi(\xi;\psi)\big) \exp\left(2\pi i\, \frac{\mathrm{NA}}{\lambda_{\mathrm{em}}} \mathbf{r}^\mathsf{T}\xi\right), \qquad k(\psi)[\mathbf{r}] = \frac{|h_\psi(\mathbf{r})|^2}{\sum_{\mathbf{r}'} |h_\psi(\mathbf{r}')|^2}. \tag{69}$$

The detector grid has 65×65 samples with $0.1\,\mu\mathrm{m}$ spacing. The discrete kernel is normalized on this grid before circular convolution.

Aberration distribution and UAL. The five pupil-phase coefficients follow the prescribed Gaussian distribution

$$q = \mathcal{N}(\mu, \Sigma_\psi), \qquad \mu = (0.10, -0.10, 0.05, 0, 0.10)^\mathsf{T}, \qquad \Sigma_\psi = 0.15^2 I_5. \tag{70}$$

This distribution generates the synthetic aberrations and provides the mean and covariance used by UAL. We linearize at $\psi_0 = \mu$, so the first-order mean correction vanishes. Independent observation noise is added to each measurement. One draw $\psi^\star \sim q$ and one noisy measurement are fixed per section and shared by both sampler versions and all 12 trajectories.

To propagate phase uncertainty, we differentiate the optical intensity and its normalization. Define $I_\psi(\mathbf{r}) = |h_\psi(\mathbf{r})|^2$ and $T_\psi = \sum_{\mathbf{r}} I_\psi(\mathbf{r})$. The pupil-field derivative at μ is

$$h_j(\mathbf{r}) = \partial_{\psi_j} h_\psi(\mathbf{r})\big|_\mu = \mathrm{i} \sum_{\xi\in\Omega} Z_{\iota_j}(\xi) e^{\mathrm{i}\phi(\xi;\mu)} e^{2\pi\mathrm{i}(\mathrm{NA}/\lambda_{\mathrm{em}})\mathbf{r}^\mathsf{T}\xi}.$$

The intensity and normalized-kernel derivatives are

$$I_j(\mathbf{r}) = 2\,\mathrm{Re}\Big[\overline{h_\mu(\mathbf{r})} h_j(\mathbf{r})\Big],$$
$$\kappa_j(\mathbf{r}) = \frac{I_j(\mathbf{r}) - k(\mu)[\mathbf{r}] \sum_{\mathbf{r}'} I_j(\mathbf{r}')}{T_\mu}, \qquad b_j = 0.15\kappa_j.$$

Each κ_j sums to zero, so it changes the PSF shape while preserving unit sum. For $a \sim \mathcal{N}(0, I_5)$, linearization gives $k(\mu + 0.15a) \approx k(\mu) + \sum_j a_j b_j$. Under this linearization, the conditional measurement model is

$$y \mid x, a \sim \mathcal{N}\left(A_{k(\mu)}x + \sum_{j=1}^{5} a_j A_{b_j} x, \sigma_e^2 I\right).$$

Integrating over a gives

$$\bar{y}(x) = A_{k(\mu)}x, \qquad U(x) = [A_{b_1}x, \ldots, A_{b_5}x], \qquad S(x) = \sigma_e^2 I + U(x)U(x)^\mathsf{T}. \tag{71}$$

We prepare the reference kernel and its five scaled derivative kernels b_j once. The reduced system is 5×5 and is evaluated by Cholesky factorization in double precision.

Released sampler and UAL update. Both versions use PnP-DM (Wu et al., 2024) with the ED-Diff CARE checkpoint (Luo et al., 2026) and the same mean kernel $k(\mu)$. For candidate image u, with $d(u) = y - A_{k(\mu)}u$, the UAL likelihood step uses

$$\ell_{\mathrm{UAL}}(u; y) = \tfrac{1}{2} d(u)^\mathsf{T} S(u)^{-1} d(u) + \tfrac{1}{2} \log\det S(u) + \tfrac{m}{2} \log(2\pi),$$

with $S(u)$ from (71). At outer iterate $x^{(k)}$, the conditional target is proportional to $\exp[-\ell_{\mathrm{UAL}}(u; y) - \|u - x^{(k)}\|^2/(2\rho_k^2)]$. Its Langevin gradient includes the covariance derivatives, as in Algorithm 4. The UAL likelihood step takes 100 steps of size 10^{-4} per likelihood update.

Algorithm 3 gives the shared outer sampler; the diffusion-prior update is unchanged from the released method. Both versions use 100 outer iterations in total, including 40 burn-in iterations, and the variance-preserving SDE mode with 100 prior steps. The coupling scale ρ_k starts at 10, decays by a factor of 0.9 per iteration, and has minimum 0.3. For each seed, we retain the last sample returned by PnP-DM. Twelve paired seeds, 0 through 11, give 600 reconstructions per version. The realized aberration coefficients and true PSFs are used only in measurement generation and evaluation.

Diffusion prior, scoring, and data access. The released ED-Diff model has one input channel, 64 base channels, three residual blocks per resolution level, and a 1,000-step linear diffusion schedule. For each prior update, we map the requested noise level to the nearest level on this schedule and convert the pretrained noise prediction to a clean-signal estimate, clipped to $[-1, 1]$. No retraining is performed. PSNR and SSIM use data range 2 and are computed per trajectory before averaging within sections and across the 50 sections. The reference sections come from CARE's original test stacks. The released ED-Diff prior is pretrained on high-SNR images (Luo et al., 2026).

CARE data are distributed through the Edmond repository, doi:10.17617/3.FDFZOF, whose metadata specify CC0 1.0. The ED-Diff code and checkpoint are obtained through the authors' release. The examined release does not specify a separate redistribution license for the checkpoint.

C.7 Evaluation metrics and reporting

Reported 95% intervals use 20,000 paired bootstrap resamples of images for deblurring, detector rows within the same acquisition for CT, and scans for MRI, after computing each case's reconstruction and ensemble scores. Ablation error bars are standard errors across the ablation set. The microscopy whiskers in Figure 1 resample the 50 sections in pairs and describe variation within these sections. Per-stack gains provide a separate descriptive comparison. Ultrasound whiskers in Figure 1 show one standard error of the paired differences across the 12 contrast-phantom trajectories. Following the DRUS evaluation protocol, we report contrast-to-noise ratio (CNR), generalized contrast-to-noise ratio (gCNR), and speckle SNR on the contrast phantom, and axial and lateral full width at half maximum (FWHM) on the resolution phantom.

For one case, let $\widehat{x}^{(1)}, \dots, \widehat{x}^{(R)} \in \mathbb{R}^n$ be the vectorized reconstructions in scoring coordinates with $R \geq 2$, and let $x^\star$ be the reference. We distinguish the average trajectory score from the score of the ensemble-mean reconstruction:

$$\overline{\text{metric}} = \frac{1}{R}\sum_{r=1}^{R} \text{metric}(\widehat{x}^{(r)}, x^\star), \qquad \text{metric}_{\text{ens}} = \text{metric}(\widehat{x}_{\text{ens}}, x^\star). \tag{72}$$

This distinction applies to both PSNR and SSIM. For the real-valued tasks, $\widehat{x}_{\text{ens}} = R^{-1}\sum_r \widehat{x}^{(r)}$. For MRI, individual magnitude reconstructions are used for per-trajectory PSNR and SSIM and for the three fair ensemble scores. For ensemble-mean PSNR and SSIM, we first average the complex reconstructions and then take the magnitude. Case-level scores are averaged over the reported cases.

For a reconstruction z, we use

$$\text{MSE}(z, x^\star) = \frac{1}{n}\|z - x^\star\|_2^2, \qquad \text{PSNR}(z, x^\star) = 10\log_{10}\frac{D^2}{\text{MSE}(z, x^\star)}, \tag{73}$$

where D is the task-specific PSNR peak. If $\overline{\Delta}_{\text{PSNR}}$ is the mean paired PSNR gain, the geometric mean of the UAL-to-original MSE ratios is $10^{-\overline{\Delta}_{\text{PSNR}}/10}$. We use this relation for the relative MSE reductions in the Abstract and Introduction. For deblurring and microscopy, PSNR and SSIM use the data range 2 for intensities in $[-1, 1]$. CT uses the reference-image range for both; MRI uses the reference-image maximum for PSNR and its range for SSIM. SSIM uses the scikit-image default window settings; color-image scores are averaged over channels.

The fair finite-ensemble estimators of the continuous ranked probability score (CRPS) and the energy score (Gneiting & Raftery, 2007; Ferro, 2014) use off-diagonal ensemble pairs:

$$\text{CRPS}_{\text{f}} = \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{R}\sum_{r=1}^{R}|\widehat{x}_i^{(r)} - x_i^\star| - \frac{1}{R(R-1)}\sum_{r<s}|\widehat{x}_i^{(r)} - \widehat{x}_i^{(s)}|\right], \tag{74}$$

$$\text{ES}_{\text{f}} = \frac{1}{R}\sum_{r=1}^{R}\|\widehat{x}^{(r)} - x^\star\|_2 - \frac{1}{R(R-1)}\sum_{r<s}\|\widehat{x}^{(r)} - \widehat{x}^{(s)}\|_2. \tag{75}$$

The energy score uses the Euclidean norm of the full image vector, without an additional dimension normalization.

