Dear reviewers,

We would like to thank you for careful reading our manuscript and providing constructive comments. The comments led us to reconsider not only several technical details, but also the way the study was framed and evaluated. We therefore made substantial revisions throughout the manuscript. The main changes are summarized below:

(1) We reformulated ANNSim as a learned conditional stochastic simulator that complements SGSim rather than replacing it, and clarified that its sequential realizations are generated from a learned conditioned joint model rather than being exact realizations of the unknown underlying random field.

(2) We removed the previous "neural process" framing and now describe the model directly through its actual components.

(3) We replaced the Gaussian output with a one-dimensional rational-quadratic-spline normalizing flow to avoid restricting the conditional distribution to a Gaussian family. This allows ANNSim to represent more flexible conditional density shapes while retaining an explicit normalized probability density for likelihood-based training and stochastic sampling.

(4) We added a clearer formulation of sequential simulation, context masking, path-sensitivity analysis, and probabilistic calibration.

(5) We redesigned the main experiments around controlled covariance-model misspecification and amortized computational cost, so that the contribution of the proposed model is made more clear.

(6) We added a real Cu geochemical case study to demonstrate the practical applicability of ANNSim and moved the detailed comparison of eight spatial representations to the Supplement.

Despite these revisions, the central idea of ANNSim remains unchanged: spatial dependence can be learned directly from local observations and then used in sequential conditional stochastic simulation.

We hope you will be satisfied with the revisions.

Best regards,

Jian Wang, Renguang Zuo, Dazheng Huang, Mingang Liu

Response to Referee #2

General Comments

The manuscript presents an interesting and potentially valuable framework for data-driven stochastic simulation of spatial fields. The proposed combination of neural networks and geostatistical concepts is promising. However, several major issues regarding methodological positioning, uncertainty evaluation, interpretation of non-Gaussian modeling capability, and computational comparison need to be addressed. I therefore recommend major revision.

Major Comments

(1) Methodology section: the authors describe ANNSim as a neural-process framework for stochastic simulation. However, the methodological novelty is currently not sufficiently distinguished from existing machine-learning-based spatial models. The proposed framework consists of spatial feature encoding, transformer-based attention mechanism, MLP decoder, and Gaussian conditional sampling. While this architecture is reasonable, it remains unclear what fundamental methodological advance is introduced compared with existing approaches mentioned in the Introduction section. The authors are encouraged to clarify whether ANNSim is intended to learn the underlying spatial stochastic process, and/or the conditional probability distribution, and/or only a flexible approximation of the conditional mean and variance. I recommend adding a dedicated discussion comparing ANNSim with existing approaches. A summary table comparing spatial dependency representation, uncertainty modeling, stochastic realization generation, distributional assumptions, requirement of covariance/variogram modeling, would significantly improve the positioning of this work.

Reply: We appreciate this comment because it led us to change the way the method is defined. The revised manuscript no longer describes ANNSim as a neural-process framework, and we do not present neural-network prediction itself as the novelty. The modelling object is now stated explicitly in Sect. 2: ANNSim learns a local conditional density $q_\theta(z_i \mid C_i)$ from neighbouring values and their spatial relationships. The per-neighbour encoder and attention module represent the conditioning configuration, and the former Gaussian mean/variance output has been replaced by a one-dimensional RQS normalizing flow. The model therefore estimates a full normalized conditional density rather than only a conditional mean and variance. Sect. 2.3 then separates this local conditional model from the sequential construction of a joint distribution along a visiting path. We make no claim that the architecture by itself recovers the unknown true random-field law. The Introduction has also been rewritten to distinguish existing machine-learning approaches for spatial prediction from the conditional stochastic simulation problem addressed here, and Bai and Tahmasebi's SGSim acceleration work is discussed directly. We considered the suggested broad comparison table, but after narrowing the paper we chose to make this distinction in the main text rather than add another taxonomy of spatial learning methods. The revised contribution is the learned conditional simulation framework and its tested behaviour under covariance misspecification and repeated simulation.

(2) The authors provide code availability information, which is appreciated. However, the current manuscript does not provide sufficient details for independent reproduction. Important missing information includes, complete network architecture, number of parameters, initialization strategy, random seed, training epochs, convergence criteria, hyperparameter optimization procedure. Since this is a model-development paper, the above details should be included.

Reply: Thank you for this comment. The revised Sect. 2.2 now gives the neighbour input variables, encoder dimensions, attention construction, conditional-flow form, masking rule and optimization objective. Supplement Sect. S2 provides the fixed architecture and implementation settings. Specific quantities for experiments such as sample sizes, neighbourhood sizes, train/test design, realization counts and timing protocol are given in Sect. 3. We believe that the architecture and main training choices that define the reported model are now explicit, and the code archive remains available for the implementation details of the proposed model ANNSim.

