

The Role of Geological Prior Information in Diffusion-Based Probabilistic Gravity Inversion

Omid Mahmoodi

o.mahmoodi@icloud.com

Independent Researcher

Research Article

Keywords: gravity inversion, geological priors, diffusion models, posterior sampling, uncertainty quantification, structural evaluation

Posted Date: October 9th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11134959/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

The Role of Geological Prior Information in Diffusion-Based Probabilistic Gravity Inversion

Omid Mahmoodi

Independent Researcher, Mountain View, California, USA

Correspondence: o.mahmoodi@icloud.com

Abstract

Gravity inversion is inherently non-unique, so the recovered density distribution depends strongly on geological prior assumptions. Diffusion models can encode complex geological structure, but the effect of how much geological information is available before inversion remains insufficiently quantified. We isolate this effect within one measurement-guided diffusion workflow by comparing three information states: an unconditional prior for unknown geology, a correctly family-conditioned prior for known geology, and an equal-weight ensemble across six plausible families for uncertain geology. Experiments use six procedurally generated geological families and 100 fixed in-distribution (ID) plus 100 fixed within-family out-of-distribution (OOD) cases. Relative to a smooth SimPEG L2 baseline, all learned-prior settings reduce mean root-mean-square error (RMSE) on both splits. Correct family conditioning provides the strongest structural recovery, with the highest mean structural similarity and anomaly overlap, whereas equal-weight scenario pooling achieves the lowest ID RMSE but produces smoother aggregate estimates. The scenario ensemble also gives the broadest empirical interval coverage, although all finite ensembles remain under-dispersed. These results show that geological-prior information changes reconstruction error, recovered structure, and ensemble spread. The inversion favored by RMSE can differ from that favored by structure-aware metrics, so geological prior information should be treated explicitly when evaluating probabilistic gravity inversion.

Keywords gravity inversion; geological priors; diffusion models; posterior sampling; uncertainty quantification; structural evaluation

1 Introduction

Gravity inversion estimates subsurface density contrasts from potential-field observations but is inherently non-unique; prior assumptions therefore materially influence the recovered model. Conventional deterministic approaches commonly stabilize the inverse problem with smoothness or minimum-structure regularization. Such constraints are computationally effective but do not directly encode contacts, faults, dykes, or compact intrusive geometries. Modular inversion frameworks such as SimPEG make these assumptions explicit and reproducible [1].

Conceptual geology has long been introduced through training images, multiple-point statistics, morphology summaries, and learned latent priors. Training-image statistics can constrain posterior morphology even when multiple models fit the measurements [2], while assembled generative priors can represent several plausible spatial concepts and reduce

bias associated with selecting one conceptual pattern in advance [3]. These studies establish geological-prior specification as a central component of probabilistic inversion.

Diffusion probabilistic models learn complex data distributions through progressive noising and denoising [4, 5]. Pretrained diffusion priors have subsequently been coupled to physical measurements in inverse problems through methods including denoising diffusion restoration models (DDRM) [6] and diffusion posterior sampling (DPS) [7]. In geophysics, geological and well-log conditioning has been applied to full-waveform inversion (FWI) [8], diffusion-generated priors have supported Bayesian FWI [9], and diffusion networks have been developed for supervised gravity inversion [10].

These developments establish diffusion-based inversion and geological conditioning, but they do not isolate how the amount of geological information available before inversion changes posterior behavior. Here, geological prior information means knowledge of the conceptual geological family from which the density model is assumed to originate. We study three information states under an otherwise identical inversion workflow: unknown geology, known geology, and multiple equally plausible geological interpretations (Fig. 1).

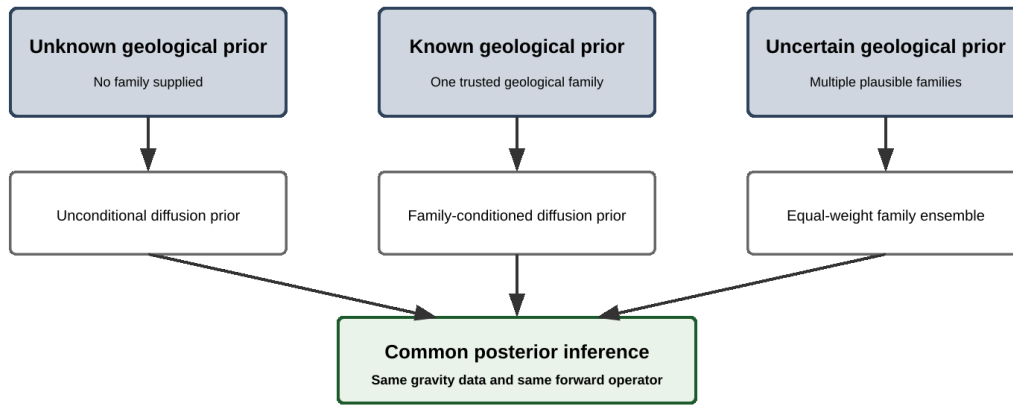


Fig. 1 Three geological-information states investigated under a common gravity-inversion framework. The observations, forward operator, and posterior sampler are fixed; only the information used to define the geological prior changes across the three experiments

This study makes four contributions: (1) a controlled hierarchy of geological-information states implemented with fixed unconditional and conditional diffusion priors; (2) a common measurement-guided gravity-inversion benchmark with ID and OOD tests; (3) a simple equal-weight scenario-ensemble baseline for unresolved geological interpretation; and (4) an evaluation showing that pixelwise RMSE and structure-aware measures such as the structural similarity index (SSIM) [11] and Intersection over Union (IoU) can rank posterior estimates differently.

2 Related work

2.1 Geological priors in probabilistic inversion

Training-image and geostatistical approaches encode conceptual structure that cannot be represented by generic smoothness penalties. Lochbühler et al. used frequency-domain parameter reduction and summary statistics derived from training-image realizations to constrain Bayesian inversion [2]. Lopez-Alvis et al. later demonstrated that a variational autoencoder trained on several plausible spatial patterns can act as an assembled prior, preserving broader conceptual uncertainty than a single-pattern prior [3]. Our equal-weight conditional ensemble is conceptually related, but it preserves explicit family labels and pools family-specific reconstruction ensembles rather than blending all concepts into one latent prior. Because the family weights are fixed rather than inferred from the observations, it is not an evidence-weighted Bayesian model average.