For the local variogram score, define the pair features $v_{ij}^{(r)} = |\widehat{x}_i^{(r)} - \widehat{x}_j^{(r)}|^{1/2}$ and $v_{ij}^\star = |x_i^\star - x_j^\star|^{1/2}$ using the variogram construction of Scheuerer & Hamill (2015). Let $\bar{v}_{ij} = R^{-1}\sum_r v_{ij}^{(r)}$. For each

offset g, let $\mathcal{P}_g$ contain the valid within-channel pixel pairs at that offset. We use lags 1, 2, 4, 8, and 16 in the horizontal, vertical, and two diagonal directions, giving 20 offset groups. We use the fair estimator

$$\mathrm{LVS_f} = \frac{1}{20}\sum_g \frac{1}{|\mathcal{P}_g|}\sum_{(i,j)\in\mathcal{P}_g}\left[(\bar{v}_{ij}-v^\star_{ij})^2 - \frac{1}{R(R-1)}\sum_{r=1}^{R}(v^{(r)}_{ij}-\bar{v}_{ij})^2\right]. \tag{76}$$

Thus valid pairs and channels are averaged within each offset group, and the groups receive equal weight. Lower values of all three ensemble scores are better. Metrics are computed in double precision, without additional clipping or rescaling.

Table 6: Pointwise 95% paired-bootstrap confidence intervals for the primary differences in Table 1. PSNR gain is UAL minus Published for mean per-trajectory PSNR (dB); CRPS reduction is Published minus UAL, scaled by 10^3.

Sampler / task	PSNR gain 95% CI	CRPS reduction 95% CI $\times 10^3$
Gaussian blur: mean width 3 pixels, standard deviation 0.5 pixels		
DAPS	$[0.20, 0.61]$	$[-0.10, 1.65]$
PnP-DM	$[.26, .84]$	$[1.79, 3.49]$
DiffPIR	$[.59, 1.46]$	$[2.35, 5.31]$
BlindDPS	$[.80, .99]$	$[1.91, 2.51]$
Motion blur		
DAPS	$[10.48, 11.85]$	$[97.27, 137.38]$
PnP-DM	$[8.88, 10.13]$	$[95.12, 133.09]$
DiffPIR	$[2.50, 3.83]$	$[85.82, 108.75]$
BlindDPS	$[-0.27, -0.07]$	$[0.80, 2.70]$
Fast Diffusion EM	$[5.71, 6.94]$	$[45.31, 56.27]$
Imaging applications		
CT	$[.173, .236]$	$[.516, .671]$
MRI	$[0.0097, 0.0160]$	$[0.0239, 0.0435]$
Microscopy	$[0.238, 0.649]$	$[0.671, 2.426]$

D Additional experimental results

This appendix gives the supporting comparisons and full metric summaries for Section 4.

D.1 Numerical likelihood reference

For Figure 2, we integrate the nonlinear Gaussian-blur likelihood over two independent positive-truncated-normal width distributions, with mean width 3 pixels and standard deviations $\tau \in \{0.25, 0.5, 1, 1.5\}$ pixels. At fixed candidate x and measurement y, we center adaptive tensor-product Gauss–Hermite quadrature at a mode of $p(\psi \mid x, y) \propto q(\psi)p(y \mid x, \psi)$. We find the mode by L-BFGS-B with positive width bounds, starting from $(3, 3)$ and $(3+\tau, 3-\tau)$ pixels and retaining the solution with the lower negative log posterior. The Gaussian quadrature proposal h has covariance given by the inverse of the symmetrized Hessian of $-\log[q(\psi)p(y \mid x, \psi)]$, computed by centered finite differences of the nuisance gradient at the mode. The transformed quadrature weights include $q(\psi)/h(\psi)$, with zero weight outside the positive-width support. The likelihood calculations use the same kernel grid, boundary convention, crop, and noise standard deviation 0.02 as the paired reconstructions.

Writing $g = \nabla_x \log p(y \mid x)$ and $\ell = -\log p(y \mid x)$, the plotted errors are $100\|g - g_{\mathrm{ref}}\|_2/\|g_{\mathrm{ref}}\|_2$ percent and $|\ell - \ell_{\mathrm{ref}}|/m$ nats per measurement, with $m = 3 \times 256 \times 256$. The reference score is the nuisance-posterior-weighted average of the conditional image scores. At each uncertainty level, all three likelihoods use the same candidate and observation. We average errors over trajectories within each image before computing the mean and SE across images.

We checked numerical convergence for all 768 final candidate–observation pairs in Figure 2a,b by increasing the quadrature order from 5 to 9 per width coordinate. The largest observed changes were

5.82×10^{-11} nats in energy and 5.77×10^{-12} in the relative ℓ_2 score change. These changes with quadrature order are far below the plotted approximation errors. To compare likelihood accuracy during sampling, we reran the published PnP-DM sampler on the 16 ablation images with seed 0 at each τ. We saved its prior-step outputs at zero-based outer iterations 0, 25, 50, 74, and 99 of 100 and evaluated all three likelihoods at these same candidates against the quadrature reference. Relative to the original final reconstructions, the reruns change the measured score error by at most 0.0121 percentage points and the energy error by at most 6.82×10^{-5} nats per measurement. Table 7 reports the relative score error; UAL also has the lowest mean energy error at every sampling state and uncertainty level. At the first saved state, UAL's mean relative image-score error ranges from 17.5% to 19.9% across uncertainty levels.

Table 7: Relative image-score error (percent) at five saved PnP-DM sampling states in Gaussian deblurring. Entries are mean $\pm$ one SE across 16 images, with one trajectory per image at each τ. Candidates are prior-step outputs from Published; iteration indices are zero-based. Figure 2c labels the sampling states nominally as 1, 25, 50, 75, and 100. Numerical marginalization in Appendix D.1 provides the reference.

τ	Likelihood	iter. 0	iter. 25	iter. 50	iter. 74	iter. 99
0.25	Published	36 ± 5	8.1 ± 0.9	6.1 ± 0.7	6.0 ± 0.6	6.2 ± 0.6
	Isotropic	33 ± 5	7.8 ± 0.7	6.9 ± 0.5	6.8 ± 0.5	6.9 ± 0.5
	UAL	17 ± 4	0.35 ± 0.06	0.14 ± 0.03	0.11 ± 0.01	0.12 ± 0.01
0.5	Published	37 ± 5	11 ± 1	8.2 ± 0.9	7.9 ± 0.8	8.3 ± 0.8
	Isotropic	34 ± 4	16 ± 1	17 ± 1	17 ± 1	17 ± 1
	UAL	18 ± 4	0.58 ± 0.08	0.24 ± 0.04	0.21 ± 0.03	0.22 ± 0.03
1	Published	38 ± 5	15 ± 2	12 ± 1	11 ± 1	11 ± 1
	Isotropic	58 ± 2	40 ± 2	42 ± 3	42 ± 3	42 ± 3
	UAL	19 ± 4	1.2 ± 0.2	0.51 ± 0.07	0.46 ± 0.07	0.43 ± 0.07
1.5	Published	39 ± 5	19 ± 2	15 ± 2	15 ± 2	15 ± 2
	Isotropic	80 ± 3	58 ± 3	61 ± 4	62 ± 4	62 ± 4
	UAL	20 ± 4	2.0 ± 0.4	0.97 ± 0.16	0.99 ± 0.23	0.91 ± 0.16

D.2 Deblurring comparisons

At $(\bar{s}, \tau) = (3, 0.5)$, UAL increases per-trajectory PSNR for all four Gaussian-deblurring samplers (Table 1). The gains range from 0.39 dB for DAPS, with a 95% interval of $[0.20, 0.61]$, to 0.99 dB for DiffPIR, with interval $[0.59, 1.46]$. The fraction of images with higher mean per-trajectory PSNR ranges from 66% for DAPS to 96% for DiffPIR and BlindDPS. Mean CRPS decreases for all four samplers; Table 6 reports the paired confidence intervals.

In motion deblurring, UAL raises mean per-trajectory PSNR from 17.68 to 28.84 dB for DAPS, from 17.04 to 26.54 dB for PnP-DM, from 11.44 to 14.55 dB for DiffPIR, and from 22.76 to 29.10 dB for Fast Diffusion EM. With the true measurement kernel, the published PnP-DM sampler gives 27.65 dB (Table 8). UAL closes 89.5% of the per-trajectory PSNR gap from Published to this oracle reference, with 1.11 dB remaining. For ensemble-mean PSNR, UAL closes 90.3% of the gap. For DAPS, the oracle gives 30.65 dB mean per-trajectory PSNR; UAL closes 86.0% of the Published-to-Oracle PSNR gap, with 1.81 dB remaining. With the true kernel, the published DiffPIR sampler gives 30.90 dB mean per-trajectory PSNR; UAL closes 16.0% of its Published-to-Oracle PSNR gap (Table 8).

