Response to Reviewer's Comments on Manuscript ID: egusphere-2026-1087

Dear Dr. Poulet and Reviewers,

Thank you for your letter and for regarding our manuscript entitled "Latent-Compression-Free Generative Diffusion with Geological Priors and Geophysical Regularization for Implicit Structural Modeling" (ID: egusphere-2026-1087). We sincerely appreciate the time and effort you have dedicated to reviewing our work.

We thank both reviewers for their careful and constructive engagement with our work over the course of the review process. We are grateful to Reviewer 2 for the positive assessment of our previous revision and for these two final points, both of which were precise and have led to genuine improvements in the manuscript. We have addressed them in full, as detailed below. All changes are marked in the revised manuscript.

Response to Reviewer 2:

Comment 1: Noise scale: definition and internal inconsistency. Section 3.2 (l. 247) states that the noise scale is set to 0.2 when constructing noisy samples during training, whereas the caption of Fig. 6 states that the model is trained with noise scale NS = 2.0 and that only the inference-time noise scale is varied. These two statements are inconsistent. The discussion in Sec. 4 suggests that 0.2 is the correct training value and that the figure caption has not been updated. More importantly, the "noise scale" is never defined in Sec. 2.1. Equation (1) is written with a standard Gaussian, and no scale parameter appears anywhere in the methods. Since the paper identifies this quantity as a critical factor for implicit structural modeling, please introduce it explicitly in Sec. 2.1, state precisely how it enters the construction of z_t, and make the training and inference values consistent throughout.

Response: We thank the reviewer for catching this. We realize that our original text did not make an important point clear: the two statements in fact refer to two different models, and both values are correct. The caption of Fig. 6 was therefore not simply out of date.

The actual sequence of our experiments was as follows. We first trained an exploratory model using a noise scale of 2.0, a value commonly adopted in natural-image diffusion, and found that it produced overly noisy generations. Before committing to a full retraining, we used that model to scan the **inference-time** noise scale while keeping all other settings fixed; this is the experiment shown in Fig. 6, and it revealed that noise scales of 0.2 and 0.1 gave markedly smoother and more geologically reasonable fields. On the basis of this observation we retrained the model with a noise scale of 0.2, and all other results reported in the paper use that final model.

We fully agree that presenting Fig. 6 under the heading "Impact of noise scale" without stating this made it read as an ablation of the final model, which is a legitimate source of confusion. We have therefore revised the caption of Fig. 6 and the corresponding paragraph in Sec. 4 to state explicitly that the figure was obtained with an early exploratory model trained with NS = 2.0, that it characterizes a deliberate train–inference mismatch rather than an ablation of the training noise scale, and that it is reported because it motivated our final choice of NS = 0.2.

Second, we agree that the noise scale was never defined. We have now introduced it explicitly in Sec. 2.1. The base variable is obtained by scaling a standard Gaussian, $\varepsilon = \sigma \cdot \varepsilon_0$ with $\varepsilon_0 \sim N(0, I)$, where the scalar $\sigma > 0$ is the noise scale (NS) and scales the amplitude of the Gaussian noise. Substituting this into Eq. (1) gives $z_t = t \cdot x + (1-t) \cdot \sigma \cdot \varepsilon_0$, which we now state explicitly, together with the fact that the ODE is initialized from $z_0 = \sigma \cdot \varepsilon_0$ at inference and that the same value of σ is used at training and inference time. The default value $\sigma = 0.2$ is stated in Sec. 2.1, and the training and inference values are now consistent throughout the manuscript.

Comment 2: Interpretation of Fig. 7. The new experiment is a useful addition, but the conclusion drawn from it is stronger than the evidence supports. Because z_t = tx + (1-t)eps, the input at t = 0.1 still contains a fraction of the ground-truth field, so Fig. 7 characterises single-step denoising given partial ground truth rather than the regime the network encounters early in generation. This is reflected in the numbers: the reported MSE of 9.77e-04 at t = 0.1 is roughly an order of magnitude lower than the full 50-step generation MSE of 2.93e-02 in Table 4, indicating that the diagnostic is not measuring the same quantity as the generation task. I would ask the authors either to extend the analysis to smaller timesteps (e.g., t = 0.01 and t = 0.05), where the justification for the prior losses is weakest, or to soften the claim that applying these losses at high noise levels "remains reasonably well-justified" so that it acknowledges the untested low-t limit.

Response: We fully agree with the reviewer's analysis. Because z_t retains a fraction t of the ground-truth field, Fig. 7 characterizes single-step denoising given partial ground truth, which is not the same quantity as the end-to-end generation error; the gap between 9.77×10^{-4} and the 2.93×10^{-2} of Table 4 correctly reflects this, and our original wording did not acknowledge the distinction. We have followed both of the reviewer's suggestions.

First, we have extended the analysis to the smaller timesteps suggested by the reviewer. Figure 7 now includes panels at t = 0.01 and t = 0.05, with prediction errors of 1.89×10^{-2} and 1.39×10^{-3} respectively. At t = 0.01 the single-step error is thus comparable to the full 50-step generation MSE of 2.93×10^{-2}, confirming the reviewer's point that the previously reported range was substantially easier than the regime the network encounters early in generation. At the same time, the corresponding panels show that even at t = 0.01, where z_t is almost pure noise and the pixel-wise error is largest, the network still recovers a coherent, fault-aligned stratigraphic architecture rather than a severely blurred or incoherent field. This is the property that bears on whether the prior losses act on a meaningful target, and we now frame our argument in these terms rather than in terms of prediction accuracy.

Second, we have softened the claim accordingly. The revised text no longer states that applying the prior losses at high noise levels "remains reasonably well-justified". It instead states that the prior losses act on a structurally meaningful target across the whole range of t, that their justification is nevertheless strongest at low noise levels, and that applying them with a uniform weight across all t should be regarded as an empirically motivated choice rather than one that is equally well-founded at every noise level. We have also added a clarification in Sec. 4 that this diagnostic measures single-step denoising given a partially informative z_t and is therefore not directly comparable to the end-to-end generation error reported in Table 4. This strengthens the case for the t-dependent weighting schedule that we already identify as a direction for future work.

We thank the reviewer again for these two comments, which have improved the precision of the manuscript, and we thank the editor for handling the revision personally.

Sincerely,

Zhixiang Guo, on behalf of all authors