2.2 Diffusion priors for inverse problems

Denoising diffusion probabilistic models (DDPMs) and score-based stochastic differential equations provide generative mechanisms for sampling learned high-dimensional distributions [4, 5]. DDRM exploits a pretrained diffusion prior for linear image-restoration problems [6]. DPS extends measurement-guided sampling to noisy linear and nonlinear inverse problems by approximating the likelihood score through a denoised estimate [7]. We adopt DPS as the inference mechanism but use a delayed and bounded guidance update for numerical stability.

2.3 Diffusion models in geophysical inversion

Recent geophysical studies have introduced geological and well information into conditional diffusion models for FWI [8], used diffusion-generated priors to accelerate Bayesian FWI [9], and developed a supervised diffusion network for gravity inversion [10]. Our contribution is not the first use of diffusion or geological conditioning. It is the controlled comparison of unknown, known, and uncertain geological-prior states within one gravity-inversion framework.

3 Methodology

3.1 Problem formulation

Let m denote a discretized density-contrast model and d the observed gravity anomaly. The forward problem is

$$d = Gm + \varepsilon \quad (1)$$

where G is the gravity sensitivity operator and ε is additive Gaussian noise. The posterior is

$$p(m | d) \propto p(d | m)p(m) \quad (2)$$

The central experimental variable is the geological information used to define $p(m)$.

3.2 Synthetic geological benchmark

The benchmark contains six single-family geological classes: blobs, horizontal layers, dipping layers, faulted layers, dykes, and intrusions (Fig. 2). These families were selected to span three morphology axes that materially affect gravity inversion: distributed versus compact support, subhorizontal versus inclined boundaries, and continuous versus disrupted interfaces. Blobs and intrusions represent compact bodies with irregular and coherent shapes; dykes represent narrow, high-aspect-ratio targets; horizontal and dipping layers represent laterally extensive interfaces with different orientations; and faulted layers introduce discontinuous stratigraphy. The set is not intended to represent the full diversity of natural geology. It provides distinct structural regimes in which the effect of family information can be tested under a common forward problem.

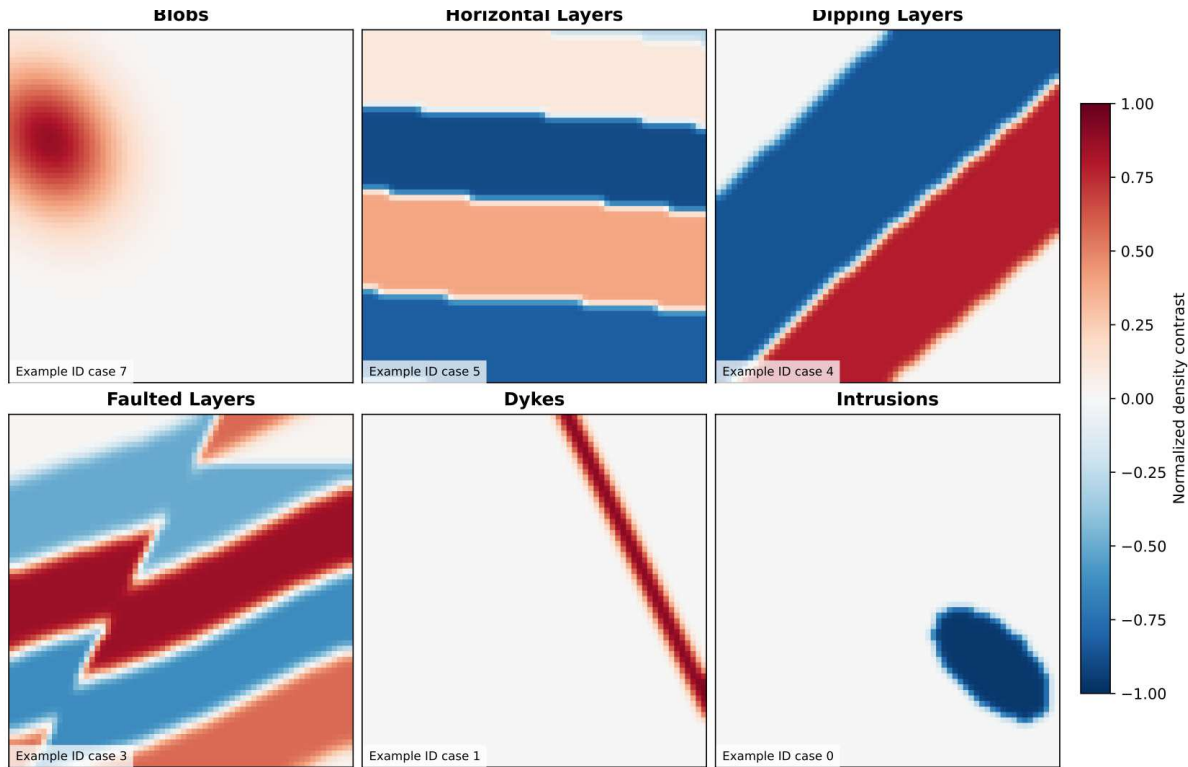


Fig. 2 Representative realizations from the six-family benchmark. The families span compact, elongated, layered, inclined, and fault-disrupted morphologies while retaining one family label per realization

The six families were produced with randomized procedural generators. Each family was divided into independent training, validation, ID, and OOD sets. The complete generated dataset contained 12,000 training realizations, 1,000 validation realizations, 1,000 ID test realizations, and 1,000 OOD test realizations. The inversion benchmark used 100 fixed ID and 100 fixed OOD cases on a 64×64 grid, with family counts ranging from 12-20 cases in ID and 10-25 cases in OOD. Overall metrics therefore summarize the realized benchmark composition rather than an equal-weight average across families; family-stratified results are reported separately to expose heterogeneity.

The OOD sets were generated by shifting procedural-generator parameter ranges beyond the corresponding training ranges while retaining the same conceptual family. Consequently, OOD performance tests within-family geometric extrapolation, not generalization to unseen geological concepts. This distinction limits the scope of the OOD conclusions and is retained throughout the interpretation.