For BlindDPS in motion deblurring, UAL improves all three ensemble scores and raises ensemble-mean PSNR from 23.21 to 23.33 dB. Mean per-trajectory PSNR is 21.95 dB versus 22.12 dB for Published, with corresponding SSIM values of 0.5449 and 0.5559.

For DAPS, PnP-DM, and DiffPIR in motion deblurring, UAL raises mean per-trajectory PSNR on every image. Figures 5 and 6 show selected individual reconstructions for the same three images.

Comparison with blind samplers. In motion deblurring, we compare DAPS, PnP-DM, and Fast Diffusion EM with UAL against the published blind samplers BlindDPS and Fast Diffusion EM on the same 50 observations, including the same noise realizations. The blind samplers sample

Table 8: DAPS, PnP-DM, and DiffPIR motion deblurring on 50 images with 12 paired trajectories per image. Published uses the population-mean kernel; Oracle uses the true measurement kernel with the published sampler. UAL uses $P = 256$ covariance modes. T/M denotes mean per-trajectory score / score of the ensemble mean. PSNR is in dB; LVS is scaled by 10^3.

Sampler	Treatment	PSNR T/M ↑	SSIM T/M ↑	CRPS ↓	Energy ↓	LVS ↓
DAPS	Published	17.68 / 18.10	.3303 / .3779	.15043	96.013	57.076
	UAL	28.84 / 29.58	.7103 / .7726	.03502	23.388	12.774
	Oracle	30.65 / 32.50	.7698 / .8537	.02265	15.048	8.783
PnP-DM	Published	17.04 / 17.90	.2766 / .3897	.14485	91.587	61.032
	UAL	26.54 / 29.43	.5552 / .7628	.03236	20.943	14.959
	Oracle	27.65 / 30.68	.5752 / .7816	.02878	17.908	13.728
DiffPIR	Published	11.44 / 12.17	.0701 / .0877	.34406	172.270	179.800
	UAL	14.55 / 15.19	.1440 / .1666	.24688	130.857	100.688
	Oracle	30.90 / 33.27	.7936 / .8738	.02041	13.380	7.447

or estimate a kernel for each image; the UAL versions use the population-mean kernel and the propagated kernel covariance. Table 9 reports paired image-level bootstrap intervals with 20,000 resamples. All intervals favor UAL except per-trajectory SSIM for PnP-DM with UAL against BlindDPS, whose interval contains zero.

Table 9: Paired improvements over published blind samplers in motion deblurring, 50 images with 12 trajectories each: mean [95% paired bootstrap interval] and number of images improved. Values are UAL minus the blind baseline for PSNR and SSIM, and the blind baseline minus UAL for CRPS, energy, and LVS, so positive values denote improvement; CRPS and LVS are scaled by 10^3. The first method in each pair uses UAL.

Pair	PSNR M (dB)	PSNR T (dB)	SSIM M	SSIM T	CRPS	Energy	LVS
DAPS with UAL vs BlindDPS	+6.37 [5.81, 6.95] 50	+6.72 [6.25, 7.19] 50	+0.1463 [0.1330, 0.1606] 50	+0.1544 [0.1385, 0.1713] 50	+28.90 [25.12, 32.81] 50	+23.24 [20.35, 26.31] 50	+5.52 [4.51, 6.58] 48
DAPS with UAL vs Fast Diffusion EM	+6.31 [5.69, 6.87] 48	+6.08 [5.53, 6.60] 49	+0.2201 [0.1960, 0.2435] 48	+0.1868 [0.1621, 0.2114] 48	+42.87 [37.43, 48.27] 48	+28.48 [24.98, 31.91] 48	+11.50 [10.15, 12.83] 49
PnP-DM with UAL vs BlindDPS	+6.22 [5.69, 6.75] 50	+4.42 [3.98, 4.85] 49	+0.1364 [0.1237, 0.1500] 50	−0.0007 [−0.0199, 0.0190] 23	+31.55 [28.30, 34.94] 50	+25.69 [23.11, 28.33] 50	+3.33 [2.17, 4.45] 42
PnP-DM with UAL vs Fast Diffusion EM	+6.16 [5.58, 6.70] 50	+3.78 [3.25, 4.29] 47	+0.2103 [0.1859, 0.2338] 49	+0.0316 [0.0052, 0.0582] 31	+45.53 [40.55, 50.38] 50	+30.92 [27.58, 34.11] 50	+9.31 [7.73, 10.88] 47
Fast Diffusion EM with UAL vs BlindDPS	+7.52 [6.97, 8.05] 50	+6.98 [6.46, 7.51] 50	+0.1845 [0.1717, 0.1981] 50	+0.2076 [0.1941, 0.2215] 50	+36.84 [33.50, 40.45] 50	+28.30 [25.60, 31.16] 50	+8.66 [7.68, 9.74] 50
Fast Diffusion EM with UAL vs Fast Diffusion EM	+7.45 [6.83, 8.02] 50	+6.34 [5.71, 6.93] 48	+0.2583 [0.2329, 0.2832] 50	+0.2399 [0.2150, 0.2640] 49	+50.81 [45.32, 56.30] 50	+33.53 [30.09, 36.86] 50	+14.64 [13.21, 16.05] 50

D.3 Application details

In the main CT comparison, UAL gains 0.204 dB in per-trajectory PSNR, with a 95% interval of $[0.173, 0.236]$ dB, and improves SSIM and all three ensemble scores. The gain persists when both versions use the same mean flat and dark reference frames: UAL raises per-trajectory PSNR from 25.942 to 26.145 dB and lowers CRPS from 0.021955 to 0.021352. The paired PSNR gain is 0.2034 dB with a 95% interval of $[0.1735, 0.2356]$, and all 50 detector rows improve. Using those mean reference frames without UAL changes per-trajectory SSIM from 0.42478 to 0.42434, with a 95% paired interval of $[-0.00070, -0.00021]$ for the change. Figure 7 shows paired reconstructions and differences from the reference.

For MRI, mean per-trajectory PSNR improves on 47 of 50 scans, with a mean paired gain of 0.0127 dB and a 95% paired-bootstrap confidence interval of $[0.0097, 0.0160]$ dB (paired $t = 7.83$). The corresponding mean SSIM gain is 0.00031. All seven reported metrics favor UAL, with their pointwise 95% paired-bootstrap intervals excluding zero (Table 10). Figure 8 shows paired magnitude reconstructions.

In microscopy with five uncertain pupil-phase coefficients, UAL increases per-trajectory PSNR by 0.425 dB and SSIM by 0.0153. The mean PSNR gain is positive in each of the four source stacks, ranging from 0.251 to 0.724 dB. Using the ensemble mean, PSNR rises from 33.67 to 34.22 dB. The CRPS, energy, and local variogram scores decrease by 0.00148, 0.901, and 0.000881, respectively. The uncertainty model and synthetic measurement construction are specified in Appendix C.6. Figures 9 and 10 show paired reconstructions and ensemble spread. For the section in Figure 10, UAL reduces the root-mean-square error of the ensemble mean from 0.085 to 0.052, while the root-mean-square sample standard deviation changes from 0.063 to 0.059.

