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

Dear 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.

Response to Reviewer1:

1. Geophysical vs. Geological Data: The manuscript frequently refers to "geophysical data," yet the model constraints and inputs appear to consist exclusively of geological interpretations (e.g., horizons and faults). While these are presumably derived from seismic data, they are geological interpretations rather than geophysical data. Please either clarify what is meant by "geophysical" or modify the text to remove references to "geophysical data".

Response:
We sincerely thank you for this valuable comment and for pointing out the ambiguity in our use of the terms "geophysical data" and "geological interpretations". We agree that the conditioning inputs used in this study consist of interpreted geological structures (i.e., horizons and faults) rather than direct geophysical measurements. Following your suggestion, we carefully revised the manuscript to clearly distinguish geological interpretations from geophysical data. In particular, we replaced ambiguous references to "geophysical data" with more precise terminology, such as "implicit structural models", where appropriate. We believe these revisions improve the clarity and precision of the manuscript.

Line3: However, existing diffusion transformer pipelines scale poorly to high-dimensional structural modeling data because noise- or velocity-prediction objectives are often unstable at large patch sizes

Line18: LFD offers new insights into deploying diffusion models for high-dimensional implicit structural modeling problems

Line73: However, deploying flow matching for high-dimensional implicit structural modeling faces key challenges.

Line125: which greatly increases token length and attention cost for high-dimensional data

Line127: but this is less appealing for implicit structural models because it may compromise fine structural details and makes it harder to explicitly impose geological priors in the latent space.

Line303: This finding provides practical guidance for deploying diffusion models in geophysics: the noise scale should be tuned to the target task and data characteristics to achieve optimal generation quality.

Line329: Overall, LFD demonstrates that diffusion models can effectively generate implicit structural models while directly honoring geological constraints in data space.

2) Ablation study and uncertainty: Please extend the results or discussion section to include an ablation study showing how the model's predictions change as the horizon constraints become increasingly sparse. A multi-realisation comparison (using metrics such as information entropy) to demonstrate how well the generated realisations cover the plausible, data-consistent model space is also needed. Ideally this uncertainty analysis should be explicitly linked to the sparse horizon ablation study, to illustrate how realisation variance increases as the amount of conditioning data is reduced.

Response:
Thank you for your valuable suggestion. You correctly point out that an ablation study on horizon sparsity and a multi-realization uncertainty analysis is needed to better characterize the model's behavior under sparse conditioning. To address this, we have added a new subsection (Sec. 3.5) in the Experiments section. Specifically, we generate 20 independent realizations by sampling different initial Gaussian noise fields while keeping the fault and horizon conditioning fixed, under four horizon configurations (6, 4, 2, and 1 input horizons). We compute the pixel-wise variance across realizations to quantify uncertainty. The results show that the mean variance increases monotonically as the number of input horizons decreases, and that high-variance regions expand spatially in areas lacking horizon constraints. These findings confirm that realization spread is directly controlled by the density of horizon constraints, and that the model appropriately broadens its output distribution to reflect increased geological ambiguity when conditioning data are sparse. The corresponding figure has been added as Figure 4 in the revised manuscript.

3.5 Ablation on Horizon Sparsity and Uncertainty Analysis
To investigate how the number of input horizon constraints affects generation quality and realization variability, we conduct an ablation study in which the number of input horizons is systematically reduced from 6 to 4, 2, and 1 (The top row in Figure 4). For each configuration, we generate 20 independent realizations by sampling different initial Gaussian noise fields while keeping the fault and horizon conditioning fixed, and compute the pixel-wise variance across realizations to quantify uncertainty.

The middle row of Figure 4 shows the mean implicit structural models over 20 realizations generated under each sparsity level. As the number of input horizons decreases, the generated models remain geologically plausible but exhibit increasing variability in stratigraphic geometry, particularly in regions far from the remaining horizon constraints. This reflects the model's learned prior filling in the unconstrained space with a broader range of geologically consistent solutions.

The bottom row of Figure 4 shows the spatial distribution of pixel-wise variance across 20 realizations for each horizon configuration. With 6 input horizons, variance is uniformly low across the

model domain, indicating that dense conditioning effectively anchors the generation. As the number of horizons is reduced to 4, 2, and finally 1, high-variance regions progressively expand, concentrating first in areas between horizons and eventually spanning the majority of the model domain. This spatial pattern is consistent with the intuition that uncertainty grows where conditioning data are absent. The mean variance plot (bottom-left panel of Figure 4) further confirms this trend quantitatively: mean variance increases monotonically from approximately 0.0013 (6 horizons) to 0.0073 (1 horizon), demonstrating that realization spread is directly controlled by the density of horizon constraints. Collectively, these results suggest that LFD generates a data-consistent distribution: when conditioning data are abundant, the model converges to consistent solutions; when data are sparse, the model appropriately broadens its output distribution to reflect the increased geological ambiguity, rather than collapsing to a single overconfident prediction.

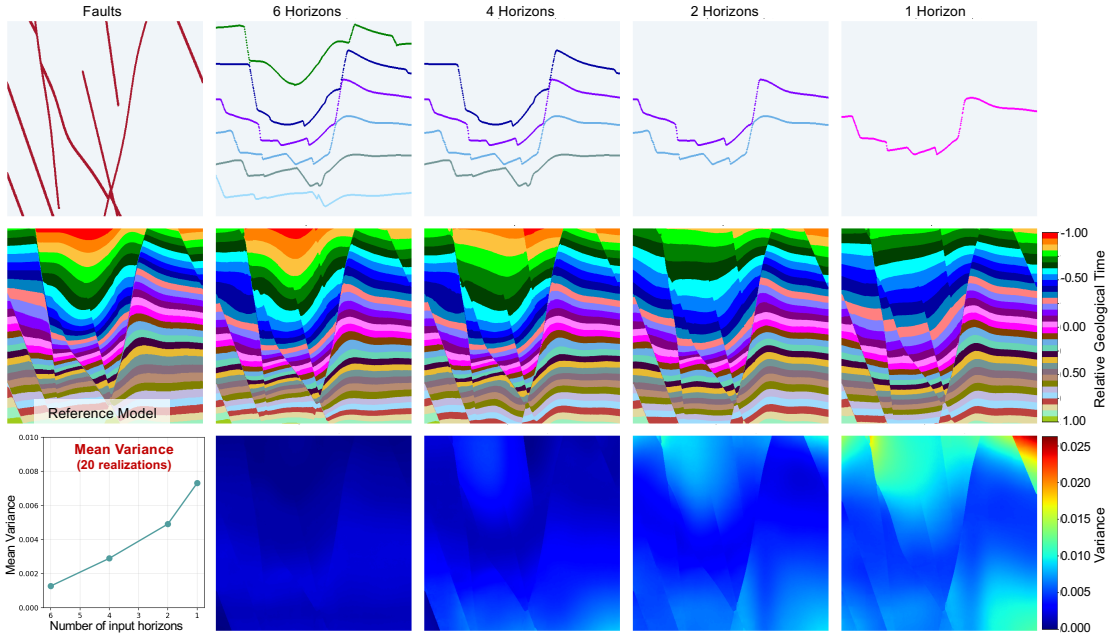

Figure4: Ablation study on horizon sparsity and realization uncertainty. Top row: input conditioning, showing the shared fault interpretation (leftmost) and horizon inputs with 6, 4, 2, and 1 horizons. Middle row: mean implicit structural model over 20 realizations for each horizon configuration, with the reference model shown in the leftmost panel. Bottom row: pixel-wise variance across 20 realizations (leftmost panel: mean variance as a function of the number of input horizons), demonstrating that realization variance increases monotonically as horizon constraints become sparser.

3) Expanded Discussion: The current discussion is quite short. I suggest expanding it significantly, including to address the models generalisation capacity. How close (geologically speaking) do applications need to be to the synthetic training data for the model to perform reliably? How much variability can the model produce, especially when data is quite sparse? Does it / can it produce geologically impossible results (e.g., "bubbles")?

Response:
Thank you for your valuable suggestion. We agree that the original Discussion section was too brief. Following your recommendation, we have substantially expanded the Discussion to address (1) the

model's generalization capacity, (2) Variability under sparse horizon conditioning., and (3) the possibility of geologically implausible results.

