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 #1

General Comments

This paper proposes a neural-network approach to sequential Gaussian simulation for generating conditional realisations of a 2D spatial field. The experimental section of this work is divided into two parts. First, the authors compare eight ways of encoding spatial dependence as covariates fed into a feedforward network that outputs the mean and variance of a Gaussian conditional distribution, which are then benchmarked against ordinary kriging on synthetic Gaussian fields. Second, they implement an attention-based encoder–decoder and compare its realisations to those from SGSim. While the covariate comparison is a useful reference, I have questions about the novelty and framing of the work. I also note that there no real geoscientific data analysis being carried out.

Major Comments

(1) Clarity-attention mechanism vs. MLP in the method comparison. The methodology section builds up an attentive encoder with a multi-head cross-attention mechanism, but then P7 states "Given that the model's first component inherently learns spatial correlations, we adopted only the second component, a feedforward neural network, to evaluate the performance of various methods." It is unclear to me whether the method comparison in Table 1 uses only an MLP, and not the attention model. Is the attention mechanism only used in the second experimental study? What role does the attention mechanism play in this work?

Reply: Thank you for this point. We acknowledge that the original manuscript did not separate the two parts clearly. In the revision, the comparison of eight spatial representations has been moved to Supplement Sect. S1. All learned representations in that comparison use the same feed-forward prediction head; the attention module is not part of that benchmark. The final ANNSim used in the main experiments is different. For each neighbour, it retains the observed value together with Euclidean distance and signed x and y offsets, and these per-neighbour inputs are encoded and aggregated by multi-head attention before the conditional density is estimated. Thus, the Supplement comparison is not a simplified version of ANNSim and SLF is not used directly as the ANNSim input. Its role is to show that neighbouring attributes are informative, while the location-based representations show that spatial geometry also carries useful information. This distinction is now stated in Sect. 2.1 of the revised manuscript and in Supplement S1.2.

(2) Clarity-connection with neural processes. Another point of clarification is the connection with neural processes. In neural processes, one typically trains using several realisations from the underlying process model; here one is training a network using a single spatial dataset.

Reply: We agree that the original use of the term 'neural process' was somewhat imprecise. The revised manuscript no longer describes ANNSim as a neural-process framework, and the title has been changed accordingly. The method is now defined as a learned conditional stochastic simulator. It does not require a collection of independent realizations from an underlying process model for training. Instead, one observed spatial realization provides many local training examples because each observed target can be paired with a neighbourhood context, and the same conditional operator is shared across locations. This point is stated explicitly in Sect. 2.2. The real Cu case in Sect. 3.2 and Sect. 4.4 follows this setting: the model is trained from one real spatial dataset and does not use SGSim generated realizations as training targets. We retained the useful ideas of per-neighbour encoding and attention, but removed the implication that ANNSim is a standard neural process.

(3) Clarity-fairness of the kriging comparison. Another point of clarification is whether ordinary kriging sometimes does worse because in the other methods "normalization and detrending" are used as "beneficial preprocessing choices." It is unclear whether this is an apples for apples comparison. Moreover, what do we learn from ordinary kriging outperforming the other methods on nearly all metrics?

Reply: We appreciate this valuable comment. The revised paper no longer uses the eight-encoding experiment to argue that a learned method should outperform ordinary kriging. That experiment is now placed in Supplement Sect. S1 and uses standardized Gaussian fields with zero mean and unit variance, so the earlier discussion about normalization or detrending as a possible source of advantage has been removed.

Ordinary kriging is retained as a reference in Table S1, and its strong performance is now treated as an expected result under this well-specified Gaussian setting. What we take from the comparison is more limited: neighbouring-value representations such as NO and SLF perform well, while several geometry-based representations also contain useful spatial information. Those observations motivated the final ANNSim input. They are not used as evidence that ANNSim is superior to kriging.

(4) Clarity-construction of the non-Gaussian process. There also needs to be more details on the non-Gaussian setting; how is a t correlated spatial process being constructed?

Reply: Thank you for this comment. The *t*-distributed field experiment has been removed from the revised manuscript. In the revision, we turned our attention to the experimental design on conditional stochastic simulation and covariance-model misspecification; the former heavy-tail experiment did not directly address these main questions and was therefore removed.

(5) Novelty-positioning against prior work. The paper does not explicitly recognise that there are other works that already use neural networks to give a prediction and associated prediction uncertainty at unobserved locations. The most relevant of these is Wang et al. (2019). Some of the claims are also misleading. For example, the claim that "machine learning algorithms typically treat observations as independent and identically distributed data, neglecting spatial dependence among nearby observations, which might result in misleading predictions and inferences" does not acknowledge that there are several ways to indeed take into account spatial correlation when needed (e.g., Zhan and Datta, 2025). Further, my understanding is that Bai and Tahmasebi (2021) also incorporate neural networks for spatial prediction; it is not clear where the novelty of this work lies.

