Reply to referee 1

We thank the Reviewer for the comments and suggestions on our manuscript.

We have responded to most of the Reviewer's comments in our online reply dated 9 August 2025 (at the end of this letter, for reference). Following the Reviewer's suggestion, we conducted comprehensive numerical experiments using three different neural network architectures, namely a UNet with attention mechanism, a standard UNet, and a convolutional autoencoder with ResBlocks. Each network was trained both deterministically and using diffusion-based training. The numerical results are provided in Appendix D (p. 32) of the revised manuscript and are summarised below.

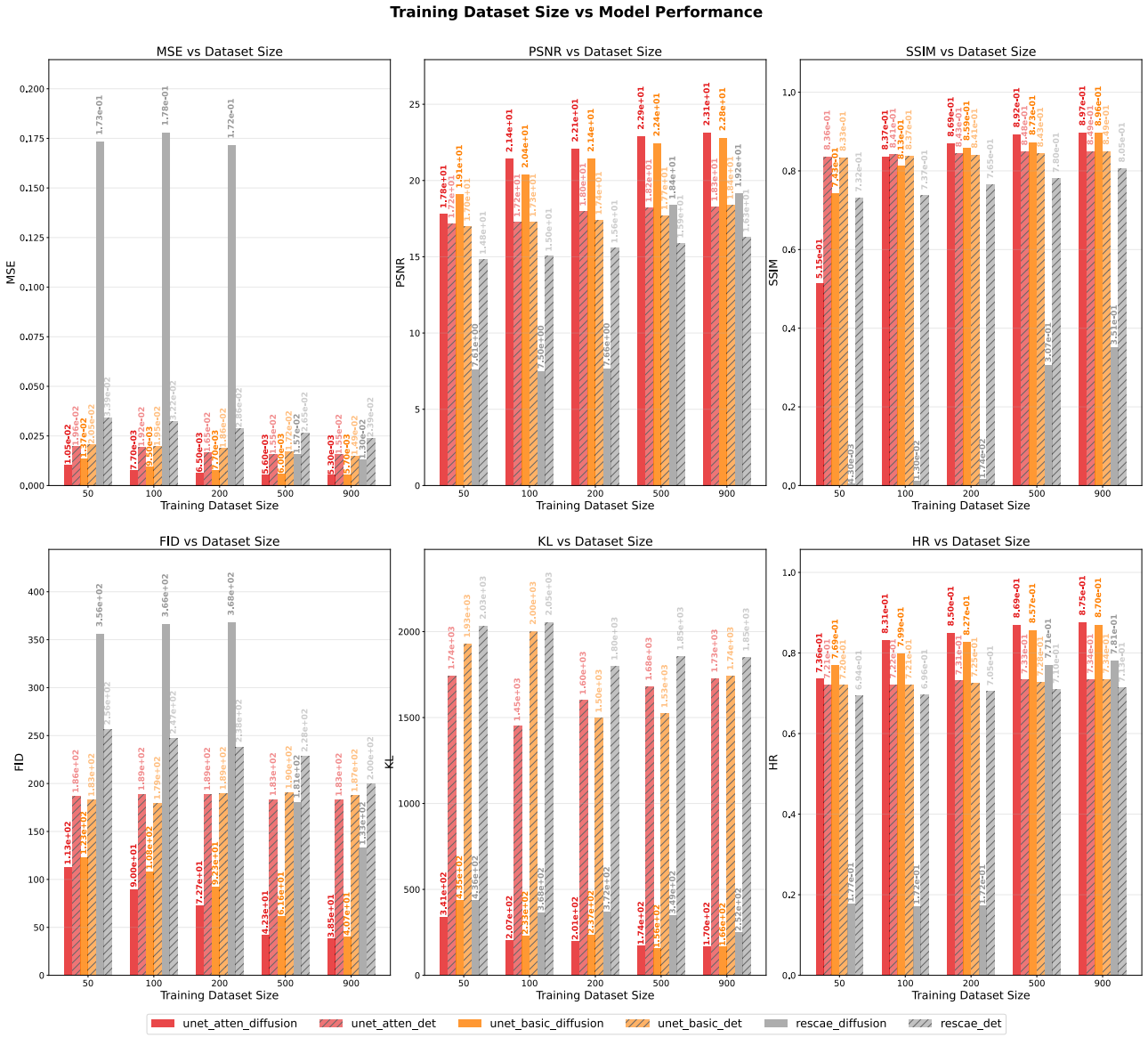

Figure 1: Ablation study comparing model performance across different training dataset sizes and architectures.

Our ablation study evaluates three distinct neural network architectures across varying training dataset sizes {50, 100, 200, 500, and 900 samples}, comparing stochastic diffusion models against deterministic approaches, as shown in Figure 1. The architectures under scrutiny include: (1) a UNet architecture with attention mechanisms (unet_atten) as employed in our

main study; (2) a UNet basic architecture (unet_basic), representing a simplified variant without attention blocks; and (3) a residual AutoEncoder architecture (rescae), which maintains a similar network structure to the UNet but removes skip connections to assess their contribution to model performance. All models were trained using identical hyperparameter configurations, including a learning rate of 1e−5 (selected from candidates $1e{-}3, 1e{-}4, 1e{-}5$ as the optimal choice), 200 training epochs, and the AdamW optimiser with weight decay of 1e−4. All models are evaluated on the ensemble test dataset of the Chimney fire event.