Specifically, we discuss how the model's performance depends on the range of structural styles represented in the synthetic training data and highlight the limitations observed for thrust-related structures that are absent from the current training set. We also incorporate the newly added horizon-sparsity ablation study (Sect. 3.5), which demonstrates increasing realization variability as conditioning horizons are progressively reduced. Finally, we discuss the potential occurrence of geologically implausible structures under extremely sparse conditioning and the role of the proposed fault-aware regularization in mitigating such artifacts, while outlining possible future improvements through additional geological constraints.

Discussion

Generalization capacity. The real-survey experiments (Figure 3) provide some indication of the model's generalization capacity. Since the synthetic training data are generated using parameterized fold, fault, and unconformity rules, the model is expected to perform reasonably well when the target geological structures fall within the range of styles represented in the training distribution. However, as demonstrated in panel (e) of Figure 3, performance degrades noticeably in the presence of pronounced thrust structures, which are absent from the training dataset. This suggests that the current model should be applied with caution in tectonic settings that deviate significantly from the training distribution, such as thrust-and-fold belts. Expanding the synthetic dataset to include such structural styles would be a natural direction for future work.

Variability under sparse horizon conditioning. The horizon-sparsity ablation study (Sec.3.5) shows that the variability of the generated realizations increases as the number of horizon constraints decreases, with mean variance rising from 0.0013 (6 horizons) to 0.0073 (1 horizon). This trend is also evident from the variance maps in Figure\ref{ablation}, where high-variance regions progressively expand as conditioning horizons are removed. Such behavior is expected, since regions that are weakly constrained by horizon information inherently admit a broader range of geologically plausible solutions. As the amount of conditioning information decreases, the model is afforded greater freedom in generating the implicit structural field, resulting in increased realization variability. These observations suggest that the model responds to sparse conditioning in a physically reasonable manner, producing more diverse structural realizations where geological constraints are limited.

Geologically impossible results. The fault-aware bending-energy loss encourages the generated implicit field to remain smooth within fault-bounded regions, which in practice is intended to suppress localized artifacts such as closed iso-surfaces (e.g., "bubbles") that would be geologically implausible. In our experiments, we did not observe such artifacts in the generated results, suggesting that the bending-energy regularization is effective in practice. Nevertheless, under very sparse conditioning (e.g., 1 input horizon), the high-variance regions visible in Figure 4 indicate that the model has considerable freedom in unconstrained areas, and the occurrence of geologically unreasonable structures cannot be entirely ruled out in more challenging scenarios. Incorporating additional

geological rules, such as stratigraphic monotonicity constraints, into the loss function could further reduce the likelihood of such artifacts and would be a worthwhile direction for future work.

4) Reproducibility: I attempted to run the code provided in the Zenodo repository but was unable to do so. To ensure the presented method is FAIR and usable, please update it to include a comprehensive README.md file explaining the code structure and how to get started. Specifically, the documentation should explain how to setup the required data, dependencies, and paths. Currently, it is unclear where to download the required ImageNet model, how to set the IMAGENET_PATH, and where to download the pretraining checkpoint (PRETRAIN_CKPT). Where to download the training and testing datasets (and so set the associated paths) is also unclear.

Response:
Thank you for this valuable comment and for attempting to reproduce our results. We agree that the original repository did not provide sufficient documentation to enable straightforward reproduction of the workflow. Following your suggestion, we have substantially revised the repository and expanded the README documentation. The updated documentation now includes detailed instructions for environment setup, dependency installation, dataset preparation, path configuration, and inference. We have also provided links to the required pretrained models and added explicit instructions for configuring PRETRAIN_CKPT. In addition, an example testing dataset has been included to facilitate reproduction of the inference workflow and the results. The Zenodo repository (https://doi.org/10.5281/zenodo.20508635) and the github repository (https://github.com/ProgrammerZXG/LFD) has been updated accordingly.

5) Figure Placement: Please check that all figures appear after their first mention in the text. Currently, several figures seem to appear very early in the manuscript and are not discussed until much later. Additionally, please clarify in Fig. 1 that the implicit field is predicted as a continuous variable. It would be useful to plot this continuous field alongside the discrete lithology/color visualisations (in at least one figure), as this gives a clearer representation of the models output.

Response:
Thank you for this valuable comment. Following your suggestion, we carefully reviewed the placement of all figures in the revised manuscript and adjusted their positions where necessary to ensure that figures appear after their first mention in the text.

We also clarified that the proposed model predicts a continuous implicit scalar field representing relative geological time rather than discrete classes. The discrete colormap visualizations used throughout the manuscript are intended solely for visualization purposes, as they make individual stratigraphic layers easier to distinguish. To make this distinction explicit, we revised the caption of Figure 1, added a clarification in the main text, and updated Figure 1 by explicitly labeling the model output as a continuous implicit field. These revisions provide a clearer representation of the model output and reduce the possibility of interpreting the displayed color bands as discrete classes.

Caption of Figure 1: Note that the implicit scalar field is a continuous-valued output, the colormap visualizations shown here use discrete color bins solely for display purposes.

The last sentence of the third paragraph in Section 3.4: Throughout this paper, we visualize the continuous implicit scalar field using a discrete colormap to better reveal the stratigraphic layering.

Minor Points

Abstract (Line 1): The phrase "diffusion models provide a promising way to model the distribution of implicit structural models" is awkward (do the models really model the distribution of implicit structural models? Which implicit structural models?). Consider rewording for clarity.

Response:
Thank you for your suggestion. We agree that the original phrasing was awkward. We have revised Line 1 of the abstract to: 'Diffusion models provide a promising way to generate geologically consistent implicit structural models by learning data-driven priors from training examples.

Abstract (Line 5): Change "Variational" to lowercase.

Response:
Thank you for this suggestion. We have corrected 'Variational' to lowercase accordingly.

Abstract (Line 11): Change "Transformer" to lowercase.

Response:
Thank you for this suggestion. We have corrected 'Transformer' to lowercase accordingly.

Line 116: Please explicitly clarify if the model predicts continuous implicit field values or discrete lithology classes. I assume the former is true, but the text and figure captions should be updated to make this unambiguous.

Response:
Thank you for this careful observation. We clarify that LFD predicts a continuous implicit scalar field representing relative geological time, not discrete lithology classes. The discrete colormap visualizations are used intentionally throughout the paper, as a continuous colormap makes it difficult to distinguish individual stratigraphic layers, whereas discrete color bins allow the layering structure to be seen more clearly. We have added explicit clarifications at two places: (1) at the end of Section 3.1, where we note that the continuous implicit scalar field is visualized using a discrete colormap to better reveal the stratigraphic layering, and (2) in the caption of Figure 1, where we explicitly state that the colormap visualizations use discrete color bins solely for display purposes.

Caption of Figure 1: Note that the implicit scalar field is a continuous-valued output, the colormap visualizations shown here use discrete color bins solely for display purposes.

The last sentence of the third paragraph in Section 3.4: Throughout this paper, we visualize the continuous implicit scalar field using a discrete colormap to better reveal the stratigraphic layering.

Line 125: Please define what is meant by "geophysical data" in the context of this study, or remove the term (as the manuscript does not currently feature integration with geophysical datasets).

Response:
Thank you for this careful observation. We agree that the term “geophysical data” was not appropriate in the context of the present study, as the model is conditioned by interpreted geological structures (horizons and faults) rather than direct geophysical measurements. Following your suggestion, we removed or revised the relevant occurrences of “geophysical data” throughout the manuscript and replaced them with more precise terminology, such as “implicit structural models”, “structural modeling”, or “geological constraints”, depending on the context. These revisions improve the clarity and accuracy of the manuscript.

Line3: However, existing diffusion transformer pipelines scale poorly to high-dimensional structural modeling data because noise- or velocity-prediction objectives are often unstable at large patch sizes

Line18: LFD offers new insights into deploying diffusion models for high-dimensional implicit structural modeling problems

Line73: However, deploying flow matching for high-dimensional implicit structural modeling faces key challenges.

Line125: which greatly increases token length and attention cost for high-dimensional data

Line127: but this is less appealing for implicit structural models because it may compromise fine structural details and makes it harder to explicitly impose geological priors in the latent space.