Reply: We agree with the reviewer that neural-network spatial prediction itself is not a new contribution. The original manuscript blurred this distinction and also made the overly broad statement that machine-learning methods generally neglect spatial dependence. We have removed that statement and rewritten the *Introduction* to acknowledge existing data-driven ways of representing spatial dependence, including coordinate, neighbouring-value, spatial-lag, eigenvector, basis-function and distance-based approaches. In particular, the revised paper now separates spatial prediction from conditional stochastic simulation, which is necessary for quantifying uncertainty propagation (e.g., subsurface system modeling, mineral prospectivity mapping).

The core question we address is whether a local conditional density can be learned directly from observed spatial configurations and then reused sequentially to generate conditional realizations, without requiring a prescribed covariance model or generating training realizations by such as SGSim. We test the fidelity in a Gaussian case, degradation under covariance-model misspecifications, and the conditions under which the fitting cost is amortized. The revised conclusion presents ANNSim as a complementary approach rather than a general replacement for covariance-based simulation.

(6) Theoretical support. The neural network trained in (1) will yield a Gaussian approximation to the (marginal) predictive distribution at any location in the domain. The network, however, is ultimately used in a different way, and conditions on "previously simulated values." It is not obvious to me that if you simulate sequentially in this way from a network trained to give the correct marginal distribution, that the sample jointly comes from the required predictive distribution, a proof is needed.

Reply: Thank you for raising this important theoretical issue, and we have changed both the formulation and the claim. The original text moved too quickly from point prediction to sequential simulation. In Sect. 2.3, the revised manuscript now writes the path-conditioned construction explicitly. For a visiting path, the implemented model has the form $q_\theta(z_U|C_O,\pi) = \prod_{i=1}^{M} q_\theta\left(z_{\pi_i}\middle|C_O, z_{\pi_{<i}}\right)$, where each local context contains selected neighbours from the original observations and values simulated earlier on the same path. Because every factor is a normalized conditional density, this product defines a valid joint probability distribution on the discretized unknown cells for that specified path. We now state clearly what this does not prove: it does not establish that the learned joint distribution is identical to the unknown true conditional random field. That correspondence has to be evaluated empirically rather than inferred from the chain rule alone. For that reason, the revision adds CRPS and interval coverage, exact conditioning checks, marginal and variogram comparisons, and a dedicated visiting-path sensitivity analysis.

Random context masking was also introduced to expose the model during training to changing context configurations, although Figure 6 shows that masking does not significantly improve every diagnostic and that residual path effects remain for some spatial statistics.

(7) No real application. The introduction motivates the work with soil geochemistry and mantle melting. Yet experiments are done on synthetic 2D Gaussian and t-distributed fields. For GMD I would expect at least one real or highly realistic dataset to demonstrate the method does what the motivation promises.

Reply: Thank you for this suggestion. We added a real geoscientific application using stream-sediment Cu data from the Shibanjing 1:200,000 mapsheet in Inner Mongolia, China. The study area and data are described in Sect. 3.2 and Fig. 3. The final model was trained from the observed spatial dataset and used to generate 50 conditional realizations with independently randomized paths. The results are presented in Sect. 4.4 and Figure 10. In addition to cross-validation, the case study reports the ensemble mean, ensemble standard deviation, the probability of exceeding the observed Cu 90th percentile, empirical distribution reproduction, and experimental semivariograms. The case is presented as a practical demonstration of the proposed model.

Other Comments

(8) P4: Xu et al. (2018) is referenced, but does not appear in the bibliography.

Reply: The detailed DLE description has been moved to Supplement Sect. S1, and the full reference for Xu et al. (2018) is now included in the Supplement reference list.

(9) P6 (Line 150): The data $y_i = z(u\alpha)$ but $u\alpha$ is not clearly defined (what is α relative to i?). Also, what is z?

Reply: The notation in this part has been rewritten completely. The revised Sect. 2 now defines the discretized locations, observed and unsampled index sets, visiting path, target value, neighbourhood and local context before the model equations are introduced. The former expression involving u_α and z is no longer used.

(10) P7 (Eq. 1, Lines 189–190): Why is there a 1/N factor in Eq. (1) but not in (2)?

Reply: We agree that the normalization was inconsistent in the original presentation. Those equations have been replaced by the conditional negative log-likelihood formulation in Sect. 2.2.3. The revised text defines the training loss consistently as the average conditional negative log likelihood over the mini-batch, so the earlier difference in the 1/N factor no longer occurs.