3.3 Unconditional and conditional diffusion priors

The unconditional model learns $p\theta(m)$ from all six families without labels. The conditional model learns $p\phi(m | g)$, where g is a family label. Both use the same U-Net backbone with base width 32 and time-embedding dimension 128; the conditional network additionally uses a 128-dimensional class embedding. Both were trained for 200 epochs with AdamW, batch size 64, learning rate 2×10^{-4} , and weight decay 10^{-6} . A 1,000-step linear variance schedule from 10^{-4} to 2×10^{-2} was used, and the checkpoint with the lowest validation loss was retained.

3.4 Measurement-guided posterior inference

Posterior samples were generated with a stabilized implementation of DPS. At reverse step t , the denoised estimate was passed through the gravity operator and the data-misfit gradient was backpropagated to the noisy state. Guidance was applied only below timestep 600 with coefficient $\zeta = 0.05$. The root-mean-square guidance update was capped at 0.5 times the prior update, and density values were bounded to $[-1, 1]$. These are numerical stabilization choices rather than a new sampling algorithm. The DPS configuration was fixed before the final benchmark evaluation and was not tuned by case or geological family.

3.5 Geological-information settings and controlled comparison

For unknown geology, inference uses $p\theta(m)$. For known geology, the true synthetic family is supplied and inference uses $p\phi(m | g)$; we refer to this setting as the oracle case because the correct family label is known in the synthetic benchmark. For uncertain geology, the conditional sampler is run under all six families and samples are pooled with fixed equal weights

$$q_{\text{ens}}(m | d) = \frac{1}{6} \sum_{g=1}^6 q_g(m | d) \quad (3)$$

where q_g is the approximate DPS reconstruction distribution under family g . Twenty-four samples were generated for each diffusion setting; the scenario ensemble used four samples per family. The equal-weight pool is a geological-scenario sensitivity analysis rather than Bayesian posterior marginalization, which would require data-dependent probabilities $p(g | d)$.

All information-state comparisons used the same gravity operator, acquisition geometry, noise model, density bounds, diffusion schedule, posterior-sampling hyperparameters, sample budget, and fixed evaluation cases. Thus, the geological information supplied to the prior is the intended experimental difference among the three diffusion settings (Fig. 3).

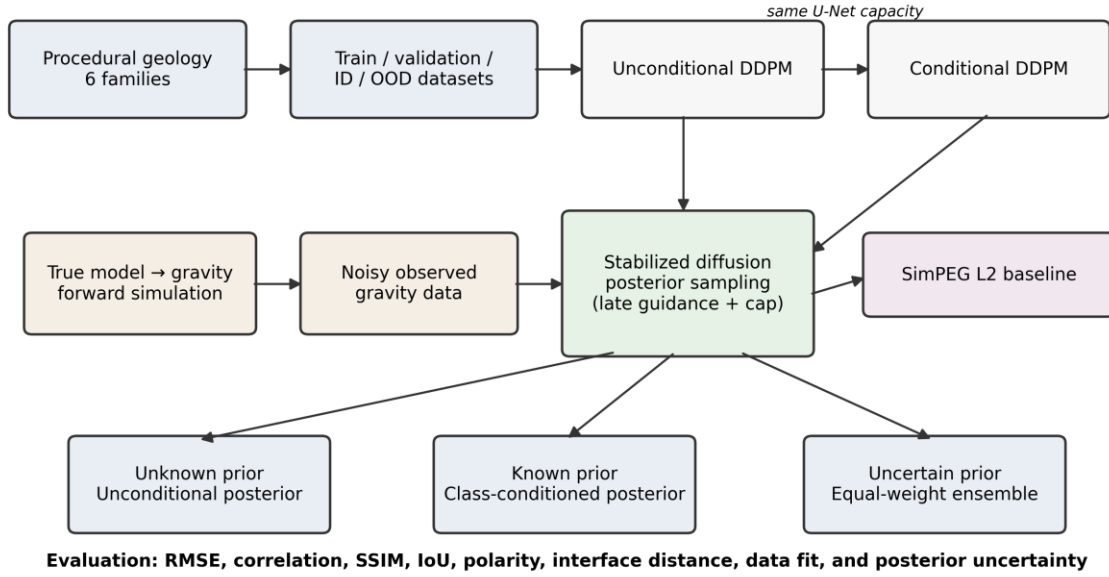


Fig. 3 Controlled workflow. The observations, forward operator, noise model, posterior-sampling configuration, test cases, and sample budget are fixed; only the geological information used by the prior changes

3.6 Gravity configuration and deterministic baseline

The gravity problem used a $64 \times 1 \times 64$ TensorMesh with 100 m cells horizontally and vertically and a 1,000 m strike width. Sixty-four receivers were spaced at 100 m and placed 1 m above the surface. Gaussian noise had standard deviation equal to 2% of the maximum absolute clean response for each case. The SimPEG L2 baseline used weighted least-squares regularization, sensitivity weighting, eigenvalue-based initial beta estimation, beta cooling, bound-constrained projected Gauss-Newton conjugate gradients, bounds $[-1, 1]$, a maximum of 20 nonlinear iterations, and target-misfit stopping [1]. The observations, forward operator, and physical bounds were identical to those used for the diffusion inversions.

3.7 Evaluation and statistical analysis

No single metric is treated as a complete measure of inversion quality. RMSE is the primary full-field reconstruction metric because it measures cellwise density error, although it can favor smooth averages of structurally incompatible solutions. Correlation measures spatial pattern agreement while being less sensitive to absolute amplitude. SSIM [11] compares local intensity, contrast, and spatial organization between the recovered and true density fields; higher SSIM indicates greater structural agreement. SSIM is evaluated over the common normalized density range $[-1, 1]$ using a fixed data range of 2.0.

For blobs, dykes, and intrusions, IoU measures the overlap of thresholded anomaly regions

$$\text{IoU} = \frac{|A_{\text{est}} \cap A_{\text{true}}|}{|A_{\text{est}} \cup A_{\text{true}}|} \quad (4)$$

where higher values indicate better spatial overlap. IoU and related object descriptors are used only for the compact-anomaly families ($n = 52$ ID; $n = 45$ OOD). For horizontal, dipping, and faulted layers, interface-distance metrics evaluate boundary recovery. These structure-aware measures complement RMSE but also depend on the chosen structural representation or threshold, so disagreements among metrics are interpreted as meaningful trade-offs rather than reduced to a single ranking.