Line303: This finding provides practical guidance for deploying diffusion models in geophysics: the noise scale should be tuned to the target task and data characteristics to achieve optimal generation quality.

Line329: Overall, LFD demonstrates that diffusion models can effectively generate implicit structural models while directly honoring geological constraints in data space.

Figure 3: Please discuss what happens to the framework's performance when the fault mask is also sparse?

Response:
Thank you for this valuable comment. We agree that the completeness of the fault mask can influence the generated structural models. In the proposed framework, fault masks serve as conditioning information and explicitly indicate locations where structural discontinuities are permitted through the fault-aware regularization. Consequently, discontinuities in the generated implicit field tend to coincide with the supplied fault constraints.

If the fault mask becomes sparse, the model receives less information regarding the location and geometry of fault-related discontinuities. In such cases, regions without fault constraints are no longer explicitly identified as discontinuous, and the model tends to preserve the smoothness of the implicit structural field there. As a result, missing fault constraints may lead to weakened or absent discontinuities in those regions.

A related trend is observed in the horizon-sparsity ablation study (Figure 4), where reducing conditioning information leads to increased realization variability and structural ambiguity. Since fault masks are introduced through the same conditional mechanism, sparse fault constraints would likewise reduce the structural information available to the model and increase ambiguity near fault zones. This behavior is consistent with the conditional nature of the framework, which is designed to honor the available structural constraints. More generally, an important goal in implicit structural modeling is to identify previously unknown discontinuities when fault information is sparse or incomplete. The present study does not explicitly investigate this capability, and extending the framework to infer missing discontinuities represents an interesting direction for future work.

We have added a corresponding discussion in the revised manuscript.

Discussion

Fault sparsity. The current framework relies on fault masks to indicate locations where structural discontinuities are permitted, and discontinuities in the generated implicit field therefore tend to coincide with the supplied fault constraints. If a fault is omitted from the conditioning data, the corresponding region is no longer explicitly identified as a discontinuity, and the model tends to preserve the smoothness of the implicit structural field there, potentially leading to weakened or absent discontinuities. A similar trend is observed in the horizon-sparsity ablation study (Figure 4), where reducing conditioning information results in increased realization variability and structural ambiguity. Since fault masks are introduced through the same conditional mechanism, sparse or incomplete fault constraints would likewise reduce the structural information available to the model, increasing ambiguity near fault zones while favoring smoother structures in unconstrained regions. This behavior is consistent with the conditional nature of the framework, which is designed to honor the available structural constraints. In many implicit structural modeling workflows, it is desirable for the model to infer previously unidentified discontinuities even when fault information is sparse. However, this capability is not explicitly investigated in the present study. Extending the framework to identify and incorporate such missing discontinuities represents an interesting direction for future work.

Figure 3: There appears to be an absence of unconformities in the demonstration data (and potentially the training set?). Please address this limitation or design choice in the discussion. Was the model able to accurately reproduce unconformitites? How were these parameterised in the implicit framework (as the implicit value above and below an unconformity is not comparable).

Response:

Thank you for this insightful comment. We agree that unconformities are important geological features and that their treatment was not sufficiently discussed in the original manuscript.

Unconformity-related geometries are present in the synthetic structural models used for training. However, unlike horizons and faults, unconformities are not explicitly represented as conditioning information in the current framework. In the present study, our primary objective was to investigate whether the proposed diffusion-based workflow could effectively generate implicit structural models under geological constraints. To keep the conditioning scheme simple and isolate the effects of the two most commonly interpreted structural elements, we considered only horizons and faults as conditioning inputs. Consequently, the present study does not provide a dedicated evaluation of the model's ability to reconstruct unconformity surfaces, which we acknowledge as a limitation of the current work.

From the perspective of implicit structural modeling, unconformities and faults share an important similarity in that both correspond to discontinuities in the implicit scalar field. However, their geological roles are different. Faults primarily represent displacement and truncation of stratigraphic layers, whereas unconformities represent stratigraphic contact relationships between different depositional packages. In the current framework, fault masks are used both as conditioning information and as guidance for the fault-aware regularization. A similar strategy could be adopted for unconformities in future work.

Specifically, unconformity surfaces could be introduced as an additional conditioning channel, analogous to the fault-mask input. Since the implicit scalar field exhibits discontinuous behavior across both faults and unconformities, unconformity masks could also be incorporated into the regularization framework to explicitly model stratigraphic contact relationships. Such a representation would allow the diffusion model to learn unconformity-related discontinuities and contact geometries directly from conditioning information.

We have added a discussion of unconformities, their current limitations in the framework, and possible extensions in the revised manuscript.

Discussion

Unconformities.

Unconformity-related geometries are present in the synthetic training models; however, unconformities are not explicitly represented as conditioning constraints in the current framework. In this study, our primary objective was to evaluate the effectiveness of the proposed diffusion-based workflow under a simplified conditioning scheme using only horizons and faults. Consequently, the ability of the model to reconstruct unconformity surfaces has not been systematically evaluated, which we acknowledge as a limitation of the current work. Similar to faults, unconformities correspond to discontinuities in the implicit scalar field, although their geological meaning differs: faults represent stratigraphic displacement, whereas unconformities represent contact relationships between distinct stratigraphic packages. This suggests that unconformities could be incorporated as an additional conditioning channel, analogous to fault masks, together with an unconformity-aware regularization to explicitly model discontinuous stratigraphic contacts. Extending the framework in this direction represents an important avenue for future work.

Line 210: When creating the structural data, was the timing of the structural events allowed to vary (e.g., scenarios involving faults that are truncated by an unconformity)? Please expand a little on the synthetic modeling logic.

Response:
Thank you for this insightful comment. In the synthetic data generation workflow, structural evolution is represented through a sequential geology-informed modeling process. Specifically, folding deformation is generated first, followed by fault construction and slip simulation. Unconformities and associated stratal termination patterns (e.g., onlap, downlap, and toplap) are then introduced into the folded–faulted framework. Therefore, unconformity-related geometries are explicitly represented in the synthetic dataset.

The timing of structural events is not treated as an independent random variable. Instead, geological diversity is achieved through extensive randomization of fold geometries, fault networks, slip distributions, unconformity surfaces, and stratigraphic termination patterns within a predefined workflow. Consequently, the objective of the synthetic dataset is not to reproduce all possible geological histories or event-order relationships (e.g., faults truncated by unconformities), but rather to generate a broad range of geologically plausible structural geometries.

For diffusion-based generative modeling, such geometric diversity is particularly important because it exposes the model to a wide variety of structural styles during training. By covering diverse combinations of folds, faults, unconformities, and stratigraphic terminations, the dataset enables the model to learn a richer distribution of structural patterns, thereby improving its generative capability and generalization to previously unseen structural scenarios.

Following your suggestion, we have expanded the description of the synthetic modeling workflow in the revised manuscript to clarify the role of folds, faults, unconformities, and their associated parameterization.

The second paragraph of Sec 3.1:
The combination of explicit fold/fault/unconformity rules and controlled parameter randomization ensures both geological plausibility and broad structural--stratigraphic diversity across the dataset. It should be noted that the synthetic dataset is generated through a geometry-based structural simulation workflow rather than an explicit reconstruction of geological event histories. This design prioritizes structural diversity for diffusion-model training rather than reproducing specific geological evolution pathways.

Line 216: Please expand upon the explanation of the random subsampling of horizons. What percentage of the horizons were retained, and how were the removed horizons selected? This more detailed description should ideally feed directly into the major ablation study requested above.

Response:
Thank you for this valuable comment. We agree that the description of the horizon subsampling procedure was insufficient in the original manuscript and have expanded it in the revised version.

For each implicit structural model, horizon constraints are generated by first randomly selecting between 1 and 6 relative geological time (RGT) values from the normalized implicit scalar field. Horizon constraints are then extracted as iso-contours corresponding to these selected RGT values. Therefore, the horizon conditioning data are not obtained by removing pre-existing interpreted horizons, but rather by randomly sampling different iso-surfaces from the target implicit structural model. This strategy allows the model to experience a wide range of conditioning densities during training and mimics realistic interpretation scenarios in which only a limited number of horizons are available.