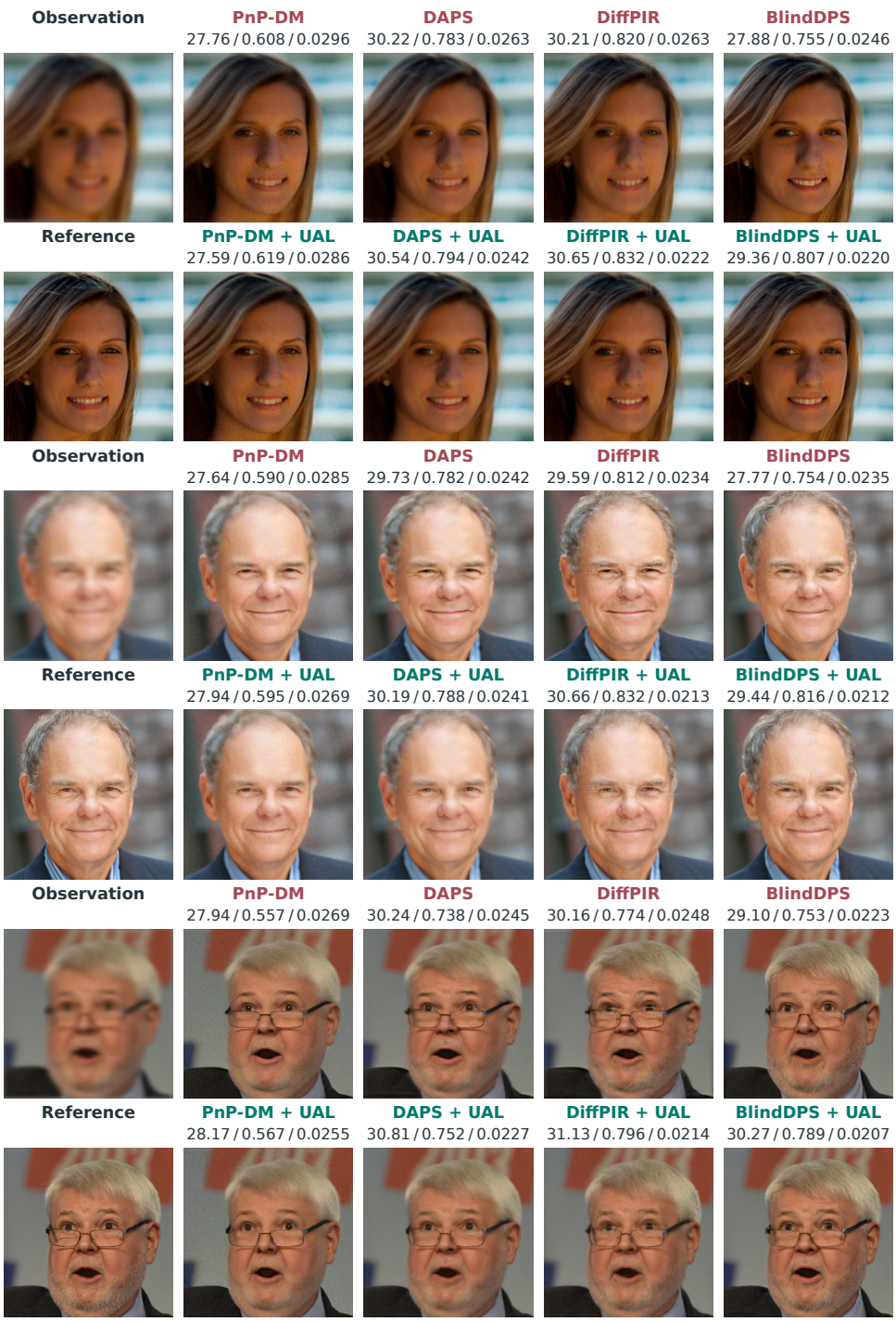

Figure 5: Selected Gaussian-deblurring reconstructions at $(\bar{s}, \tau) = (3, 0.5)$ pixels. Each row pair shows the observation and reference at left, followed by Published (upper) and UAL (lower) reconstructions from four samplers. Published and UAL use matched random seeds within each sampler. Numbers give PSNR (dB), SSIM, and CRPS, in that order. PSNR and SSIM describe the displayed trajectory; CRPS uses all 12 trajectories.

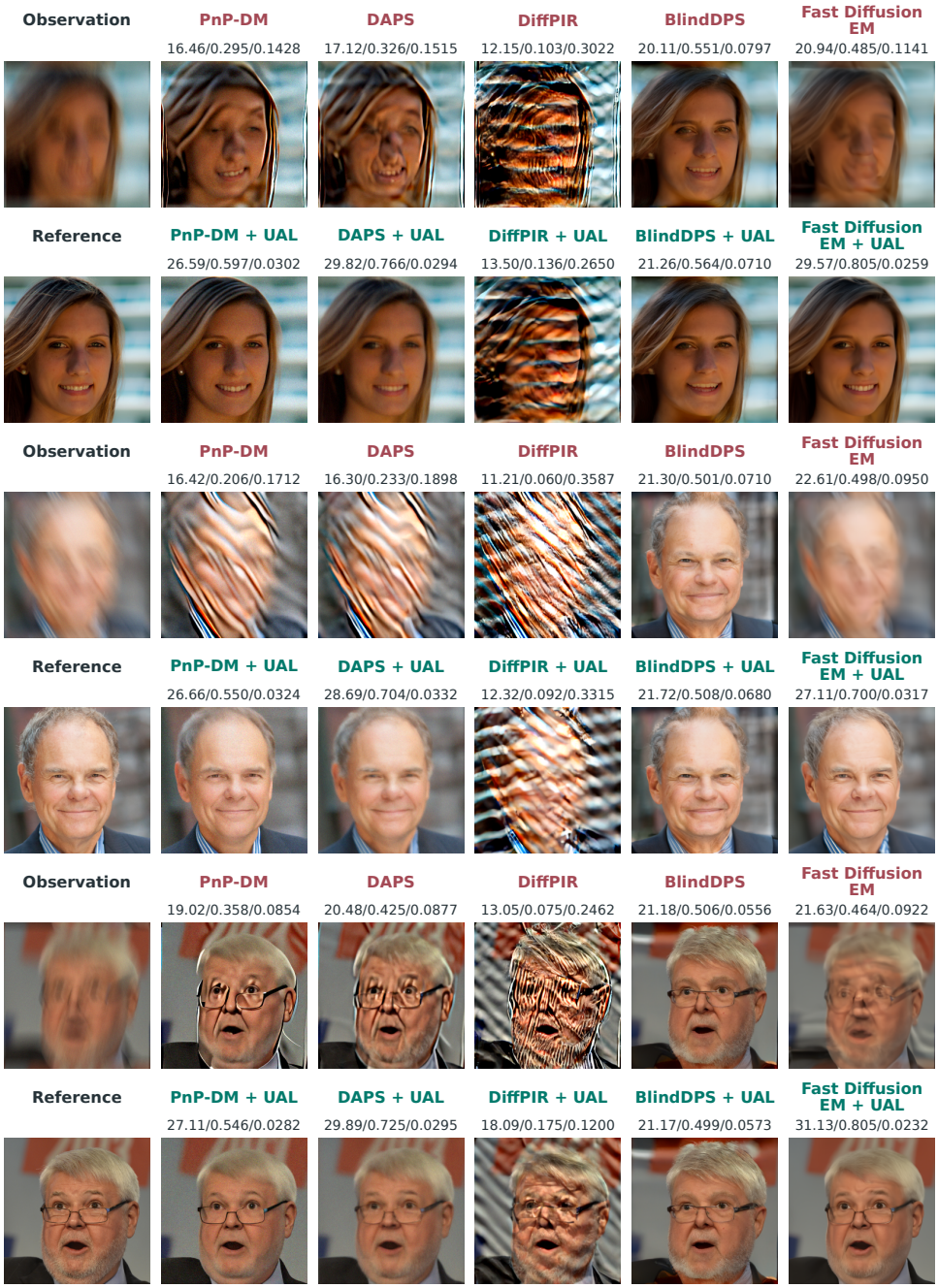

Figure 6: Selected motion-deblurring reconstructions for the same three images as Figure 5, with $P = 256$ covariance modes for UAL. Each row pair shows the observation and reference at left, followed by Published (upper) and UAL (lower) reconstructions. For each sampler, the displayed reconstructions are paired by random seed. Numbers give PSNR (dB), SSIM, and CRPS, in that order. PSNR and SSIM describe the displayed trajectory; CRPS uses all 12 trajectories.

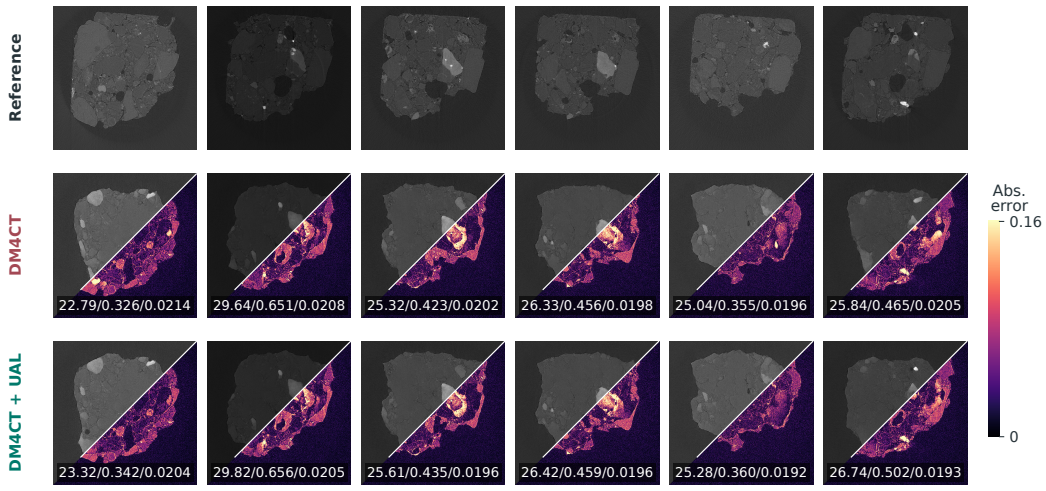

Figure 7: Selected CT cases, with reconstruction draws paired by seed. Columns show cases. Each method panel combines the reconstruction in the upper-left half with its absolute difference from the reference in the lower-right half. Difference maps share one color scale. Numbers give PSNR (dB), SSIM, and CRPS, in that order. PSNR and SSIM describe the displayed draw; CRPS uses all 12 trajectories. DM4CT denotes Published.

Table 10: MRI gains and pointwise 95% paired-bootstrap confidence intervals across 50 scans, each with 12 paired trajectories. Gains are UAL minus Published for PSNR and SSIM, and Published minus UAL for the ensemble scores. T/M follows Table 1.