Experimental setup section:

(3) The main contribution of this manuscript is stochastic simulation. However, most evaluation metrics focus on deterministic prediction accuracy, e.g., MAE, RMSE, and R^2. These metrics evaluate reconstruction capability but do not directly evaluate whether generated realizations correctly represent spatial uncertainty. A stochastic simulator should reproduce not only the expected field but also, uncertainty distribution, ensemble spread, conditional variability, and spatial correlation uncertainty. Currently, the manuscript mainly compares histograms, variograms, and single realizations, which is insufficient to determine whether ANNSim produces statistically consistent ensembles. The authors should consider adding ensemble variance maps, or uncertainty coverage analysis, or comparison between simulated and theoretical conditional distributions.

Reply: Thank you for this valuable comment. The evaluation section has been expanded substantially. Sect. 2.4 now separates point accuracy from probabilistic calibration and realization-level diagnostics. We added NLL and CRPS, nominal predictive-interval coverage, exact conditioning checks, empirical distribution comparison, experimental variograms; in addition, Wasserstein distance and short-range spatial-correlation errors are considered in some cases. The controlled Gaussian experiment reports 50% and 90% coverage together with CRPS and NLL, and Figures 4 to 6 evaluate marginal reproduction, variogram behaviour, context masking and visiting-path sensitivity. The real Cu case adds ensemble standard deviation and exceedance-probability maps in Figure 10, together with five-fold cross-validation and observed-versus-simulated distribution and variogram diagnostics. These additions changed the interpretation as well.

(4) In Line 254, the authors mentioned "a random field characterized by heavy tail distribution is constructed to investigate the model performance in non-Gaussian scenarios". However, in the "Methodology" section, the authors also mentioned that the model "estimate the parameters of the underlying conditional Gaussian distribution" (Line 155). Therefore, although the neural network may learn nonlinear spatial representations, the final sampling process remains constrained by Gaussian likelihood assumptions, which creates a conceptual inconsistency between flexible representation learning and explicit non-Gaussian stochastic modeling. The authors should clarify that ANNSim can capture complex spatial dependence in non-Gaussian fields, but the current implementation does not explicitly model arbitrary non-Gaussian conditional distributions. Alternatively, the authors may discuss future extensions using Student-t likelihood, or mixture distributions, or normalizing flows.

Reply: Thank you for pointing out this inconsistency in the original formulation. We addressed it at the model level. The Gaussian mean/variance head has been replaced by a conditional one-dimensional rational-quadratic-spline normalizing flow, described in Sect. 2.2.2, which can model a wide range of conditional density via the invertible conditional transformation. At the same time, we removed the former heavy-tail *t*-field experiment and the broad claim of demonstrated non-Gaussian robustness, because that experiment was tied to the earlier Gaussian-output formulation and was not needed for the revised argument.

Conclusions Section:

(5) Lines 452–453, the authors highlight that "the data-driven method is flexible and shows less sensitivity to the non-Gaussianity and non-stationarity characteristic". However, this statement appears stronger than what is demonstrated by the experiments. The non-stationary experiment introduces deterministic trends into Gaussian random fields, while the non-Gaussian experiment mainly considers heavy-tailed distributions. These cases are useful for initial testing, but they do not fully represent the complexity of real geoscientific fields. For example, real geoscientific fields often exhibit complex and non-stationary spatial structures, such as spatially varying covariance functions, multiple spatial scales, anisotropic correlations, discontinuities, and heterogeneous patterns. Therefore, the current experiments demonstrate that ANNSim performs well for the selected synthetic scenarios, but they do not yet prove general robustness to non-stationarity and non-Gaussianity. The authors should either reduce the strength of these statements or provide additional experiments involving more challenging spatial structures.

Reply: We agree that the original conclusion was too broad. The statements about general resistance to non-Gaussianity and non-stationarity have been removed. The revised robustness test is more specific. Sect. 3.1.2 defines controlled covariance-model misspecification scenarios, and Sect. 4.2 compares the degradation of ANNSim with SGSim supplied with deliberately misspecified covariance models. The main results cover geometric anisotropy, omitted nugget, spatially varying local principal direction, and spatial regimes. Correctly specified SGSim is retained where applicable so that increased process difficulty can be separated from the effect of model misspecification. We also report a zonal-anisotropy case that did not show the same relative degradation pattern (see Supplement Sect. S4 and Fig. S1). The conclusion now says only that ANNSim showed reduced sensitivity for four of the tested misspecification scenarios and explicitly states that this is not a general advantage over covariance-based simulation.

Minor Comments

(1) Line 377, the authors mention that ANNSim generates a wider value range than SGSim. However, it remains unclear whether this represents improved uncertainty representation, better reproduction of extremes, or increased sampling noise. Additional quantitative evaluation is needed.