Following your suggestion, we have also added a dedicated horizon-sparsity ablation study (Sec 3.5), where the number of conditioning horizons is systematically reduced from 6 to 4, 2, and 1. The results demonstrate how horizon sparsity influences realization variability and structural ambiguity, directly linking the conditioning strategy to model uncertainty and generation behavior.

The third paragraph of Sec 3.1:

(2) sparse horizon constraints (h) obtained by randomly selecting between 1 and 6 relative geological time (RGT) values and extracting the corresponding iso-contours from the target implicit structural model. This strategy mimics practical interpretation scenarios where only a limited number of horizons are available as constraints (following Bi et al., 2022).

We sincerely thank you again for the careful evaluation of our work and for the constructive suggestions. We believe that the revisions made in response to these comments have significantly improved the manuscript, and we hope that the revised version adequately addresses all concerns.

Response to Reviewer2:

General Comments

1. Uncertainty quantification. The manuscript frames LFD as a generative model for the distribution of implicit structural models, which implies that the method should be capable of representing and communicating uncertainty across plausible structural realizations. This is, after all, one of the principal motivations for using a generative rather than a deterministic model in this setting. However, the manuscript does not clearly characterize the uncertainty produced by the method: it is not clear whether multiple samples from LFD given the same conditioning (horizons/faults) are shown to produce meaningfully diverse, geologically plausible realizations, how the spread of these realizations compares to the known or expected geological uncertainty in the synthetic/real test cases, and whether the prior-guided losses (horizon loss, bending-energy loss), by construction, narrow the ensemble in ways that could underrepresent genuine structural uncertainty. Given that uncertainty representation is central to the value proposition of a diffusion/flow-matching approach over deterministic implicit modeling, I recommend that the authors add an explicit quantitative and visual analysis of ensemble spread/variability (e.g., pixel-wise standard deviation maps or probability-of-occurrence maps across multiple samples) and discuss how the imposed priors affect this spread.

Response:

Thank you for your valuable suggestion. We agree that, as a generative approach, LFD should be able to represent and communicate uncertainty in the generated structural models, and that the

original manuscript lacked a quantitative analysis in this respect. To address this concern, we have added a new subsection to the Experiments section (Sect. 3.5, "Ablation Study on Uncertainty: Effects of Conditioning Sparsity and Prior-Guided Regularization Strength"), which systematically extends the uncertainty analysis along two dimensions:

(1) Quantitative characterization of multi-sample generation. With the fault and horizon conditioning fixed, we generate 20 independent realizations for each conditioning configuration by sampling different initial Gaussian noise fields, and compute the pixel-wise variance across these realizations to quantify uncertainty. We further systematically reduce the number of input horizons from 6 to 4, 2, and 1 to examine how the density of conditioning information affects realization spread. As shown in the newly added Figure 4, when horizon constraints are abundant (6 horizons), the variance remains uniformly low across the model domain, indicating that dense conditioning effectively anchors the generation. As the number of horizons decreases, high-variance regions progressively expand, first emerging in unconstrained areas between horizons and eventually spanning the majority of the model domain when only a single horizon is provided. The mean variance increases monotonically from 0.0013 (6 horizons) to 0.0073 (1 horizon), quantitatively confirming that realization spread is directly governed by conditioning density. These results demonstrate that LFD generates a data-consistent distribution: when conditioning information is abundant, the model converges to consistent solutions; when conditioning information is sparse, the model appropriately broadens its output distribution to reflect increased geological ambiguity, rather than collapsing to a single, overconfident prediction.

(2) Effect of the prior-guided regularization strength on the ensemble. To directly address whether the prior-guided losses narrow the ensemble by construction, we further investigate how the strength of the proposed fault-aware bending-energy regularization affects the realization ensemble. For each bending-energy weight $\lambda_{Bend} \in \{0, 0.1, 1\}$ ($\lambda_{Bend}=0.1$ being the default used throughout this paper), we train a separate model under the same training settings and evaluate it on the full validation set (3200 cases). For each case, we extract the central horizon of the implicit field together with the corresponding fault mask as the conditioning input, and generate 20 realizations from different noise seeds. We report the resulting pixel-wise Mean Squared Error (MSE), the Horizon Consistency Error (HCE-6, evaluated on 6 horizons uniformly extracted from the ground-truth implicit field), and the mean realization variance, averaged over the validation set, in the added Table 4.

As shown in Table 4, MSE and HCE-6 do not vary monotonically with λ_{Bend} , but both are consistently lower with the bending-energy prior enabled than without it, indicating that the regularization yields realizations that are more stable and closer to the reference model, rather than drifting into implausible solutions. In particular, $\lambda_{Bend}=0.1$ achieves the lowest pixel-wise error among the three settings. At the same time, the prior does not reduce realization variance—the strongest setting ($\lambda_{Bend}=1$) yields the highest variance among the three, showing that this improved stability does not come at the cost of the model's ability to represent structural uncertainty. Taken together with the results in (1), these findings indicate that the prior-guided losses do not narrow the ensemble in ways that would underrepresent genuine structural uncertainty.

3.5 Ablation Study on Uncertainty: Effects of Conditioning Sparsity and Prior-Guided Regularization Strength

To investigate how the number of input horizon constraints affects generation quality and realization variability, we conduct an ablation study in which the number of input horizons is systematically reduced from 6 to 4, 2, and 1 (The top row in Figure 4). For each configuration, we generate 20 independent realizations by sampling different initial Gaussian noise fields while keeping the fault and horizon conditioning fixed, and compute the pixel-wise variance across realizations to quantify uncertainty.

The middle row of Figure 4 shows the mean implicit structural models over 20 realizations generated under each sparsity level. As the number of input horizons decreases, the generated models remain geologically plausible but exhibit increasing variability in stratigraphic geometry, particularly in regions far from the remaining horizon constraints. This reflects the model's learned prior filling in the unconstrained space with a broader range of geologically consistent solutions.

The bottom row of Figure 4 shows the spatial distribution of pixel-wise variance across 20 realizations for each horizon configuration. With 6 input horizons, variance is uniformly low across the model domain, indicating that dense conditioning effectively anchors the generation. As the number of horizons is reduced to 4, 2, and finally 1, high-variance regions progressively expand, concentrating first in areas between horizons and eventually spanning the majority of the model domain. This spatial pattern is consistent with the intuition that uncertainty grows where conditioning data are absent. The mean variance plot (bottom-left panel of Figure 4) further confirms this trend quantitatively: mean variance increases monotonically from approximately 0.0013 (6 horizons) to 0.0073 (1 horizon), demonstrating that realization spread is directly controlled by the density of horizon constraints. Collectively, these results suggest that LFD generates a data-consistent distribution: when conditioning data are abundant, the model converges to consistent solutions; when data are sparse, the model appropriately broadens its output distribution to reflect the increased geological ambiguity, rather than collapsing to a single overconfident prediction.

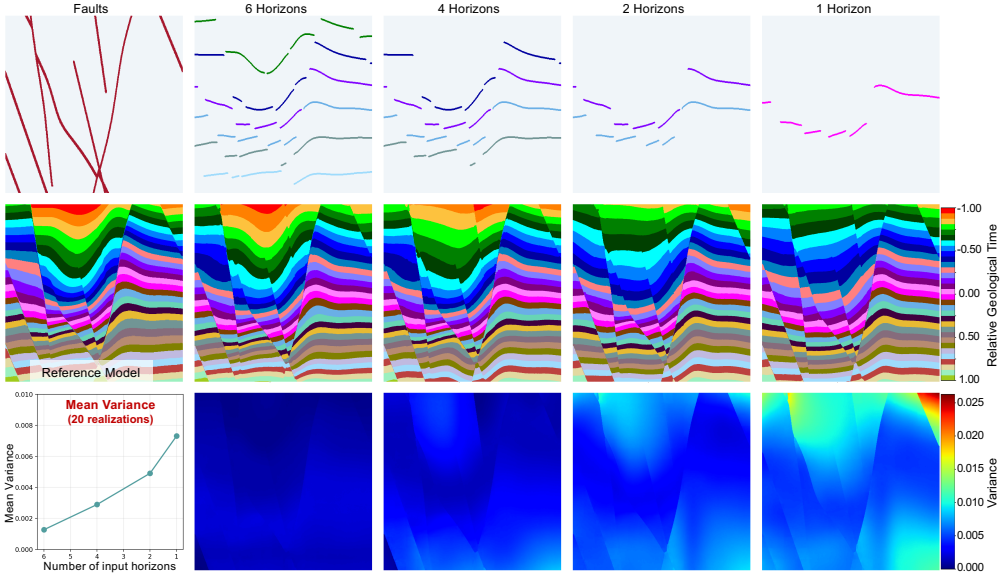

Figure 4: Ablation study on horizon sparsity and realization uncertainty. Top row: input conditioning, showing the shared fault interpretation (leftmost) and horizon inputs with 6, 4, 2, and 1 horizons. Middle row: mean implicit

structural model over 20 realizations for each horizon configuration, with the reference model shown in the leftmost panel. Bottom row: pixel-wise variance across 20 realizations (leftmost panel: mean variance as a function of the number of input horizons), demonstrating that realization variance increases monotonically as horizon constraints become sparser.