Metric	Mean gain	95% interval
PSNR T (dB)	0.0127	$[0.0097, 0.0160]$
PSNR M (dB)	0.0141	$[0.0104, 0.0183]$
SSIM T	0.000313	$[0.000213, 0.000420]$
SSIM M	0.000119	$[0.000026, 0.000231]$
CRPS	0.0000331	$[0.0000239, 0.0000435]$
Energy	0.0220	$[0.0160, 0.0290]$
LVS ($\times 10^3$)	0.0116	$[0.0086, 0.0149]$

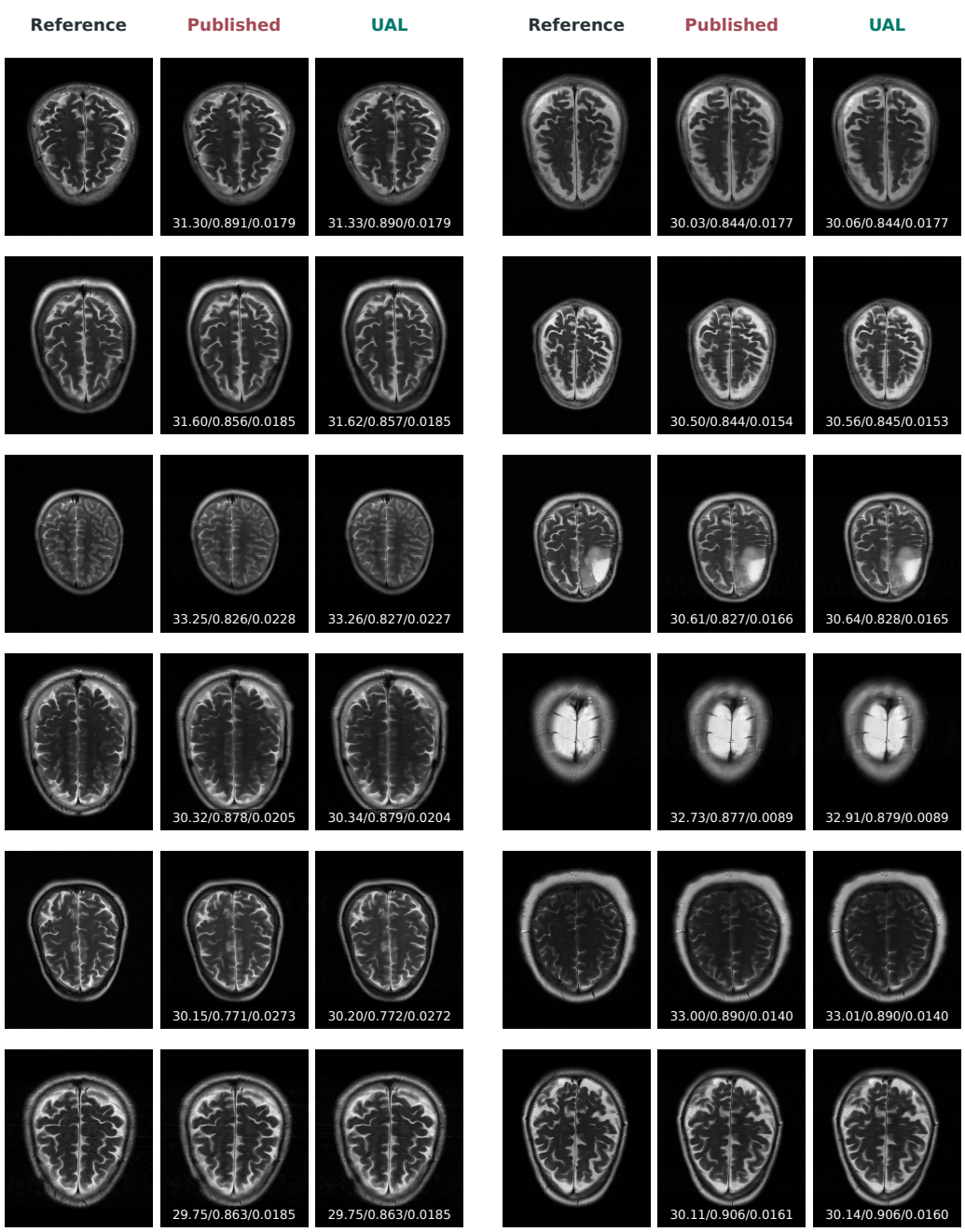

Figure 8: Magnitude reconstructions for selected MRI scans. Each triplet shows the reference, Published Ambient DPS, and Ambient DPS with UAL. Both methods use the same random seed for each scan. Numbers give PSNR (dB), SSIM, and CRPS, in that order. PSNR and SSIM describe the displayed draw; CRPS uses all 12 trajectories.

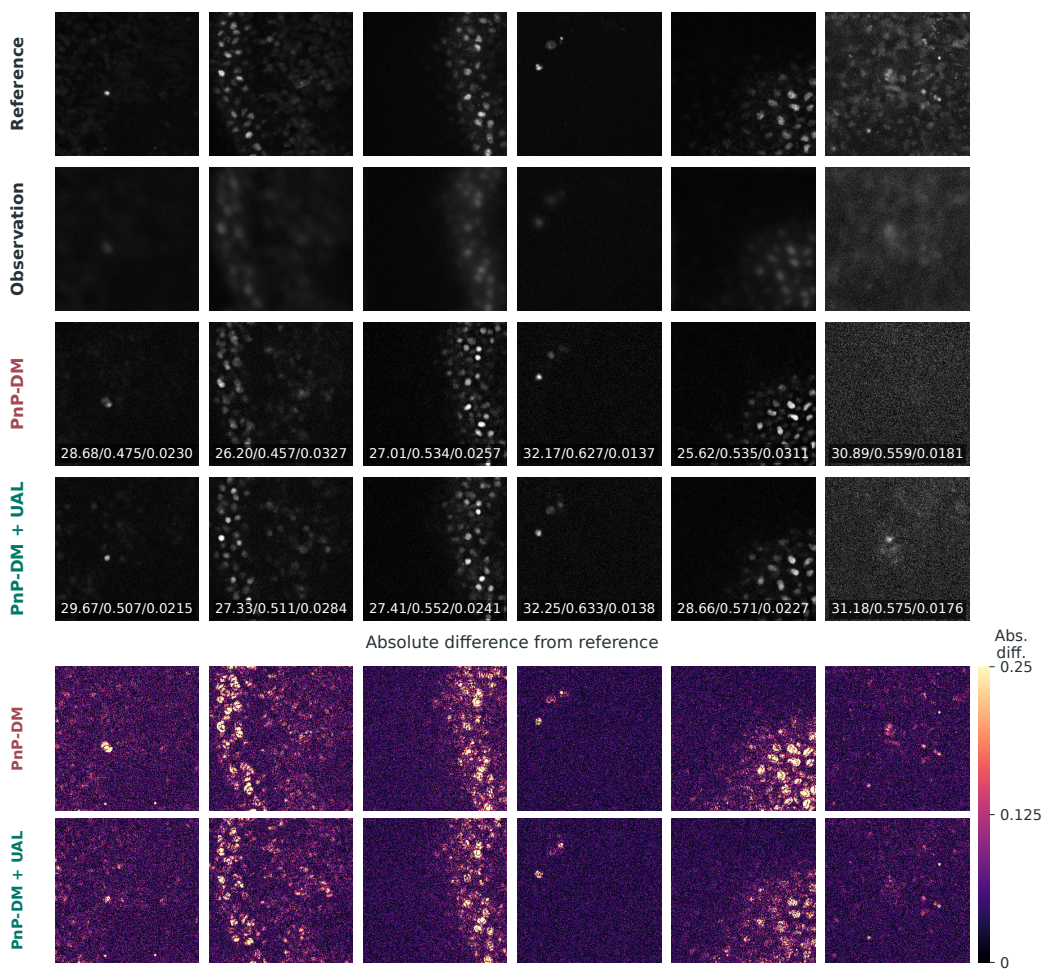

Figure 9: Selected fluorescence-microscopy sections. Reconstructions are paired by seed; difference maps share one scale. Numbers give PSNR (dB), SSIM, and CRPS. PSNR and SSIM use the displayed draw; CRPS uses all 12 trajectories.

Table 11: Native ultrasound contrast and resolution metrics, averaged over 12 trajectories per phantom. CNR, gCNR, and speckle SNR use the contrast phantom; axial and lateral FWHM use the resolution phantom. FWHM is in mm and lower values are better. Speckle SNR is compared with the reference 1.91. Bold marks better values.

Metric	Published	UAL
CNR (dB), higher	9.9292	**12.3625**
gCNR	0.9141	**0.9743**
Speckle SNR, target 1.91	2.0616	**2.0184**
Axial FWHM (mm)	0.2506	**0.2497**
Lateral FWHM (mm)	0.5729	**0.5722**

On the acquired PICMUS contrast phantom, UAL increases CNR and gCNR. On the resolution phantom, both mean FWHM changes are below 0.01 mm (Table 11). Figure 11 shows all 12 paired contrast-phantom reconstructions.