Posterior behavior is summarized by pixelwise sample standard deviation, empirical pointwise 68% and 95% interval coverage, and the within-case Spearman association between ensemble spread and absolute reconstruction error. Coverage is treated as a finite-ensemble diagnostic under the synthetic experiment, not as evidence of calibrated Bayesian credible intervals.

Statistical comparisons use matched test cases. The six primary RMSE comparisons are the three diffusion-prior settings against SimPEG on the ID and OOD benchmarks. For paired comparisons, we report the mean difference with a 95% confidence interval, paired t-test, Wilcoxon signed-rank test, and Cohen's paired effect size d_z . Holm correction is applied to the six primary RMSE Wilcoxon tests. Structural and family-stratified analyses are treated as secondary or descriptive unless paired inference is explicitly reported.

4 Results

4.1 Overall reconstruction performance

Table 1 summarizes the principal benchmark metrics. All three learned-prior settings reduced mean RMSE relative to SimPEG (Fig. 4). In ID cases, mean RMSE decreased from 0.409 for SimPEG to 0.350 for the unknown prior, 0.343 for the known prior, and 0.333 for the scenario ensemble. In OOD cases, the corresponding values were 0.424, 0.380, 0.375, and 0.375.

Table 1 Principal benchmark metrics reported as mean (standard deviation). RMSE, correlation, and SSIM use all 100 cases per split; IoU uses compact-anomaly cases only ($n = 52$ ID; $n = 45$ OOD). Lower RMSE is better; higher values are better for the remaining metrics

Split	Method	RMSE	Correlation	SSIM	IoU
ID	SimPEG L2	0.409 (0.252)	0.143 (0.323)	0.149 (0.163)	0.086 (0.136)
ID	Unknown prior	0.350 (0.213)	0.388 (0.351)	0.226 (0.216)	0.285 (0.183)
ID	Known prior	0.343 (0.196)	0.398 (0.348)	0.281 (0.266)	0.300 (0.230)

Split	Method	RMSE	Correlation	SSIM	IoU
ID	Scenario ensemble	0.333 (0.192)	0.404 (0.304)	0.207 (0.195)	0.195 (0.239)
OOD	SimPEG L2	0.424 (0.219)	0.164 (0.272)	0.114 (0.160)	0.070 (0.117)
OOD	Unknown prior	0.380 (0.195)	0.349 (0.340)	0.184 (0.211)	0.305 (0.248)
OOD	Known prior	0.375 (0.189)	0.345 (0.368)	0.231 (0.268)	0.341 (0.275)
OOD	Scenario ensemble	0.375 (0.184)	0.334 (0.320)	0.148 (0.176)	0.219 (0.247)

Paired comparisons confirm that these reductions are not driven only by differences in aggregate means (Table 2). The paired mean RMSE reductions were 0.044-0.075, with moderate paired effect sizes ($|dz| = 0.38$ -0.50), and all six learned-prior-versus-SimPEG comparisons remained significant after Holm correction. These results support a learned-prior advantage over the stated smooth L2 baseline for this fixed synthetic benchmark; they do not imply universal superiority over deterministic inversion.

Table 2 Paired RMSE differences relative to the SimPEG baseline. Differences are learned-prior method minus SimPEG, so negative values indicate lower RMSE. CI denotes the 95% confidence interval for the paired mean difference, dz is Cohen's paired-sample effect size, and pHolm is the Holm-adjusted Wilcoxon signed-rank p-value across the six primary comparisons

Split	Method	Mean difference	95% CI	dz	pHolm
ID	Unknown prior	-0.0585	[-0.0877, -0.0293]	-0.40	3.23e-03
ID	Known prior	-0.0652	[-0.0936, -0.0368]	-0.46	1.41e-04
ID	Scenario ensemble	-0.0751	[-0.1049, -0.0452]	-0.50	3.15e-07
OOD	Unknown prior	-0.0444	[-0.0673, -0.0214]	-0.38	3.23e-03
OOD	Known prior	-0.0493	[-0.0730, -0.0256]	-0.41	1.82e-03
OOD	Scenario ensemble	-0.0496	[-0.0717, -0.0276]	-0.45	4.74e-05