Beyond conditioning sparsity, we further investigate how the strength of the proposed fault-aware bending-energy regularization affects the realization ensemble. For each bending-energy weight $\lambda_{Bend} \in \{0, 0.1, 1\}$ (λ_{Bend} =0.1 being the default used throughout this paper), we train a separate model under the same training settings and evaluate it on the full validation set (3200 cases). For each case, we extract the central horizon of the implicit field together with the corresponding fault mask as the conditioning input, and generate 20 realizations from different noise seeds. We report the resulting pixel-wise Mean Squared Error (MSE), the Horizon Consistency Error (HCE-6, evaluated on 6 horizons uniformly extracted from the ground-truth implicit field), and the mean realization variance, averaged over the validation set, in Table 4.

Table 4. Ablation on fault-aware bending-energy weight λ_{Bend} under the sparsest conditioning setting (single input horizon), evaluated on the validation set (3200 cases, 20 realizations each). HCE-6 is computed on 6 horizons uniformly sampled from the ground-truth implicit field.

λ_{Bend}	MSE (1e-2) ↓	HCE-6 (1e-2) ↓	Mean Variance (1e-3)
0	3.73	3.65	7.63
0.1	2.93	2.84	7.27
1	3.50	3.36	8.61

As shown in Table 4, MSE and HCE-6 do not vary monotonically with λ_{Bend}, but both are consistently lower with the bending-energy prior enabled than without it, indicating that the regularization yields realizations that are more stable and closer to the reference model, rather than drifting into implausible solutions. In particular, λ_{Bend} =0.1 achieves the lowest pixel-wise error among the three settings. At the same time, the prior does not reduce realization variance; the strongest setting (λ_{Bend} =1) yields the highest variance among the three, showing that this improved stability does not come at the cost of the model's ability to represent structural uncertainty.

2. Reproducibility of the code. I attempted to run the code accompanying the manuscript and was unable to get it to execute. As the manuscript's central claims rely on specific implementation choices (e.g., the structure-enhanced Transformer conditioning, the fault-aware finite-difference stencils in Eq. 14, and the reported inference time of 1.56 s for a 512×512 model on an NVIDIA H20 GPU), I am not currently able to independently verify these claims. I would ask the authors to (a) verify that the uploaded code corresponds to the exact version used to produce the results in the manuscript, (b) provide a minimal working example with pinned dependency versions and, if possible, a small pretrained checkpoint or synthetic toy dataset that reproduces at least one figure/table end to end. Given GMD's emphasis on reproducibility of model development, this should be resolved before publication.

Response:
Thank you for raising this important issue, and for taking the time to attempt to run our code. We agree that the original repository lacked sufficient documentation to support reproduction of our results. In response, we have thoroughly revised the code repository as follows.

(a) We have verified and updated the repository to ensure that the uploaded code corresponds exactly to the version used to produce the results reported in the manuscript.

(b) We have substantially expanded the README documentation to include detailed environment setup instructions, covering pinned dependency versions, data preparation procedures, path configuration, and inference guidelines. We have also uploaded a pretrained checkpoint together with the example real-survey data used in the manuscript to Zenodo, enabling the reviewer and other readers to reproduce the inference workflow and results without retraining.

The revised repository is now available on Zenodo (https://zenodo.org/records/20508635) and GitHub (https://github.com/ProgrammerZXG/LFD).

3. Uniform-grid assumption underlying Eq. 14. The fault-aware bending-energy term in Eq. 14 is written in terms of pixel-index derivatives ($\partial^2\hat{x}/\partial i^2$, $\partial^2\hat{x}/\partial i\partial j$, $\partial^2\hat{x}/\partial j^2$). This formulation implicitly assumes that the grid spacing is uniform and identical in both the i- and j-directions (i.e., an isotropic, regular sampling grid), since otherwise the index-space second differences do not correspond to a consistent physical curvature. This assumption should be stated explicitly, since implicit structural models are frequently defined on grids with unequal vertical/horizontal spacing, or on locally warped/flattened grids. I would ask the authors to: (a) explicitly state whether the ViT patch grid used by LFD is always uniform and isotropic in physical units; (b) if not, clarify whether the derivatives in Eq. 14 are normalized by the local grid spacing (e.g., divided by Δi^2, $\Delta i\Delta j$, Δj^2, respectively); and (c) provide at least one experiment or discussion addressing whether/how the method would behave on non-uniform grids, since this materially affects the physical meaning of the bending-energy penalty and, by extension, the geological plausibility of the generated structures.

Response:
Thank you for this important observation. We agree that Eq. 14, as originally written in terms of pixel-index derivatives, is not fully rigorous, since it does not explicitly account for the physical grid spacing in the i- and j-directions. We have revised Eq. 14 as follows:

$$\mathcal{L}_{\text{Bend}} = \mathrm{E}_{(i,j)\notin\Omega_{\text{Faults}}}\left[\left(\frac{1}{\Delta i^2}\frac{\partial^2\hat{\mathbf{x}}}{\partial i^2}\right)^2 + 2\left(\frac{1}{\Delta i\,\Delta j}\frac{\partial^2\hat{\mathbf{x}}}{\partial i\,\partial j}\right)^2 + \left(\frac{1}{\Delta j^2}\frac{\partial^2\hat{\mathbf{x}}}{\partial j^2}\right)^2\right],$$

where Δi and Δj denote the grid spacing in the i- and j-directions, respectively.

This revision makes the bending-energy term correspond to a consistent physical curvature. We clarify that the present study focuses on validating the effectiveness of the proposed x-prediction formulation, structure-enhanced conditioning, and prior-guided losses under a uniformly sampled grid, consistent with the fixed-size patch embedding used by the ViT backbone. Extending the framework to non-uniform or locally warped grids is an important direction, but would require a

different patch embedding scheme to account for spatially varying grid spacing, which is beyond this scope. We have noted this as a limitation and a direction for future work in the Discussion section.

Limitation and 3D extension.

Similarly, the present study assumes a uniformly sampled grid, consistent with the fixed-size patch embedding used by the ViT backbone; extending the framework to non-uniform grids, where spacing varies spatially (e.g., locally warped or flattened grids), would require rethinking the patch embedding scheme, and represents another promising direction for future work.

4. Sensitivity to noise level, number of sampling steps, and initial state. The manuscript reports a single operating point for the flow-matching/diffusion sampler (e.g., a fixed number of integration steps and a fixed initialization) without a systematic sensitivity analysis. Since LFD is positioned as a generative model whose outputs should be both efficient and geologically reliable, I believe such an analysis is necessary. Specifically: (a) Noise level/integration step size: flow-matching and diffusion samplers are known to be sensitive to the discretization of the noise/time schedule, and coarser schedules can introduce integration error that is not obviously distinguishable from genuine multi-modal geological variability. The authors should report how the quality of the generated structural models (fidelity to horizons/faults, bending-energy statistics, and any other structural metrics used) varies as the step size/noise schedule is coarsened or refined. (b) Number of sampling steps: relatedly, the manuscript should report how output quality and inference time trade off as the number of steps is varied, since the headline efficiency result (1.56 s for a 512×512 model) is only meaningful in the context of a stated accuracy level, and readers need to know whether fewer steps could be used without materially degrading geological plausibility, or whether more steps are required to reach convergence. (c) Initial state: because LFD generates directly in data space rather than in a compact latent space, the choice of initial noise/state may have a more direct and visible effect on the final structural model than it would in latent diffusion. The authors should clarify whether the initial state is drawn from a fixed prior (e.g., standard Gaussian noise) for every sample, and, if so, present an ablation showing how the diversity and quality of generated structural models depend on the initial state. Without this sensitivity analysis, it is difficult to assess how robust the reported results are to these sampler-level choices, and whether the reported timing/quality figures represent a carefully chosen operating point or a somewhat arbitrary one.