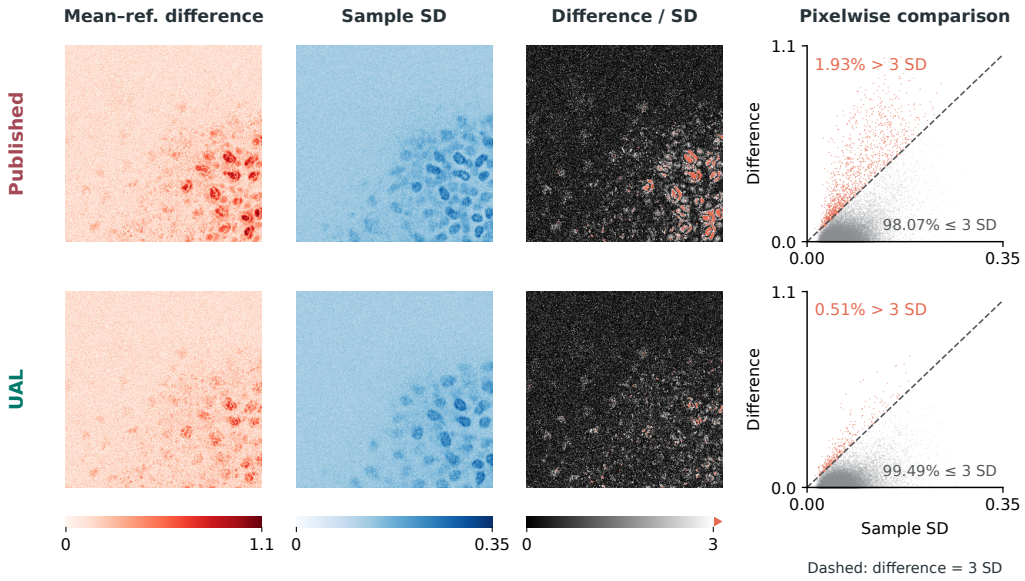

Figure 10: Pixelwise statistics for the microscopy section in Figure 9, using all 12 paired random seeds. The left columns show the absolute difference between the ensemble mean and reference and the sample standard deviation (SD). The right columns show their ratio and a scatter plot of difference against SD. Red marks pixels above the 3-SD threshold, indicated by the dashed line; percentages give the pixel fractions on either side. UAL reduces the fraction above this threshold from 1.93% to 0.51%, so large reference differences relative to the ensemble spread are less frequent in this section.

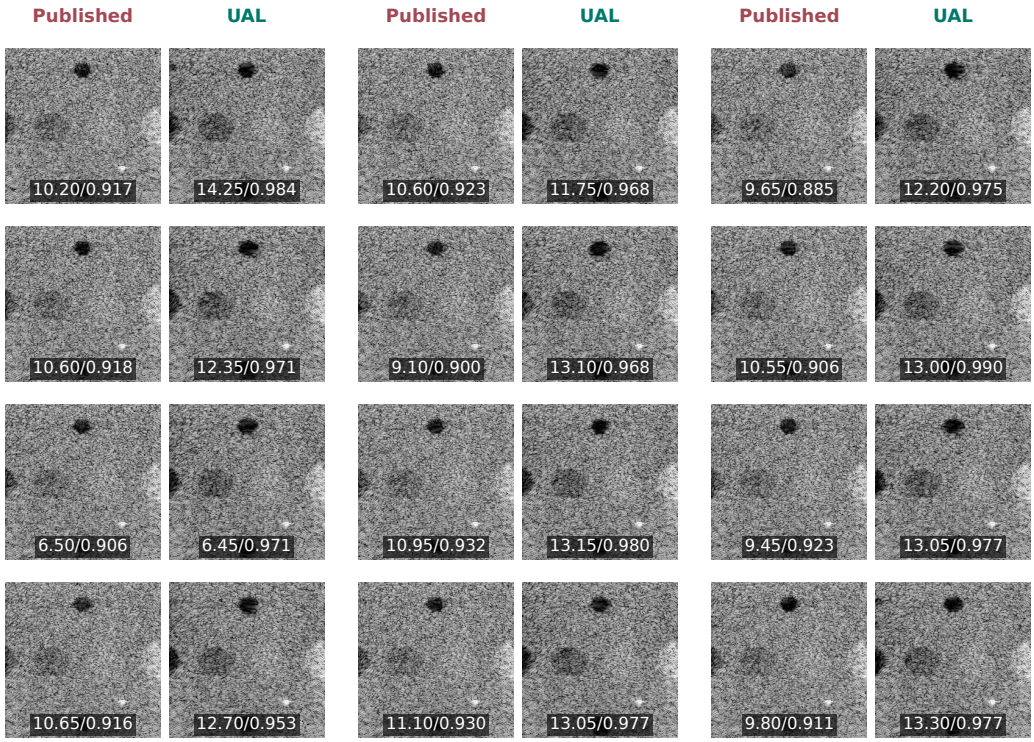

Figure 11: Paired DRUS reconstructions of the acquired PICMUS contrast phantom. Each pair shows Published on the left and UAL on the right, ordered by seed from left to right and top to bottom. Numbers give CNR (dB) and gCNR for the displayed reconstruction. All panels use the same $[-60, 0]$ dB B-mode display range.

D.4 COVARIANCE-STRUCTURE CONTROL AND SECONDARY METRICS

For Gaussian deblurring with PnP-DM at $(\bar{s}, \tau) = (3, 0.5)$, structured UAL improves reconstruction over an isotropic likelihood with the same whitened covariance trace. The isotropic covariance is

$$S_{\text{iso}}(x) = [1 + c(x)]\Sigma_e, \qquad c(x) = \frac{1}{m}\,\mathrm{tr}\big(\Sigma_e^{-1} J_\psi(x)\Sigma_\psi J_\psi(x)^\mathsf{T}\big), \tag{77}$$

where m is the measurement dimension. Here $c(x)$ is the average added variance after noise whitening, so $1 + c(x) = v_\star(x)$. Both likelihoods include Gaussian normalization and differentiate their signal-dependent covariance. Structured UAL gains 0.455 ± 0.197 dB over the isotropic control in per-trajectory PSNR and reduces CRPS by 0.00260 ± 0.00056 (mean $\pm$ standard error). The isotropic control produces little change from the published sampler (Table 13). Structured UAL improves per-trajectory PSNR over isotropic inflation on 13 of 16 Gaussian-deblurring images.

In motion deblurring, isotropic inflation improves mean per-trajectory PSNR by 6.212 ± 0.398 dB over the published sampler. Structured UAL adds 2.557 ± 0.317 dB in per-trajectory PSNR and reduces CRPS by 0.02706 ± 0.00336 relative to the isotropic control. The corresponding ensemble-mean PSNR margin is 4.431 ± 0.325 dB. Structured UAL improves per-trajectory PSNR over isotropic inflation on 15 of 16 motion-deblurring images. These results support using the nuisance-induced covariance directions to weight the residual.

In Gaussian deblurring with PnP-DM, recomputing covariance preserves the reconstruction gain when its derivatives are stopped. The stopped-derivative control removes both covariance terms in (6), so it tests their combined effect. Mean per-trajectory PSNR is 26.944 dB for the published sampler, 27.506 dB with covariance derivatives stopped, and 27.475 dB with UAL. Fixing covariance at the published sampler's ensemble mean gives 26.964 dB, with UAL higher by 0.512 ± 0.113 dB; uncertainties are standard errors across the 16 images. With fixed covariance, ensemble-mean PSNR is 29.088 dB and CRPS is 0.03484, compared with 29.388 dB and 0.03253 for the published sampler. Refreshing covariance with its derivatives stopped gives 30.304 dB and 0.02976, close to 30.259 dB and 0.02981 for UAL.

We also select a fixed noise-variance multiplier by mean per-trajectory PSNR. For the scalar noise-variance multipliers $v = 1, 1.5, 2, 3, 5$, mean per-trajectory PSNR is 26.944, 26.991, 27.007, 27.029, and 26.995 dB, respectively.

UAL exceeds the selected scalar control by 0.446 ± 0.123 dB, with higher PSNR on 15 of 16 images; the uncertainty is one standard error of the paired image differences. In motion deblurring, the best tested multiplier, $v = 48$, gives 23.369 dB mean per-trajectory PSNR and CRPS 0.05773, compared with 26.641 dB and 0.03149 for UAL (Table 12). UAL gives higher mean per-trajectory PSNR on 15 of 16 images. For DAPS, multiplying the data-term variance τ_{D}^2 by $v = 48$ gives the best tested value, 24.413 dB, compared with 28.631 dB for UAL; UAL is higher by 4.22 ± 0.27 dB (mean $\pm$ SE across the 16 images), with higher PSNR on every image.

Table 12: Fixed data-term variance multipliers in motion deblurring: mean per-trajectory PSNR (dB) on the ablation set. For PnP-DM, v scales the observation-noise variance; for DAPS, it scales the data-fitting variance τ_{D}^2. $v = 1$ is the published sampler; UAL gives 26.641 dB for PnP-DM and 28.631 dB for DAPS.

v	1	3	6	12	24	48	96	192	384
PnP-DM	17.872	20.429	21.652	22.569	23.243	23.369	23.210	22.503	21.419
DAPS	18.441	21.098	22.431	23.448	24.113	24.413	24.360	23.965	23.245

Table 13: Gaussian-deblurring covariance-structure control. T/M follows Table 1; PSNR is in dB, and CRPS and LVS are scaled by 10^3. Lower ensemble scores are better.