The results in Figure 1 demonstrate that the UNet Architecture with attention mechanisms generally outperforms the UNet basic architecture without attention blocks, though the improvement varies across metrics and dataset sizes. For instance, at training dataset size 500, the attention-enhanced UNet achieves superior performance in MSE (5.60×10^{-3} vs 6.00×10^{-2}), SSIM (8.92×10^{-1} vs 8.73×10^{-1}), and FID (4.23×10^{1} vs 6.16×10^{1}) whilst showing comparable performance in other metrics such as PSNR. However, at smaller dataset sizes like 100, the performance differences become less pronounced, with some metrics showing marginal improvements whilst others exhibit comparable or slightly inferior performance. In contrast, the comparison between UNet architectures and the residual AutoEncoder architecture reveals more substantial performance differences. The UNet structures consistently demonstrate significant improvements across multiple metrics, highlighting the importance of skip connections in preserving fine-grained information throughout the encoding-decoding process.

The related source code, scripts, data, and experimental results have been uploaded to Zenodo (`https://zenodo.org/records/15699653`) (Yu et al., 2025). The experimental results can be found in the `out` directory.

Following our previous response, it is important to clarify that the primary focus of this study is not to establish the superiority of various model architectures, but rather to investigate the advantages of diffusion-trained stochastic models over deterministic models in simulating wildfire uncertainty. This finding is further supported by our new numerical experiments: across all three neural network architectures, diffusion-based ensemble predictions (bar chart with slashes) substantially outperform their deterministic counterparts (bar chart without slashes) as shown in Figure 1.

Our previous response (copied here for reference)

We thank the reviewer for the detailed comments and suggestions on our manuscript. However, we believe that some key aspects of our work may have been overlooked.

1. The reviewer has repeatedly suggested that a benchmark comparing "DDPM vs. ConvLSTM (or other NNs)" is necessary for the manuscript. We would like to first clarify that sampling algorithms (e.g., DDPM or DDIM) and neural network architectures (e.g., ConvLSTM or U-Net) are fundamentally two different things. Sampling algorithms define how noise is added during the forward diffusion process and removed during the reverse (denoising) process (e.g., markovian in the case of DDPM and deterministic non-markovian in the case of DDIM, see Austin et al. (2021); Song et al. (2022), whereas different neural network architectures (such as U-Net or Transformer) could be chosen to train this denoising procedure. We believe the reviewer may be confused about this fundamental concept. We can not compare a sampling/denoising method against a neural network structure.

The main objective of our experiments here is to compare a diffusion-based generative training algorithm with the deterministic training method (based on MSE) for wildfire prediction,

rather than to evaluate different neural network architectures. Therefore, we compared the performance of a conditional diffusion model based on U-Net to that of the same U-Net trained using a deterministic approach. In addition, using a different neural network architecture might improve the accuracy of deterministic training, but it would not provide probabilistic predictions or capture the uncertainty of fire propagation. And also, the new network architecture will likely improve the diffusion model's performance as well. This does not qualitatively affect our comparison of diffusion and deterministic training.

We thank the reviewer for this question and will clarify the differences between neural network architectures and diffusion sampling algorithms for non-expert readers in ML.

2. Regarding the novelty of our work, although we agree that diffusion models have recently been applied in geoscience, to the authors' knowledge, this is the first study to apply diffusion-based generative AI to wildfire spread prediction (see a recent review paper by Xu et al. (2025)). In fact, to our knowledge, only one previous GMD publication (Elena Tomasi et al., April 2025) has applied a latent diffusion model to a downscaling task. Therefore, we believe that our paper is the first to use a conditional diffusion model for dynamical-system prediction in GMD.

More importantly, our diffusion model is trained using data generated from a stochastic simulator of wildfire. Therefore, we examine if the ensemble generated by the diffusion model could represent the stochasticity of the original physics model, which brings a unique contribution and insight to the community. We have also designed a specific validation procedure to compare the two ensembles generated by the stochastic physics model and the diffusion AI model, as described in Section 2.2.2 and illustrated in Figures 3 and 7 of our manuscript.

We believe that developing a surrogate model using diffusion-based generative method to capture uncertainties in stochastic physics simulators is novel within geoscience, if not in the broader computational physics field.

Following the reviewer's suggestion, we will perform additional hyperparameter tuning in the revised manuscript to improve our diffusion model's performance. However, as noted, our primary objective is to demonstrate a generative diffusion model's ability to capture the stochasticity of the physics-based model, which our current results already successfully achieve.

3. The reviewer repeatedly refers to DDPM as our denoising approach and points out its computational inefficiency. However, in our manuscript we employ the DDIM algorithm, as clearly stated in the first sentence of Section 3.1.4, in Equation 9 on page 14, and in Algorithm 2 on page 15. We also explain our choice of DDIM over DDPM, indeed specifically for its superior computational efficiency, in Section 3.1 on page 15. Thus, we believe the reviewer may have overlooked some important statements in our methodology section.

Bibliography

Austin, J., Johnson, D. D., Ho, J., Tarlow, D., and van den Berg, R.: Structured Denoising Diffusion Models in Discrete State-Spaces, in: Advances in Neural Information Processing Systems, vol. 34, pp. 17 981–17 993, Curran Associates, Inc., 2021.

Song, J., Meng, C., and Ermon, S.: Denoising Diffusion Implicit Models, https://doi.org/10.48550/arXiv.2010.02502, 2022.

Xu, Z., Li, J., Cheng, S., Rui, X., Zhao, Y., He, H., Guan, H., Sharma, A., Erxleben, M., Chang, R., et al.: Deep learning for wildfire risk prediction: Integrating remote sensing and

environmental data, ISPRS Journal of Photogrammetry and Remote Sensing, 227, 632–677, 2025.

Yu, W., Ghosh, A., Finn, T., Arcucci, R., Bocquet, M., and Cheng, S.: A Probabilistic Approach to Wildfire Spread Prediction Using a Denoising Diffusion Surrogate Model, https://doi.org/10.5281/zenodo.15699653, 2025.