Response:

Thank you for this insightful and important comment. We agree that reporting a single operating point without a systematic sensitivity analysis makes it difficult for readers to assess whether the reported efficiency and quality figures represent a carefully considered trade-off or a somewhat arbitrary choice, particularly given that LFD is positioned as a generative model expected to be both efficient and geologically reliable. We address each point in turn.

(a) Noise level/integration step size. We agree that flow-matching and diffusion samplers can indeed exhibit different sensitivities under different time-scheduling strategies. In our implementation, the ODE time interval is discretized uniformly ($\Delta t=1/K$, Eq. 6), so the integration step size and the number of sampling steps are jointly determined by the same parameter K: specifying one uniquely

determines the other, and the two are not independent variables. We address this point through the systematic ablation on K presented in (b).

In addition, the original manuscript already includes an ablation on the noise scale (see "Impact of noise scale" in the Discussion section), in which we vary the noise scale from 2.0 to 0.1 while keeping all other settings fixed. The results show that decreasing the noise scale yields noticeably smoother and more geologically reasonable implicit fields, whereas larger noise scales commonly adopted in natural-image diffusion (e.g., 1.0 or 2.0) lead to visibly noisier generations. This finding also provides a useful reference for other generative tasks in geophysics: the noise scale should be tuned to the characteristics of the target data rather than directly adopted from conventions established for natural images.

Impact of noise scale in Sec Discussion

We find that the noise scale is a critical factor for implicit structural modeling. In early experiments, using commonly adopted noise scales for natural-image diffusion (e.g., 1.0 or 2.0) often led to overly noisy generations even with prolonged training. This is likely because implicit structural models exhibit relatively smooth distributions, for which excessively strong corruption can hinder stable denoising. We perform an ablation study on the noise scale and observe that decreasing the noise scale yields noticeably smoother and more geologically reasonable implicit fields. This finding provides practical guidance for deploying diffusion models in geophysics: the noise scale should be tuned to the target task and data characteristics to achieve optimal generation quality.

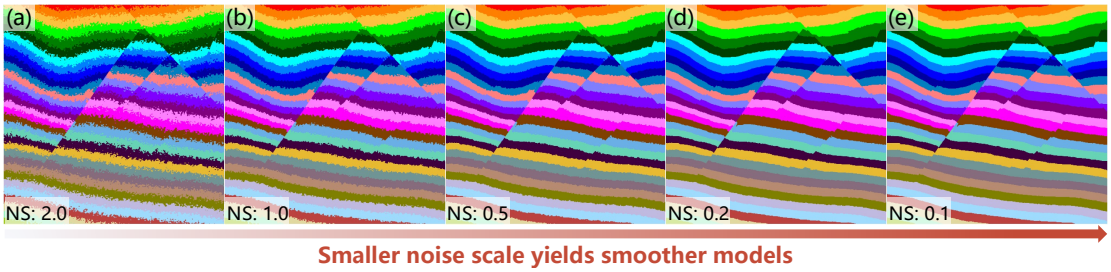

Figure 6. Effect of noise scale (NS) on generation. The model is trained with noise scale NS=2.0. During inference, we vary only the noise scale while keeping all other settings fixed. Panels (a--e) correspond to NS=2.0,1.0,0.5,0.2,0.1. As NS decreases, the generated implicit structural models become smoother and exhibit less residual noise.

(b) Number of sampling steps. We conduct an ablation on the number of ODE sampling steps $K \in \{5,10,20,30,50\}$, evaluated on the full validation set, using the fault mask together with the central horizon extracted from the implicit field as conditioning, and reporting the average inference time (see "Impact of sampling steps" in the Discussion section). The results show that inference time scales approximately linearly with K, while both MSE and HCE improve only marginally beyond K=20, with diminishing returns thereafter, indicating that geologically plausible, high-quality results can already be obtained with substantially fewer steps (e.g., 10 or 20). The default setting used throughout this paper (K=50) is intended to showcase the best achievable generation quality; in practice, the number of sampling steps can be reduced to achieve substantially faster inference while maintaining comparable quality. We thank the reviewer for raising this point, which has led us to characterize the applicable range of the reported efficiency figure more precisely and rigorously.

Impact of sampling steps in Sec Discussion

In addition to the noise scale, we examine the sensitivity of generation quality to the number of ODE sampling steps K∈{5,10,20,30,50}, evaluated on the full validation set, using the fault mask together with the central horizon extracted from the implicit field as conditioning, and reporting the average inference time. As shown in Table 5, inference time scales approximately linearly with K, while both MSE and HCE-6 improve only marginally beyond K=20, with diminishing returns thereafter. The default setting used throughout this paper (K=50) is chosen to showcase the best achievable generation quality; in practice, the number of sampling steps can be reduced to obtain substantially faster inference while maintaining comparable quality.

Table 5. Ablation on the number of ODE sampling steps K.

K	MSE (1e-2) ↓	HCE-6 (1e-2) ↓	Time (s)
5	3.18	3.016	0.15
10	3.16	3.021	0.32
20	3.13	3.025	0.64
30	3.08	2.950	0.97
50	3.04	2.967	1.56

(c) Initial state. We confirm that the initial state z_0 is drawn from a fixed standard Gaussian prior, $z_0 \sim N(0, I)$ for every generated sample. The horizon-sparsity ablation presented in **Sec 3.5 Ablation Study on Uncertainty** already provides the relevant analysis: for each conditioning configuration, we generate 20 realizations from different initial noise samples and report the resulting pixel-wise variance. The results show that these realizations exhibit meaningful diversity that scales with the sparsity of the conditioning, rather than collapsing to a near-identical output regardless of the sampled initial state, indicating that the model's generation quality is not overly sensitive to the specific noise realization drawn at inference time.

The Second Paragraph of Sec 3.5

We note that, since these 20 realizations are generated from different initial noise samples, this also indicates that the sampled initial noise has a visible influence on the resulting structural details, with different noise samples leading to different specific realizations, while the overall outputs remain geologically plausible.

Specific Comment: The Second-Order Finite-Difference Bending-Energy Loss (Eq. 14)

1. Grid metric is not accounted for. As noted above, Eq. 14 is written entirely in index space. Even setting aside non-uniform grids, vertical and horizontal sampling intervals in seismic/structural volumes are typically different by construction, so the mixed term implicitly assumes an isotropic metric that generally does not hold. The authors should clarify whether/how the loss accounts for anisotropic or spatially varying grid spacing, and, if it does not, discuss the resulting bias.

Response:

Thank you for raising this point. We agree that Eq. 14, as originally written in terms of pixel-index derivatives ($\partial^2\hat{x}/\partial i^2$, $\partial^2\hat{x}/\partial i\partial j$, $\partial^2\hat{x}/\partial j^2$), is not fully rigorous, since it does not explicitly account for

the physical grid spacing in the i- and j-directions and implicitly assumes an isotropic metric (i.e., $\Delta i = \Delta j$). As the reviewer points out, this assumption generally does not hold, particularly for structural volumes, where the vertical and horizontal sampling intervals typically differ by construction. We have revised Eq. 14 to incorporate the corresponding grid-spacing terms, normalizing the second-order derivatives by Δi^2, $\Delta i \cdot \Delta j$, and Δj^2, respectively, where Δi and Δj denote the grid spacing in the i- and j-directions. This revision makes the bending-energy term correspond to a consistent physical curvature, without implicitly assuming that the spacing in the two directions is equal.