Likelihood	PSNR T	PSNR M	CRPS	Energy	LVS
Published	26.944	29.388	32.53	21.327	15.22
Isotropic	27.020	29.472	32.42	21.078	15.12
UAL	**27.475**	**30.259**	**29.81**	**19.052**	**14.18**

For PnP-DM in motion deblurring, including the log determinant improves all seven metric means over the quadratic-only control (Table 14). This control uses $P = 256$ and the ablation set, as the covariance-structure comparison. We retain $\frac{1}{2}d(x)^\mathsf{T}S(x)^{-1}d(x)$, recomputing and differentiating its covariance at every candidate. This is the penalized profile energy described in Appendix A.2. UAL has higher mean per-trajectory PSNR on 15 of 16 images, with a paired advantage of 0.0257 ± 0.0066 dB (mean $\pm$ SE across images).

Table 14: Likelihood normalization in motion deblurring with PnP-DM at $P = 256$. Values are means $\pm$ one SE across the ablation set. Quadratic only omits the log determinant while retaining covariance recomputation and differentiation. T/M follows Table 1.

Metric	Published	Quadratic only	UAL
PSNR T (dB) ↑	17.8723 ± 0.7269	26.6157 ± 0.3040	26.6414 ± 0.3013
PSNR M (dB) ↑	18.7901 ± 0.7827	29.5099 ± 0.3551	29.5335 ± 0.3564
SSIM T ↑	0.30950 ± 0.02466	0.55399 ± 0.00805	0.55412 ± 0.00806
SSIM M ↑	0.43340 ± 0.03894	0.76189 ± 0.01441	0.76191 ± 0.01444
CRPS ($\times 10^3$) ↓	131.533 ± 19.396	31.534 ± 1.491	31.489 ± 1.496
Energy ↓	83.874 ± 9.441	20.406 ± 0.882	20.351 ± 0.884
LVS ($\times 10^3$) ↓	54.395 ± 7.343	14.790 ± 0.572	14.766 ± 0.564

For Fast Diffusion EM in motion deblurring, UAL improves every reported metric on all 50 images relative to covariance weighting alone (Table 15).

Table 15: Motion deblurring with Fast Diffusion EM on 50 images with 12 paired trajectories per image and $P = 256$. All three adapted arms use the population-mean kernel and propagated covariance; covariance weighting holds the covariance fixed during differentiation and omits the log determinant, and the quadratic-only arm recomputes and differentiates the covariance but omits the log determinant. T/M follows Table 1.

Metric	Published	Covariance weighting	Quadratic only	UAL
PSNR T (dB) ↑	22.76	23.80	28.49	29.10
PSNR M (dB) ↑	23.27	25.30	30.64	30.72
SSIM T ↑	.5236	.5663	.7537	.7635
SSIM M ↑	.5525	.6243	.8131	.8108
CRPS ↓	.07789	.05472	.02731	.02708
Energy ↓	51.865	37.077	18.360	18.332
LVS ($\times 10^3$) ↓	24.271	17.248	9.449	9.633

D.5 Covariance spectra at final reconstructions

We evaluate UAL's covariance at the final Published, Isotropic, and UAL reconstructions, using the same parameter mean and covariance for all three. With $W = \Sigma_e^{-1/2}U$ and $G = I_r + W^\mathsf{T}W$, we report the average added variance after noise whitening, $c(x) = (\operatorname{tr} G - r)/m$, and the two leading squared singular values s_1^2 and s_2^2 of W. The trace-matched variance multiplier is $1 + c(x)$.

Each quantity is evaluated separately for each trajectory and averaged within an image; the table reports means and standard errors across images. Singular values are ordered within each trajectory before aggregation.

The Gaussian covariance concentrates its added variance in two directions, so a small average variance increase can accompany large directional variances. In motion deblurring, the added covariance trace differs across the final candidate reconstructions. These diagnostics use measurement-noise whitening; the reconstruction-risk Gram matrix Λ in Appendix B uses predictive-covariance whitening.

D.6 Covariance scale and matched uncertainty

Mean and covariance mismatch. With observations fixed and covariance correct, shifting the mean width by approximately $\pm5\%$ and $\pm20\%$ gives mean per-trajectory PSNR from 27.408 to

Table 16: Final-state covariance diagnostics, reported as mean ± standard error across images. The leading squared singular values are scaled by 10^{-4}.

Task	Candidate	$c(x)$	$s_1^2/10^4$	$s_2^2/10^4$
Gaussian blur	Published	0.189 ± 0.015	2.617 ± 0.266	1.098 ± 0.048
	Isotropic	0.185 ± 0.014	2.553 ± 0.248	1.088 ± 0.047
	UAL	0.197 ± 0.010	2.516 ± 0.134	1.357 ± 0.065
Motion blur	Published	20.085 ± 3.532	176.389 ± 44.334	57.841 ± 8.578
	Isotropic	3.244 ± 0.328	34.666 ± 4.084	12.922 ± 1.438
	UAL	4.626 ± 0.370	38.296 ± 3.533	19.090 ± 1.942

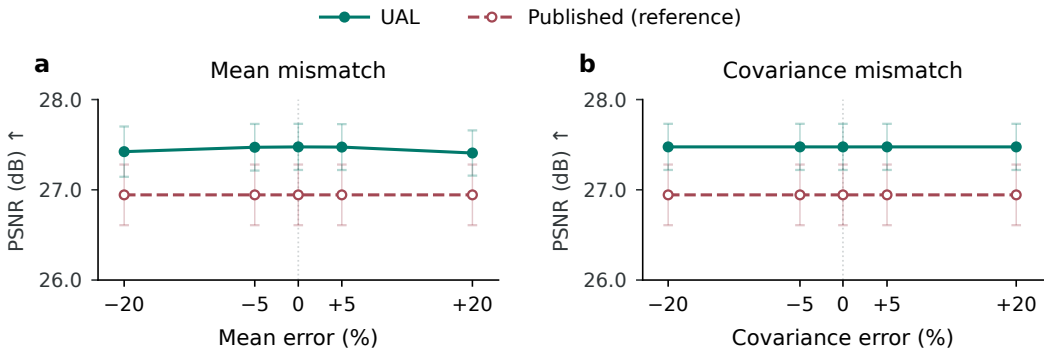

Figure 12: PnP-DM sensitivity to nuisance moments in Gaussian deblurring. (a) Mean-width error with correct covariance. (b) Covariance-scale error with correct mean. Observations stay fixed; zero denotes the correct moment. Curves show mean per-trajectory PSNR ± one SE across the ablation set. The dashed Published reference uses the correct mean in both panels. Mean PSNR changes little over the tested ranges.

27.475 dB (Figure 12a). CRPS is 0.03099 at mean width 2.4 pixels and 0.02981 at the correct mean of 3 pixels. We next perturb the covariance used by UAL while keeping the generating covariance and observations fixed. Gaussian deblurring uses the reference $\Sigma_{\psi,\mathrm{ref}} = 0.25 I_2$ in squared-pixel units. Each perturbed covariance is a scalar multiple of its task's reference. Table 17 reports paired changes from UAL with zero covariance error. We also test approximately doubled nuisance covariance. Across the nonzero Gaussian covariance settings, mean per-trajectory PSNR remains 27.475 dB to three decimals. For motion deblurring, removing the nuisance covariance gives the published sampler's PSNR of 17.872 dB; doubling it gives 26.644 dB, compared with 26.641 dB at the reference covariance.

Over the approximately $\pm 20\%$ covariance-error range, mean per-trajectory PSNR changes by less than 0.001 dB in Gaussian deblurring and at most 0.0641 dB in motion deblurring. The largest absolute relative CRPS changes are 0.0011% and 0.0572%, respectively. One trajectory accounts for most of the motion PSNR change; all trajectories remain in the paired summaries.

Matched uncertainty levels. We next vary the generating width uncertainty and the uncertainty supplied to UAL together, using $\mu = (3, 3)$ and $\Sigma_\psi = \tau^2 I_2$. Image identities, width percentiles, and noise realizations are coupled across levels; the observations change with the blur widths. Relative to Published, UAL's mean per-trajectory PSNR gain is 0.043 ± 0.067 dB at $\tau = 0.25$. It grows to 1.746 ± 0.397 dB at $\tau = 1.5$, while the reduction in per-trajectory MSE grows from 0.000099 to 0.007808 (Table 18). At the largest tested uncertainty, UAL reduces average per-trajectory MSE by approximately 38%. Table 18 also lists the trace-matched isotropic control of (77). UAL gives lower mean MSE, higher mean per-trajectory PSNR, and lower CRPS than isotropic inflation at every tested level.

Table 17: Covariance-scale mismatch with fixed observations. Each task uses the ablation set. Signed changes (perturbed minus UAL at zero covariance error) are reported as mean ± SE across images. PSNR is in dB. Relative CRPS changes divide the paired change and its SE by the reference mean CRPS.