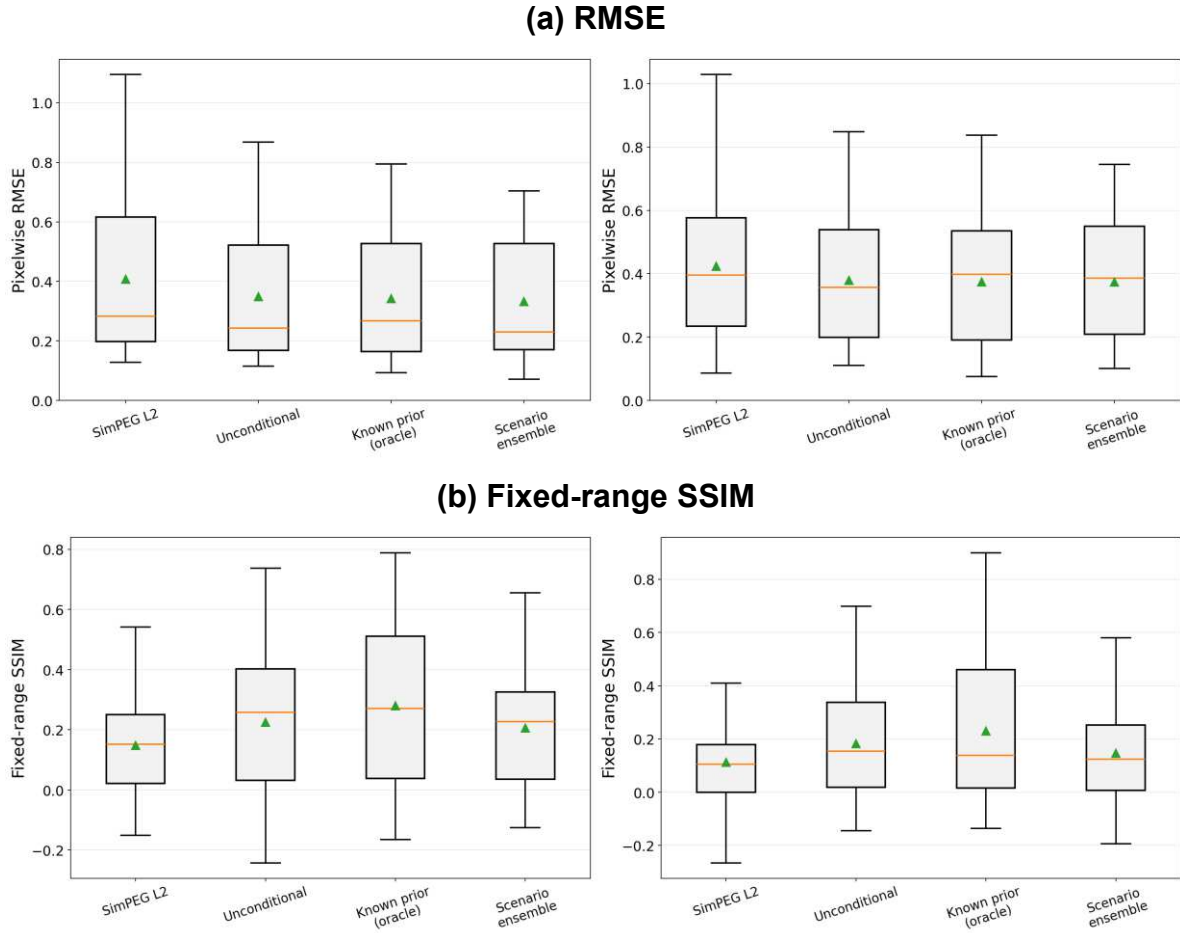


Fig. 4 Overall performance across the fixed ID and OOD benchmarks: (a) RMSE and (b) fixed-range SSIM. Learned priors reduce RMSE relative to SimPEG, while correct family conditioning gives the highest mean fixed-range SSIM

4.2 Known family information primarily improves structural recovery

Known-family conditioning lowered RMSE relative to the unknown prior in 53% of both ID and OOD cases, but the mean paired RMSE differences were not significant (ID $p = 0.40$; OOD $p = 0.45$, paired t-tests). Its clearer benefit was structural: mean SSIM increased from 0.226 to 0.281 in ID and from 0.184 to 0.231 in OOD. Relative to the unknown prior, paired SSIM effect sizes were $d_z = 0.39$ (ID) and 0.31 (OOD). For anomaly-family cases, known-prior IoU exceeded the scenario ensemble in both splits, whereas its IoU advantage over the unknown prior was not statistically established. The evidence therefore supports improved structural fidelity from correct family information, but not a general RMSE advantage over the unconditional prior.

4.3 Effects vary by geological morphology

The family breakdown reveals substantial heterogeneity (Fig. 5). Known information was helpful for compact and elongated targets in several comparisons, including ID dykes and OOD intrusions, whereas OOD horizontal layers were recovered better in RMSE by

the unknown prior. These are descriptive family-level results because the strata contain only 10-25 cases. They indicate that the value of family labels depends on morphology and extrapolation regime rather than acting as a uniform performance shift.