We further clarify that the revised formulation assumes the grid is regular (i.e., Δi and Δj are constant throughout the domain) but does not require it to be isotropic (i.e., $\Delta i = \Delta j$), which is consistent with the synthetic and field datasets used in this study. Non-uniform grids, where spacing varies spatially (e.g., locally warped or flattened grids), remain outside the scope of the current formulation. We have explicitly noted this as a limitation and a direction for future work in the Discussion section ("Limitation and 3D extension").

2. The "clean" prediction is still an uncertain estimate at high noise levels, and this is where curvature penalties are most fragile. The manuscript states that "as the model directly predicts the clean data, geological observations and prior constraints can be imposed in a straightforward and differentiable manner", and applies this prediction. I would like to point out that even though is nominally the "clean" target, at early sampling steps / high noise levels the network's prediction of given a heavily corrupted input is fundamentally underdetermined: many plausible clean structural models are consistent with the same noisy observation, and the network's point estimate at this stage is typically a blurred, low-confidence average over these possibilities rather than a single coherent geological structure. Its discrete second-order curvature at this stage therefore reflects the shape of the network's predictive uncertainty, not a genuine geological feature (or lack thereof). Penalizing that curvature is a much less well-posed operation than penalizing the curvature of a final, high-confidence output, and could bias the network toward predicting overly smooth estimates at high noise levels purely to reduce this loss term, independent of whether the final sample should be smooth. I would ask the authors to clarify: (a) at which noise levels/timesteps it is applied during training (all, or only near the clean end of the schedule), and (b) whether the loss weighting is scheduled/annealed as a function of noise level to account for the fact that it is far less reliable at high noise levels. If the loss is applied uniformly across all noise levels, I would ask for evidence (e.g., a comparison of results with vs. without this weighting) that it is not simply driving the network toward generically smoother predictions rather than toward geologically faithful ones.

Response:
Thank you for raising this insightful point, which prompted us to take a closer look at how this loss term behaves throughout training.

(a) At which noise levels/timesteps this loss is applied. In the current implementation, the bending-energy loss is computed and applied at every training iteration, regardless of the sampled timestep t (i.e., the noise level) — that is, the loss is applied uniformly across all noise levels rather than only near the clean end of the schedule.

(b) Whether the loss weighting is scheduled/annealed as a function of noise level. In the current implementation, the weight λ_{Bend} is a fixed constant and is not scheduled or annealed with respect to noise level.

We agree with this mechanistic observation: at high noise levels, the network's prediction of the clean data $\hat{x}$ is essentially an averaged estimate over multiple plausible structures, which is mathematically expected behavior. To illustrate this phenomenon directly, we performed a qualitative analysis in which the same ground-truth sample x and the same noise sample ε are fixed while only the timestep t is varied, examining the network's single-step prediction of $\hat{x}$ at different noise levels (see Figure 7). The results show that as t increases (i.e., as the noise level decreases), the prediction error decreases monotonically, and the clarity and structural detail of the prediction improve accordingly, confirming that predictions are indeed relatively more blurred at high noise levels.

Nevertheless, our results show that, even at high noise levels, the network is still able to recover a geologically plausible structure rather than a severely blurred or incoherent one, indicating that the reliability of the prediction at this stage, while lower than at low noise levels, remains sufficient to provide a meaningful geological signal. This suggests that applying the bending-energy regularization at high noise levels is still meaningful and does not severely compromise the final results. This is further supported quantitatively: as shown in Table 4, reconstruction metrics (MSE, HCE-6) improve rather than deteriorate as the loss weight increases from 0 to 1, while the variance across realizations does not decrease monotonically with increasing weight, with $\lambda_{Bend}=1$ in fact yielding the highest variance among the three settings. This indicates that the loss does not exhibit a systematic tendency to drive the network toward globally smoother outputs, and that its practical impact under the current weighting and training configuration does not compromise the geological plausibility of the final generated results.

We also acknowledge that a more refined weighting schedule, in which the strength of the bending-energy regularization decays with the timestep, could enable finer control over this trade-off. The corresponding discussion has been added to the Discussion section.

Reliability of clean-data prediction at high noise levels. Since LFD directly predicts the clean data $\hat{x}$ at every timestep, it is worth examining how the reliability of this prediction varies with the noise level, as this bears on the validity of applying the horizon and bending-energy losses uniformly across all timesteps during training. Using a single validation sample, we fix the same implicit field and noise realization while varying only the timestep t, and visualize the resulting single-step prediction $\hat{x}$ (Figure 7). As t increases from 0.1 to 0.9 (i.e., as the noise level decreases), the prediction error decreases by nearly two orders of magnitude (from 9.77×10^{-4} to 1.47×10^{-5}), and the predicted field becomes visibly sharper. This confirms that predictions are more blurred at high noise levels. In principle, geological and geophysical regularization terms such as the horizon and bending-energy losses would be more directly justified at low noise levels, where $\hat{x}$ more closely approximates the implicit model. However, our results show that, even at high noise levels, the network is still able to recover a geologically plausible structure rather than a severely blurred or incoherent one. This suggests that, for the implicit structural modeling task considered here, applying these prior losses at high noise levels remains reasonably well-justified. In this paper, the same regularization

weight is applied uniformly across all noise levels t during training. Building on this observation, a promising direction for future work is to design a more refined weighting schedule in which the strength of the bending-energy regularization decays with the timestep, which may enable finer control over the trade-off between reconstruction fidelity and geological plausibility.

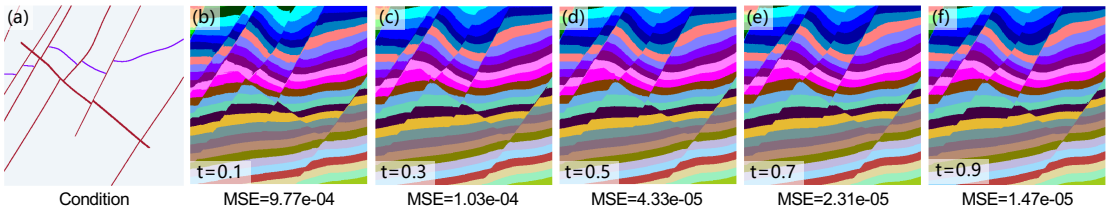

Figure 7 Single-step clean-data prediction at different noise levels. (a) Input conditioning. (b-f) Predicted x at t = 0.1, 0.3, 0.5, 0.7, 0.9, using the same ground-truth sample and the same noise realization across all panels. Prediction error (MSE, relative to the ground-truth field) decreases monotonically as t increases, indicating that predictions become progressively more reliable as the noise level decreases.

3. Inconsistent stencil order near fault-mask boundaries. "Fault-aware stencils that do not cross fault pixels" implies one-sided or truncated finite differences adjacent to. Standard central differences are second-order accurate, whereas one-sided or truncated stencils are typically only first-order accurate or have a different truncation-error constant. This means the effective smoothness penalty is not spatially homogeneous: pixels near a fault are regularized with a different bias/strength than interior pixels, purely as an artifact of discretization rather than of the underlying geology. Please specify the exact one-sided/modified stencils used near fault boundaries and confirm that they remain numerically consistent (i.e., that they converge to the correct derivative as), rather than being a central stencil with missing terms simply set to zero, which would not be consistent.

Response:

Thank you for raising this point. We clarify that the fault-aware treatment adopted in this work is not a one-sided or truncated finite-difference scheme, as the reviewer's comment surmised, but rather a stencil-validity masking strategy: for each second-order derivative term ($\partial^2\hat{x}/\partial i^2$, $\partial^2\hat{x}/\partial i\partial j$, $\partial^2\hat{x}/\partial j^2$), the curvature contribution at a given location is computed only if all sampling points involved in the corresponding stencil lie in the non-fault region; if any point of the stencil falls on a fault pixel, the curvature term at that location is set to zero and does not contribute to the loss.

Specifically, we do not replace the interior stencil with a lower-order one-sided or truncated approximation near fault boundaries. Consequently, the concern raised by the reviewer, namely that pixels near a fault would be regularized with a different bias or strength than interior pixels purely as an artifact of discretization, does not arise in our implementation. Pixels near a fault either use the exact same standard central-difference stencil as interior pixels, or, when the stencil crosses the fault, are entirely excluded from this loss term. In neither case does the accuracy of the stencil itself change near a fault.