(11) P7 (Lines 196–198): "replace the output $\mu(x)$ . . . with $\hat{\mu}(x_i) = \lambda^T x_i = \Sigma_{j=1}^{n}\lambda_j x_{ij}$, then the constraints as defined in kriging can also be incorporated". This is unclear as kriging applies weights to data and not to 'engineered covariates'.

Reply: We agree with this comment. The original sentence mixed kriging weights applied to observations with weights on engineered covariates and was not a clear or necessary part of the method. We have removed that paragraph. The revised loss function is now defined only for the learned conditional density in Sect. 2.2.3.

(12) Table 1: the EDF and EBD encodings give $R^2 \approx 0$ or negative across all n. A sentence explaining why this is happening would be useful.

Reply: The full comparison is now reported in Supplement Table S1 rather than in the main text. We have also added an explanation for the low R^2 values. EBD contains distance information but no neighbouring attribute values, while EDF mainly represents absolute location and broad spatial geometry. For the stationary Gaussian benchmark used here, these inputs alone provide little information about the local field value. Consequently, their prediction errors are close to, or in the case of EBD slightly larger than, those obtained by simply predicting the test-set mean, which results in R^2 values near zero or slightly below zero. The revised paper uses this benchmark only to motivate the combined neighbour-value and relative-geometry input adopted by ANNSim.

(13) Fig. 5b: the fitted variogram is written "$\gamma(h) = 1.05 \cdot \mathrm{Exp}(23.15)$". This notation is ambiguous (is 1.05 the sill, 23.15 the range, and where is h on the RHS?).

Reply: Thank you for pointing this out. The fitted-variogram expression in the former figure has been removed. The revised Gaussian experiment specifies the generating exponential covariance range directly in Sect. 3.1.1, and Figure 4 compares the experimental variograms of the observations and ANNSim realizations without the ambiguous fitted-model notation.

(14) P17 (Line 404): states SGSim scales "approximately with $O(n^2\text{neighbors})$ due to repeated covariance matrix operations", but solving the kriging system at each node would require an nneighbors × nneighbors matrix, which is $O(n^3\text{neighbors})$, not $O(n^2\text{neighbors})$.

Reply: Thank you. The $O\left(n_{\text{neighbors}}^2\right)$ statement in the original manuscript was not a sound description of a dense local kriging solve, and we have removed the asymptotic complexity claim altogether. The computational study was redesigned as an empirical end-to-end timing experiment. Sect. 3.1.3 now includes ANNSim fitting, encoder tracing, common neighbourhood setup and sequential inference, while the SGSim timing includes the same common setup, local covariance construction, Cholesky factorization and solution, and sequential sampling. The benchmark covers eight neighbourhood sizes and eight realization counts, giving 64 combinations, and all comparisons use paired paths and neighbour plans. Figure 9 therefore reports measured crossover behaviour rather than inferring runtime from a simplified complexity expression.

(15) P18 (Conclusions, Line 442): the model is called "an autoencoder architecture inspired by neural processes", but it is not; this is an attentive conditional encoder–decoder.

Reply: We agree with your comment after careful check. The term autoencoder has been removed throughout the revised manuscript. ANNSim is now described directly as a learned conditional stochastic simulator with a per-neighbour encoder, attention aggregation, a conditional RQS flow, and sequential sampling. The revised title also no longer refers to a neural-process framework.

(16) It is never explicitly stated what ANNSim actually is, what the covariates it uses are, etc. This needs to be clearer.

Reply: Thank you, and this part has been rewritten substantially. In the revised manuscript, Sect. 2 starts by defining ANNSim as a learned conditional stochastic simulator and then defines the conditioning context in Sect. 2.1. For each neighbour, the model uses the neighbouring value, Euclidean distance to the target, and signed x and y offsets. These inputs are encoded separately and aggregated by attention in Sect. 2.2.1, then mapped to a conditional density by the RQS flow in Sect. 2.2.2. Sect. 2.3 describes how the learned conditional density is used during sequential simulation. Figure 2 shows the full architecture, and Supplement Table S2 gives the final hyperparameter settings.

(17) Calibration of the predictive variance needs to be shown.

Reply: We agree that uncertainty calibration needed to be evaluated. The revised model now assesses calibration of the full conditional distribution rather than a single variance parameter. Sect. 2.4.1 now includes negative log likelihood, CRPS and empirical predictive-interval coverage. In the controlled Gaussian experiment, the nominal 50% and 90% intervals are reported and discussed. These diagnostics are used alongside marginal-distribution and variogram reproduction.