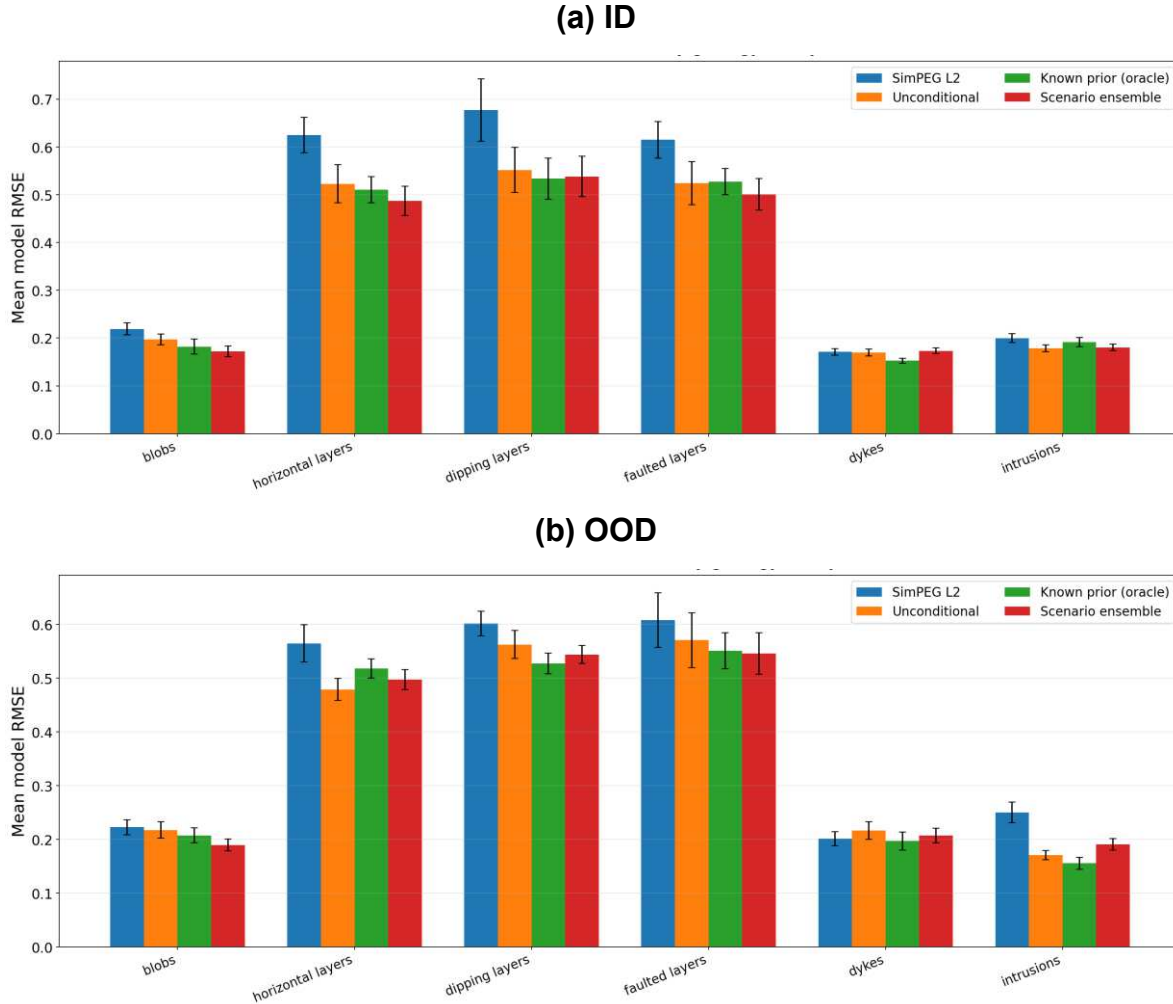


Fig. 5 Mean RMSE by geological family for (a) ID and (b) OOD cases. Family counts are unequal and modest, so these panels diagnose heterogeneity rather than provide confirmatory family-wise tests

4.4 Equal-weight scenario pooling trades sharpness for lower mean error

The equal-weight scenario ensemble achieved the lowest ID mean RMSE and highest ID correlation. Its RMSE improvement over the unknown prior was significant in ID (mean paired difference -0.0165; Wilcoxon $p = 0.0046$; paired t-test $p = 0.0099$) but not OOD. The ensemble did not lead the structural metrics: known-prior SSIM was 0.281 versus 0.207 in ID and 0.231 versus 0.148 in OOD; anomaly-family IoU was 0.300 versus 0.195 in ID and 0.341 versus 0.219 in OOD. Selected representative cases illustrate the trade-off: averaging across incompatible family-conditioned reconstructions can reduce cellwise squared error while attenuating contacts and compact bodies (Figs. 6 and 7).

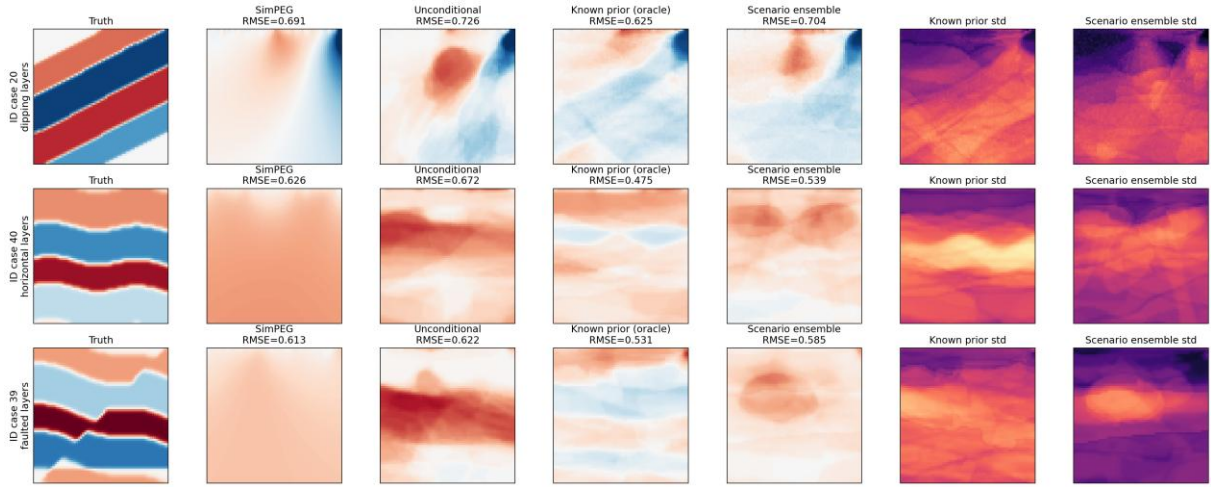


Fig. 6 Selected cases in which the known prior gives stronger structural recovery than the equal-weight scenario ensemble. The examples emphasize layered and fault-disrupted settings in which family-consistent contacts are retained by the known prior while the equal-weight ensemble becomes smoother

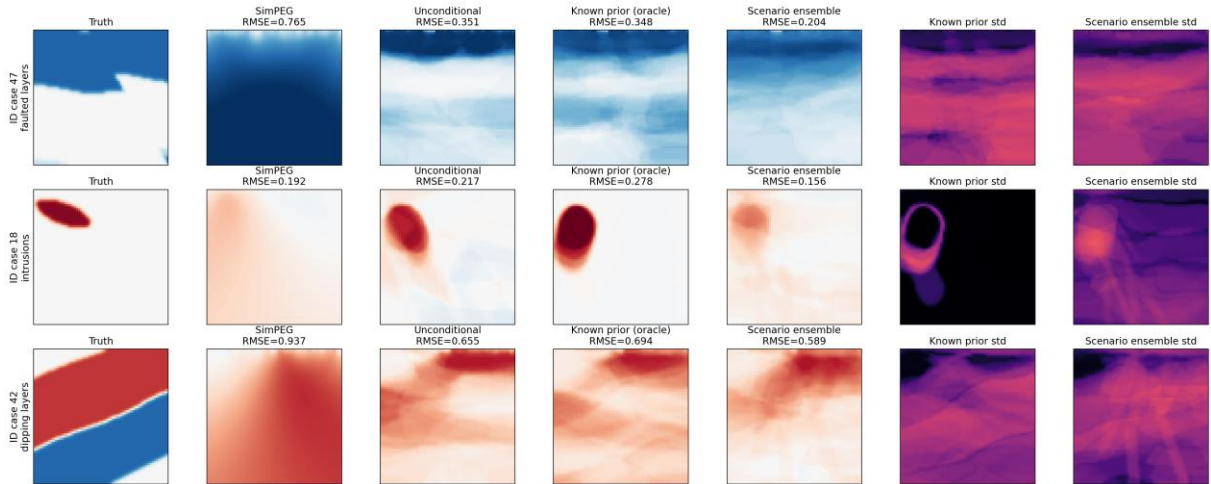


Fig. 7 Selected cases in which the equal-weight scenario ensemble has lower RMSE than the known prior. Lower cellwise error can coincide with a smoother and less sharply segmented aggregate estimate, particularly when averaging softens compact bodies or sharp interfaces