Reply: Thank you for this comment. The earlier interpretation of a wider simulated range as improved uncertainty representation has been removed. In the revised Figure 5 discussion, the wider local range is described only as a difference in short-scale realization behaviour. Uncertainty is now evaluated quantitatively through CRPS and interval coverage, and the real Cu case explicitly notes that the simulated ensemble is more dispersed than the observations rather than treating that dispersion as an advantage.

(2) Time complexity section, the scalability experiment is limited to relatively small domains. Since computational efficiency is one of the main motivations of this work, additional tests with larger grids would strengthen the manuscript.

Reply: Thank you for this point, which led us to redesign the timing experiment. The revised experiment keeps the 100 × 100 domain fixed and instead measures the quantity that is central to the revised computational claim: when the one-time fitting cost is amortized as neighbourhood size and the number of realizations increase. Sect. 3.1.3 tests eight neighbourhood sizes from 20 to 160 and eight realization counts from 1 to 128, for 64 combinations. ANNSim training is included in total time, paths and Euclidean neighbour plans are paired, and Figure 9 reports both the measured time ratios and the empirical crossover trajectory. Also, note that we only present it as an empirical result for the tested 2D CPU setting.

(3) Although synthetic experiments provide controlled evaluation, at least one real geoscientific application would significantly improve the manuscript. Potential datasets could include soil properties, geological attributes, environmental variables, geochemical observations.

Reply: A real geoscientific case has now been added. Sect. 3.2 introduces stream-sediment Cu data from the Shibanjing 1:200,000 mapsheet in Inner Mongolia, China, and Sect. 4.4 presents the conditional realizations together with cross-validation, ensemble spread, exceedance probability, empirical distributions and variograms. Please refer to the study area map in Figure 3 and the simulation results summarized in Figure 10.

(4) Line 15, “A series of simulation experiments was” should be “A series of simulation experiments are”.

Reply: Thank you for pointing out this tense issue. The abstract was rewritten during the revision, so the original sentence no longer appears there.

(5) Line 33, “generate spatially continuous distribution of the quantity of interest” should be “generate a spatially continuous distribution of the quantity of interest”.

Reply: Corrected. The revised Introduction now reads “generate a spatially continuous distribution of the quantity of interest.”

(6) Line 124, “using, such as, contiguity or distance-based approaches” should be “by using approaches such as contiguity-based or distance-based methods”.

Reply: The original sentence has been removed from the main manuscript to Supplement Sect. S1. And we have revised the expression following your suggestion.

(7) Line 274, “the number of levels of basis functions” should be “the number of basis-function resolution levels”.

Reply: This sentence was removed during the restructuring of the spatial-encoding comparison. The MRBF details and complete benchmark are now confined to Supplement Sect. S1.

(8) Line 303, “where the random field distributed with a heavy tail” should be “where the random field was distributed with a heavy tail”.

Reply: The heavy-tail experiment has been removed from the revised manuscript, so this sentence no longer appears.

(9) Line 307, for clarity, it would be better if the authors change word “variability” to ”“uncertainty.

Reply: The terminology has been revised. Where the text refers to predictive or simulated spread, we now use uncertainty or conditional uncertainty; variability is retained only where it refers to the statistical variability of a field or ensemble.

(10) Figure 8 caption, “(b) and (c) represent the capability of reproducing the histogram and variogram of the observations, respectively for ANNSim; (d) model accuracy for SGSim, (b) and (c) represent”, the last “(b) and (c)” should be “(e) and (f)”.

Reply: The former Fig. 8 has been replaced as part of the revised evaluation. The new figures separate Gaussian fidelity, masking and path effects, covariance misspecification, timing, and the real case, so the duplicated panel references in the original caption no longer occur.

(11) Lines 452–453, “non-Gaussianity and non-stationarity characteristic” should be “non-Gaussianity and non-stationarity characteristics”.

Reply: Thank you for pointing out this issue. The original sentence and the associated broad non-Gaussianity/non-stationarity claim have been removed.

(12) Line 423–424, “more and more location encoding techniques are being emerging” should be “more and more location encoding techniques are emerging”.

Reply: Thank you for pointing out this issue. The sentence was removed when the Implications and limitations section was rewritten.

(13) Line 477, “The authors acknowledged” should be “The authors acknowledge”.

Reply: Corrected.

(14) Line 479, redundant “from the financial support”, please remove.

Reply: Corrected.

(15) The abbreviation SGSim is inconsistently used throughout the manuscript (SGSsim in Line 84). Please standardize the terminology.

Reply: Corrected. We standardized the abbreviation to “SGSim” throughout the revised manuscript.