We agree that this "exclusion" strategy, as opposed to substituting a lower-order one-sided scheme, means that a narrow band of pixels immediately adjacent to a fault (spanning the width of the finite-

difference stencil) is not constrained by this regularization term. We note, however, that this effect is highly localized, spanning only a few pixels corresponding to the stencil width. Moreover, this region remains constrained by the other training objectives, including the flow matching loss and the horizon loss (where a horizon constraint passes through the region). We further note that a fault itself represents a genuine discontinuity in the geological structure, so not enforcing a continuous-curvature constraint immediately at the fault also avoids imposing an incorrect continuity assumption at a location where the field is not expected to be continuous. Consistent with this, we do not observe any visible artifacts or geologically implausible structures near fault boundaries in our results on field data (Figure 3), suggesting that this localized effect does not compromise the geological plausibility of the generated structures in practice. We have clarified this implementation detail in the revised manuscript.

Specifically, for each second-order derivative term, we adopt a stencil-validity masking strategy: the curvature contribution at a given location is computed only if all sampling points involved in the corresponding standard central-difference stencil lie in the non-fault region; if any point of the stencil falls on a fault pixel, the curvature term at that location is set to zero rather than being approximated with a lower-order one-sided or truncated scheme. As a result, the accuracy of the stencil itself does not vary near fault boundaries.

4. Risk of over-smoothing genuine, non-fault geological curvature. A quadratic penalty on bending energy favors globally low curvature (thin-plate-like behavior). However, fold hinges, drag folds, and tight synclines/anticlines away from faults are geologically real, legitimate high-curvature features that are not associated with a discontinuity. A blanket quadratic penalty risks over-smoothing these features unless the loss weight is very carefully tuned. I would ask for an ablation showing the sensitivity of the results to the weight of , and specifically its effect on fold-heavy synthetic or real test cases, together with a discussion of whether a more robust penalty (e.g., an L1 curvature term or edge-preserving weighting) was considered and, if so, why it was not adopted.

Response:
Thank you for raising this important point. We agree that a quadratic bending-energy penalty inherently favors globally low curvature (thin-plate-like behavior), and that, if the loss weight is not carefully chosen, there is in principle a risk of over-smoothing genuine, non-fault-related high-curvature geological features, such as fold hinges, drag folds, and tight synclines/anticlines.

Sensitivity ablation on the loss weight and its effect on high-curvature scenarios. We conducted a systematic ablation on the bending-energy weight $\lambda_{Bend} \in \{0, 0.1, 1\}$ under the sparsest conditioning setting (a single input horizon) (Table 4). The results show that reconstruction accuracy metrics (MSE, HCE-6) improve as the weight increases, while the realization variance does not decrease monotonically with increasing weight. In addition, we constructed several high-curvature test cases in which the fold hinges are located away from any fault (including tight anticlines and asymmetric folds), and generated results under the same three weights for comparison (Figure 8). The results show that, even under the strongest setting (λ_{Bend}=1), the model is still able to accurately preserve the high-curvature geometry at these fold hinges, with no noticeable over-smoothing or structural

collapse observed. This indicates that, within the range of weights tested in this study, the regularizer does not come at the cost of genuine, non-fault-related high-curvature geological features.

Whether a more robust penalty (e.g., an L1 curvature term or edge-preserving weighting) was considered. We acknowledge that this work does not adopt an L1 curvature penalty or edge-preserving weighting, nor have we systematically compared such alternatives against the current quadratic penalty. The quadratic (thin-plate-like) bending-energy penalty adopted here follows the conventional form of this type of regularizer commonly used in implicit structural modeling (e.g., Renaudeau et al., 2019; Belhachmi et al., 2025). We agree that L1 curvature terms or edge-preserving weighting schemes, which are more tolerant of localized high-curvature features, could further reduce the risk of over-smoothing genuine fold structures, and represent a promising direction for systematic investigation. We have identified this comparison as a direction for future work and noted it accordingly in the Discussion section.

Preservation of high-curvature features. While the bending-energy loss suppresses geologically implausible artifacts, a quadratic penalty on curvature could, in principle, also over-smooth genuine, non-fault high-curvature features such as fold hinges. To examine this risk, we constructed several high-curvature test cases in which the fold hinges are located away from any fault (including a tight anticline, an asymmetric fold, and a double-hinge fold without any fault present), and generated results under three bending-energy weights ($\lambda_{Bend} \in \{0, 0.1, 1\}$) for comparison (Figure 8). Even under the strongest setting ($\lambda_{Bend} = 1$), the model accurately preserves the high-curvature geometry at these fold hinges, with no noticeable over-smoothing or structural collapse observed. Together with the quantitative results in Table 4, where reconstruction accuracy improves and realization variance does not decrease monotonically with λ_{Bend}, this indicates that, within the tested weight range, the regularizer does not compromise high-curvature geological features while still preserving meaningful realization diversity.

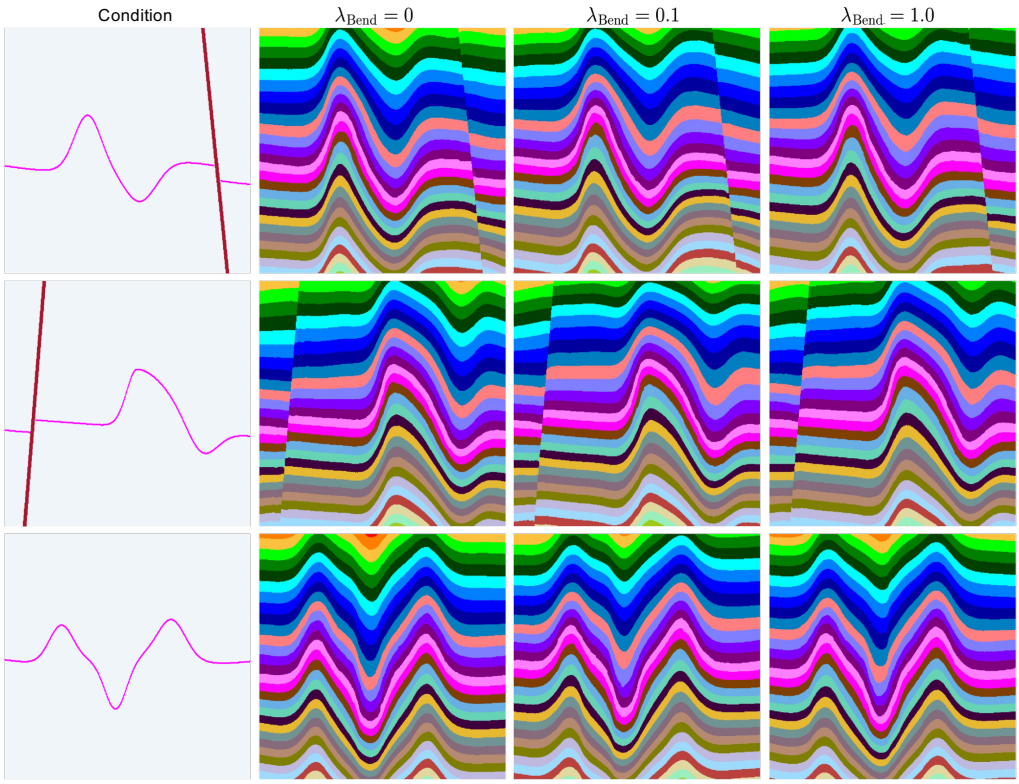

Figure 8. Sensitivity of high-curvature fold hinges to the bending-energy weight. Left column: input conditioning for three test cases, each featuring a fold hinge positioned sufficiently far from any fault: a tight anticline (top), an asymmetric fold (middle), and a double-hinge fold without any fault present (bottom). Right three columns: ensemble-mean structural models, averaged over 20 realizations from different noise seeds, generated under $\lambda_{Bend} = 0$, 0.1, 1.0, respectively.

We would like to express our sincere appreciation for your thorough assessment of our manuscript and for the valuable comments and suggestions provided. We have carefully revised the manuscript accordingly and believe that these changes have substantially enhanced its clarity and overall quality. We hope that the revised manuscript has satisfactorily addressed all the concerns raised.