Task	Covariance error	ΔPSNR T	ΔPSNR M	ΔCRPS (%)
Gaussian	-20.1%	$+0.0002 \pm 0.0002$	$+0.0001 \pm 0.0002$	-0.0003 ± 0.0008
Gaussian	-4.9%	$+0.0002 \pm 0.0002$	$+0.0002 \pm 0.0001$	-0.0006 ± 0.0007
Gaussian	$+5.1\%$	$+0.0002 \pm 0.0001$	$+0.0002 \pm 0.0001$	-0.0010 ± 0.0006
Gaussian	$+19.9\%$	$+0.0001 \pm 0.0001$	$+0.0002 \pm 0.0002$	-0.0011 ± 0.0008
Gaussian	$+99.9\%$	$+0.0001 \pm 0.0001$	$+0.0002 \pm 0.0002$	-0.0009 ± 0.0010
Motion	-20.0%	$+0.0609 \pm 0.0633$	$+0.0257 \pm 0.0291$	$+0.0572 \pm 0.0392$
Motion	-5.0%	-0.0005 ± 0.0002	-0.0006 ± 0.0002	$+0.0040 \pm 0.0024$
Motion	$+5.0\%$	$+0.0004 \pm 0.0002$	$+0.0005 \pm 0.0003$	-0.0043 ± 0.0026
Motion	$+20.0\%$	$+0.0641 \pm 0.0634$	$+0.0308 \pm 0.0292$	$+0.0252 \pm 0.0366$
Motion	$+100.0\%$	$+0.0030 \pm 0.0034$	$+0.0108 \pm 0.0078$	-0.0495 ± 0.0404

The comparison shows how the empirical benefit changes when generating uncertainty and the uncertainty used by UAL vary together; the conditional crossover in Appendix B instead holds the nuisance model q fixed and varies the nuisance realization.

Table 18: Matched generating and UAL width uncertainty, on the ablation set. MSE and PSNR (dB) are averaged per trajectory; CRPS is scaled by 10^3. Isotropic is the trace-matched control of (77), run at each level.

	MSE ↓			PSNR ↑			CRPS ↓		
τ (pixels)	Published	Isotropic	UAL	Published	Isotropic	UAL	Published	Isotropic	UAL
.25	.007301	.007312	.007202	27.508	27.502	27.551	29.93	29.97	29.23
.50	.008459	.008263	.007348	26.944	27.020	27.475	32.53	32.42	29.81
1.00	.014066	.010826	.008800	25.368	26.018	26.818	40.95	38.24	33.26
1.50	.020463	.012810	.012655	24.043	25.318	25.789	49.58	43.24	39.90

D.7 Full Gaussian-width results

Across the three Gaussian-width settings, UAL gives higher mean per-trajectory PSNR and lower CRPS than Published for all four samplers (Table 19). These comparisons use all 50 images and 12 paired trajectories per image; Table 1 gives all seven scores for the main setting; Table 20 gives them for the other two.

Table 19: Gaussian-width effects on reconstruction and ensemble prediction on 50 images, with 12 paired trajectories per image. Entries are per-trajectory PSNR (dB) and CRPS ($\times 10^3$) for the published sampler and UAL; $(\bar{s}, \tau)$ is in pixels. Bold marks better UAL values at the displayed precision.

		DAPS		PnP-DM		DiffPIR		BlindDPS	
$(\bar{s}, \tau)$	Version	PSNR ↑	CRPS ↓	PSNR ↑	CRPS ↓	PSNR ↑	CRPS ↓	PSNR ↑	CRPS ↓
$(3, 0.5)$	Published	29.49	28.71	26.97	32.09	29.13	29.48	28.31	26.98
	UAL	**29.88**	**27.96**	**27.50**	**29.56**	**30.12**	**25.83**	**29.21**	**24.76**
$(5, 0.5)$	Published	28.16	32.89	26.35	35.58	27.52	32.67	26.96	32.38
	UAL	**28.65**	**32.22**	**26.84**	**33.93**	**28.60**	**28.64**	**27.37**	**29.79**
$(3, 1)$	Published	28.05	35.42	25.80	38.56	26.70	43.55	28.43	27.01
	UAL	**29.56**	**30.13**	**27.16**	**31.03**	**29.63**	**29.07**	**29.00**	**25.73**

Table 20: Full Gaussian-width results for $(5, 0.5)$ and $(3, 1)$ pixels on 50 images, with 12 paired trajectories per image. T/M follows Table 1; PSNR is in dB and LVS is scaled by 10^3. Bold marks better UAL values at the displayed precision.

Sampler	Version	PSNR T/M ↑	SSIM T/M ↑	CRPS ↓	Energy ↓	LVS ↓
$(5, 0.5)$ pixels						
DAPS	Published	28.16 / 29.32	.6888 / .7685	.03289	22.999	12.581
	UAL	**28.65 / 29.67**	**.6949 / .7702**	**.03222**	**22.512**	12.631
PnP-DM	Published	26.35 / 28.55	.5325 / .7290	.03558	23.459	16.258
	UAL	**26.84 / 29.12**	**.5366 / .7338**	**.03393**	**22.008**	**15.759**
DiffPIR	Published	27.52 / 29.11	.7028 / .7712	.03267	22.846	11.400
	UAL	**28.60 / 30.07**	**.7278 / .7891**	**.02864**	**20.396**	**10.456**
BlindDPS	Published	26.96 / 28.78	.6732 / .7436	.03238	23.167	11.609
	UAL	**27.37 / 29.45**	**.6955 / .7714**	**.02979**	**21.120**	**10.365**
$(3, 1)$ pixels						
DAPS	Published	28.05 / 29.02	.7113 / .7839	.03542	25.410	13.712
	UAL	**29.56 / 30.56**	**.7347 / .8041**	**.03013**	**20.679**	**11.525**
PnP-DM	Published	25.80 / 28.13	.5388 / .7370	.03856	25.832	18.788
	UAL	**27.16 / 29.96**	**.5639 / .7683**	**.03103**	**19.965**	**14.864**
DiffPIR	Published	26.70 / 27.78	.7114 / .7667	.04355	31.323	18.119
	UAL	**29.63 / 30.67**	**.7684 / .8130**	**.02907**	**20.319**	**10.210**
BlindDPS	Published	28.43 / 30.37	.7381 / .8015	.02701	19.286	9.725
	UAL	**29.00 / 30.88**	**.7622 / .8202**	**.02573**	**18.312**	**9.155**

D.8 Covariance rank and computational cost

Table 22 reports mean wall time per trajectory on NVIDIA L40S GPUs, with one reconstruction job per GPU. DAPS deblurring times cover the sampling call, excluding setup. DiffPIR includes process initialization and output writing; DRUS excludes model loading, singular value decomposition, and covariance preparation. For Figure 4, reconstruction scores use the ablation set. Table 21 compares Published and UAL at $P = 256$ and $P = 2048$ on the ablation set to assess sensitivity to rank. At both ranks, UAL improves mean per-trajectory PSNR for the three non-blind samplers and lowers the mean CRPS, energy, and local variogram scores for all four samplers.

Table 21: Motion deblurring with the published samplers and UAL. All treatments use the ablation set. T/M follows Table 1; PSNR is in dB. CRPS, energy, and LVS are fair ensemble scores, with LVS scaled by 10^3.

Sampler	Treatment	PSNR T/M ↑	SSIM T/M ↑	CRPS ↓	Energy ↓	LVS ↓
DAPS	Published	18.44 / 18.89	.3677 / .4190	0.13913	88.578	52.203
	UAL, $P = 256$	28.63 / 29.34	.7030 / .7660	0.03572	24.046	13.079
	UAL, $P = 2048$	28.56 / 29.29	.7002 / .7659	0.03568	23.962	12.990
PnP-DM	Published	17.87 / 18.79	.3095 / .4334	0.13153	83.874	54.395
	UAL, $P = 256$	26.64 / 29.53	.5541 / .7619	0.03149	20.351	14.766
	UAL, $P = 2048$	26.25 / 29.31	.5420 / .7517	0.03209	20.788	14.997
DiffPIR	Published	11.72 / 12.50	.0853 / .1057	0.32936	165.774	170.151
	UAL, $P = 256$	15.19 / 15.87	.1681 / .1930	0.22664	120.672	90.039
	UAL, $P = 2048$	16.78 / 17.52	.2184 / .2507	0.18415	101.031	68.612
BlindDPS	Published	22.50 / 23.61	.5747 / .6418	0.06109	44.357	17.561
	UAL, $P = 256$	22.20 / 23.62	.5593 / .6391	0.06048	42.963	16.822
	UAL, $P = 2048$	22.38 / 23.78	.5616 / .6392	0.05955	42.368	16.759

Table 22: Mean wall time per trajectory on NVIDIA L40S GPUs with one reconstruction job per GPU.

Sampler / task	Published (s)	UAL (s)
Gaussian deblurring		
PnP-DM	77.3	76.9
DAPS	39.4	116.6
DiffPIR	6.0	6.1
BlindDPS	129.6	371.8
Motion deblurring		
PnP-DM	31.8	89.5
DAPS	21.8	155.4
DiffPIR	6.1	6.4
BlindDPS	126.4	70.7
Imaging applications		
DAPS (CT)	429.8	436.2
Ambient DPS (MRI)	52.9	103.2
DRUS (ultrasound)	74.3	73.9
PnP-DM (microscopy)	30.5	83.5