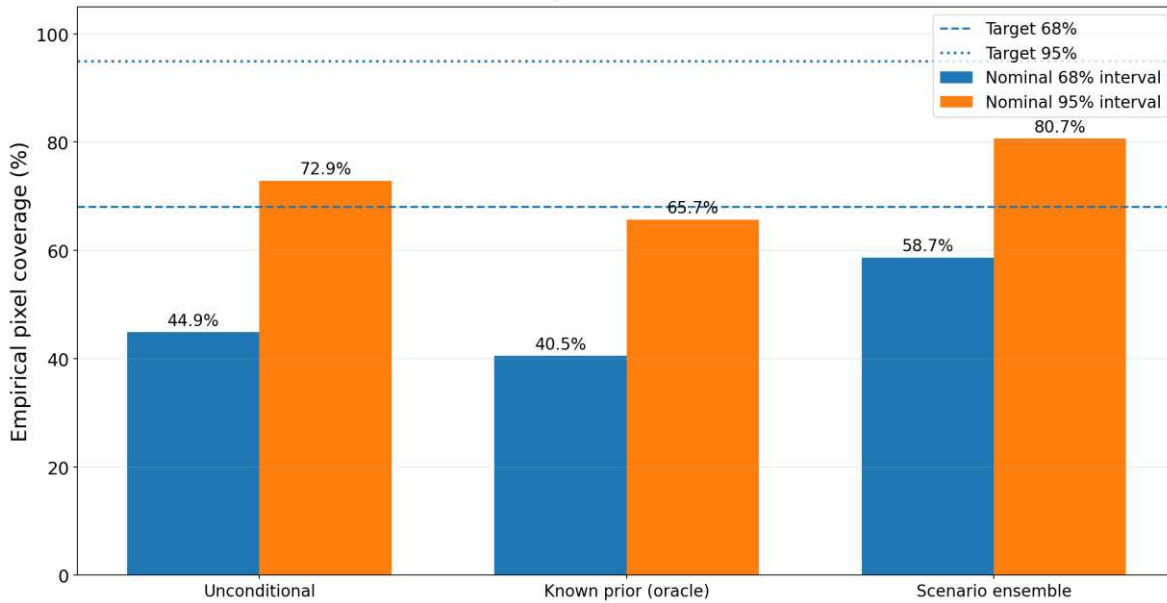
4.5 Ensemble spread is informative but under-dispersed

The scenario ensemble produced the broadest intervals and highest empirical coverage (Fig. 8). ID 68% coverage was 0.449 for the unknown prior, 0.405 for the known prior, and 0.587 for the scenario ensemble; the scenario ensemble's 95% coverage was 0.807. OOD 68% coverage decreased to 0.414, 0.346, and 0.473, respectively. All values are below their nominal levels, particularly OOD, so the intervals are under-dispersed and are not interpreted as calibrated credible intervals.

Despite this under-dispersion, ensemble spread contains useful relative information. The mean within-case Spearman association between pixelwise standard deviation and

absolute reconstruction error ranged from 0.49-0.58 across the three diffusion settings in ID and 0.43-0.50 in OOD. Higher-spread regions also tend to occur near contacts and uncertain interfaces. Spread is therefore interpreted as a relative indicator of reconstruction difficulty rather than a calibrated uncertainty magnitude.

(a) ID coverage



(b) OOD coverage

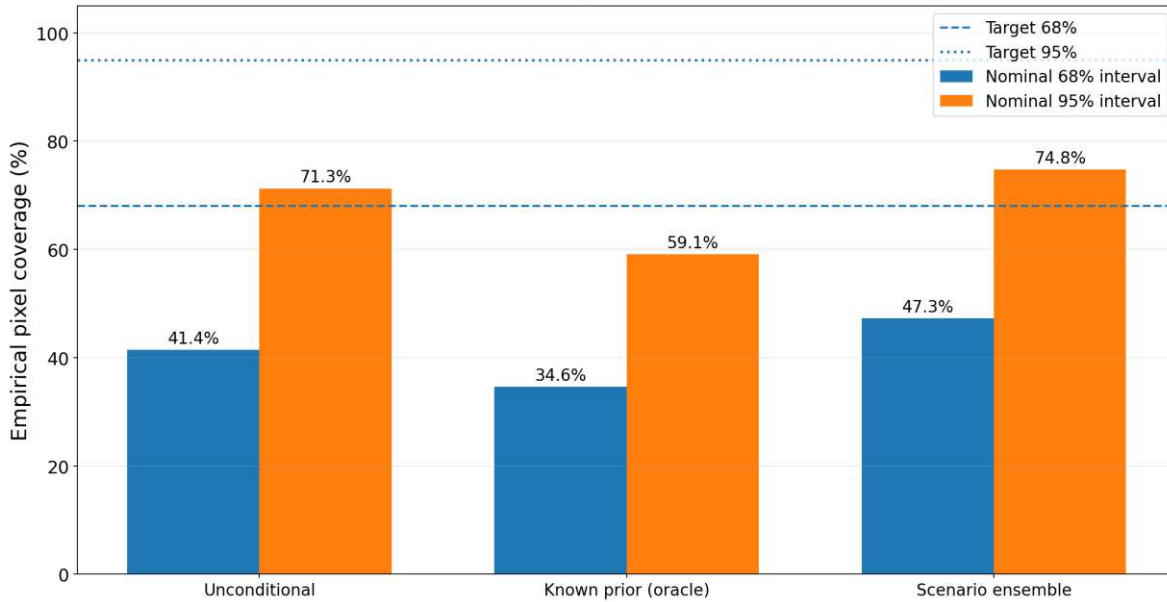


Fig. 8 Empirical 68% and 95% pixelwise interval coverage for the three diffusion-prior settings in (a) ID and (b) OOD cases. The equal-weight scenario ensemble has the broadest intervals and the highest empirical coverage, but all methods remain below nominal coverage, especially in OOD cases

5 Discussion

The controlled comparison shows that geological-prior information changes the character of the recovered solution rather than producing a single universally best method. The unconditional prior performs consistently across the mixed benchmark. Correct family conditioning restricts samples to a narrower structural manifold and improves structure-aware metrics, but it does not significantly lower average RMSE relative to the unconditional prior. Equal-weight scenario pooling can lower aggregate RMSE by averaging across alternatives, while the same averaging can blur structures that are incompatible across families.

