Review of 'A neural-process framework for stochastic simulation of spatially dependent geoscientific fields' by Wang, Zuo, Huang and Liu

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

- **Clarity.** There are several gaps in the narrative of this work, here I mention a few. 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 mexhanism play in this work? 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. 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? There also needs to be more details on the non-Gaussian setting; how is a t correlated spatial process being constructed?
- **Novelty.** The paper does not explicitly recognise that there are other works that already use neural networks to give a prediction and associated predictoin 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 correlatoin 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.
- **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.
- **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.

Other comments

- P4: Xu et al. (2018) is referenced, but does not appear in the bibliography.
- P6 (Line 150): The data $y_i = z(\mathbf{u}_\alpha)$ but $\mathbf{u}_\alpha$ is not clearly defined (what is α relative to i?). Also, what is z?
- P7 (Eq. 1, Lines 189–190): Why is there a $1/N$ factor in Eq. (1) but not in (2)?
- P7 (Lines 196–198): "*replace the output* $\mu(\mathbf{x})$ *...with* $\hat{\mu}(\mathbf{x}_i) = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}_i = \sum_{j=1}^{n} \lambda_j\, x_{ij}$*, 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' $\mathbf{x}_i$.
- 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.
- 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?).
- P17 (Line 404) states SGSim scales "approximately with $\mathcal{O}(n^2_{\text{neighbors}})$ due to repeated covariance matrix operations", but solving the kriging system at each node would require an $n_{\text{neighbors}} \times n_{\text{neighbors}}$ matrix, which is $\mathcal{O}(n^3_{\text{neighbors}})$, not $\mathcal{O}(n^2_{\text{neighbors}})$.
- 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.
- It is never explicitly stated what ANNSim actually is, what the covariates it uses are, etc. This needs to be clearer.
- Calibration of the predictive variance needs to be shown.

References

Bai, T. and Tahmasebi, P. (2021). Accelerating geostatistical modeling using geostatistics-informed machine learning. *Computers & Geosciences*, 146:104663.

Wang, H., Guan, Y., and Reich, B. (2019). Nearest-neighbor neural networks for geostatistics. In *2019 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 196–205.

Zhan, W. and Datta, A. (2025). Neural networks for geospatial data. *Journal of the American Statistical Association*, 120(549):535–547.