This trade-off has a direct interpretive consequence. When independent evidence strongly supports one geological family, family-conditioned inference can preserve the expected morphology. When several interpretations remain plausible, the family-specific reconstructions are more informative than the pooled mean alone because the mean of a multimodal scenario ensemble need not resemble any geologically plausible realization. Equal weights provide a transparent sensitivity baseline, not optimal or data-updated scenario probabilities; evidence- or expert-informed weights are a natural extension for field applications.

The benchmark was deliberately designed around six interpretable morphology regimes rather than geological completeness. Compact irregular bodies, coherent intrusions, narrow dykes, continuous layers, inclined layers, and fault-disrupted interfaces stress different aspects of gravity resolution and prior misspecification. Because the benchmark realizations are synthetic, two-dimensional, and restricted to discrete single-family geology, the conclusions should not be extrapolated directly to mixed geology, unseen geological concepts, three-dimensional surveys, or field data. Likewise, the OOD split addresses parameter extrapolation within the six known families, not unseen-family generalization.

Metric disagreement identifies a real trade-off. RMSE measures full-field density error but can favor smooth averages, whereas SSIM, IoU, and interface distance emphasize different aspects of structural preservation. These structural measures also have limitations: SSIM depends on its evaluation scale and data range, IoU depends on anomaly thresholding, and interface distance is meaningful only for appropriate geometries. The interpretation therefore relies on complementary evidence rather than a single scalar score: learned priors improve RMSE relative to the stated smooth baseline, correct family information improves structural fidelity, and equal-weight pooling can trade structural sharpness for lower mean error.

The uncertainty results require similar restraint. Pixelwise spread is positively associated with reconstruction error and is useful for identifying comparatively difficult regions, but empirical interval coverage remains below nominal levels. The finite 24-sample ensembles, approximate DPS sampler, and four-samples-per-family allocation in

the scenario ensemble further limit precision. We therefore interpret spread and coverage as diagnostics of the generated ensemble, not as calibrated Bayesian uncertainty.

Applying the framework to field problems would require realistic three-dimensional geology, irregular acquisition, forward-model discrepancy, more representative noise, and geological scenario probabilities informed by independent evidence. These extensions affect practical deployment but do not alter the controlled conclusion of the present experiment: the geological information assumed before inversion materially changes what solution is recovered and how its quality should be judged.

6 Conclusion

Under a fixed gravity-inversion workflow, changing the geological information supplied to a diffusion prior changes reconstruction error, structural fidelity, and finite-ensemble spread. All learned-prior settings reduced RMSE relative to the stated SimPEG L2 baseline. Correct family conditioning most consistently improved structural metrics, whereas equal-weight scenario pooling achieved the lowest ID mean RMSE while producing smoother aggregate estimates. The scenario ensemble should therefore be interpreted as a transparent sensitivity analysis rather than Bayesian model averaging. Geological-prior specification should be reported explicitly and evaluated with complementary pixelwise, structural, and ensemble diagnostics. Future work should extend this controlled framework to mixed and three-dimensional geology and use independent evidence to define scenario probabilities in field applications.

Declarations

Funding

No funding was received for this work.

Ethics approval and consent to participate

Not applicable. This study used synthetic data and did not involve human participants, human data, animals, or biological materials.

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Author contributions

Omid Mahmoodi is the sole author of this manuscript and approved the submitted version.

Data and code availability

The code for procedural generation, training, inversion, and analysis is available at <https://github.com/OMahmoodi/diffusion-Based-Probabilistic-Gravity-Inversion>. The

repository documents the family-specific training, validation, ID, and OOD generation definitions used in this study.

References

- [1] Cockett R, Kang S, Heagy LJ, Pidlisecky A, Oldenburg DW. SimPEG: an open source framework for simulation and gradient based parameter estimation in geophysical applications. *Computers & Geosciences*. 2015;85:142-154. <https://doi.org/10.1016/j.cageo.2015.09.015>
- [2] Lochbühler T, Vrugt JA, Sadegh M, Linde N. Summary statistics from training images as prior information in probabilistic inversion. *Geophysical Journal International*. 2015;201(1):157-171. <https://doi.org/10.1093/gji/ggv008>
- [3] Lopez-Alvis J, Nguyen F, Looms MC, Hermans T. Geophysical inversion using a variational autoencoder to model an assembled spatial prior uncertainty. *Journal of Geophysical Research: Solid Earth*. 2022;127(3):e2021JB022581. <https://doi.org/10.1029/2021JB022581>
- [4] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: *Advances in Neural Information Processing Systems*. 2020;33:6840-6851.
- [5] Song Y, Sohl-Dickstein J, Kingma DP, Kumar A, Ermon S, Poole B. Score-based generative modeling through stochastic differential equations. In: *International Conference on Learning Representations*. 2021.
- [6] Kavar B, Elad M, Ermon S, Song J. Denoising diffusion restoration models. In: *Advances in Neural Information Processing Systems*. 2022;35:23593-23606. <https://doi.org/10.52202/068431-1714>
- [7] Chung H, Kim J, McCann MT, Klasky ML, Ye JC. Diffusion posterior sampling for general noisy inverse problems. In: *International Conference on Learning Representations*. 2023.
- [8] Wang F, Alkhalifah T. Geological and well prior assisted full waveform inversion using conditional diffusion models. In: *85th EAGE Annual Conference & Exhibition*. 2024;4:2251-2255. <https://doi.org/10.3997/2214-4609.202410821>
- [9] Taufik MH, Alkhalifah T. Accelerating Bayesian full waveform inversion using reconstruction-guided diffusion sampling. *Geophysical Journal International*. 2026;245(2):ggag066. <https://doi.org/10.1093/gji/ggag066>
- [10] Xian M, Xu Z, Zhang Y, Zhdanov MS, Wang X. ResDM-Net++: an enhanced diffusion model for gravity inversion. *IEEE Transactions on Geoscience and Remote Sensing*. 2025;63:5915213. <https://doi.org/10.1109/TGRS.2025.3572823>
- [11] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004;13(4):600-612. <https://doi.org/10.1109/TIP.2003.819861>