https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.

A neural-process framework for stochastic simulation of spatially dependent geoscientific fields

Jian Wang[1], Renguang Zuo[2], Dazheng Huang[2], Mingang Liu[1]

[1] College of Geography and Planning, Chengdu University of Technology, Chengdu 610059, China
[2] State Key Laboratory of Geological Processes and Mineral Resources, China University of Geosciences, Wuhan 430074, China

Correspondence to: Renguang Zuo (zrguang@cug.edu.cn)

Abstract. Geostatistical simulation (e.g., sequential Gaussian simulation, SGSim) provides an effective framework for quantifying variability of geoscientific variables and supporting risk-informed decision-making in various scenarios. These approaches are theoretically well grounded under assumptions such as stationarity and Gaussianity, and their practical implementation typically involves explicit variogram modeling and repeated neighborhood-based computations, which may become demanding in large-scale or high-dimensional settings. Recently, data-driven modeling strategies have gained increasing attention across scientific disciplines, offering flexible mechanisms for learning spatial dependence structures directly from data. This development motivates the exploration of learning-based alternatives for stochastic simulation. In this paper, artificial neural network-based models were constructed to address the above issues. A series of simulation experiments was generated to test and validate the proposed model. Our results suggest that: (1) spatial dependence can be captured by two complementary strategies, using neighboring attributes (e.g., spatial lag features) and encoding relative positions (e.g., MEM); (2) within our experiments, the proposed data-driven model appears less sensitive to non-Gaussianity and non-stationarity; and (3) the model provides a feasible complement to SGSim by reproducing key statistics (histogram, variogram) with favorable computational cost and flexible model configuration, particularly for large conditioning neighborhoods.

1 Introduction

Reconstructing the distribution of geoscientific variables (e.g., soil geochemistry) is critical for studying the spatial patterns of natural phenomena (Reichstein et al., 2019). However, many geoscientific variables are either directly unobservable (e.g., degree of partial melting in the upper mantle) or can be constrained by only a limited number of observations, such as those obtained from soil profiles or geologic drillholes. Such scarcity of data necessitates the use of spatial predictive modeling to estimate or simulate the distributions (Griffith & Peres-Neto, 2006; Mariethoz et al., 2010; Mole et al., 2014; Szatmári et al., 2021; Talebi et al., 2022).

Numerous studies suggest that geoscientific variables inherently exhibit properties such as spatial autocorrelation, heterogeneity, and covariate correlation, which have been encapsulated in the first, second and third geographical laws, respectively (Tobler, 1970; Goodchild, 2004; Zhu et al., 2018). Based on these properties, a variety of spatial predictive models have been developed to generate spatially continuous distribution of the quantity of interest based on sparse field data and/or conditional covariates (Goovaerts, 1997; Lindgren et al., 2011; Kirkwood et al., 2016, 2022; Lary et al., 2016; Li et al., 2020; Georganos et al., 2021; Garajeh et al., 2021; Wikle & Zammit-Mangion, 2022; Zhao et al., 2023; K. Chen et al., 2024). Kriging is such a typical method for estimating the attribute at unsampled locations by accounting for spatial autocorrelation among observations (Goovaerts, 1997). Another category of predictive modeling approaches applied machine learning (ML) algorithms to capture the (non)linear relationships between covariates and the quantity of interest (Kirkwood et al., 2016; Wilford et al., 2016). However, 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 (Bergen et al., 2019; Meyer, 2019; Reichstein et al., 2019; Vance et al., 2024). A variety of studies have tried to find solutions to incorporate spatial dependence into machine learning framework. For example, Hengl et al. (2018) proposed to use Euclidean buffer distances to address spatial dependence between observations, and empirical examples indicated that buffer distance-based random forest model can produce comparatively accurate and unbiased predictions compared with ordinary kriging. In addition, there are also some other ways to represent spatial dependence, such as spatial coordinates, neighboring values, Euclidean distance fields, spatial lags, Moran eigenvector maps (Dray et al., 2006; Li et al., 2017; Behrens et al., 2018; Ahn et al., 2020; Bai & Tahmasebi, 2021; Liu et al., 2022; Kim et al., 2023). Recently, location encoding methods have been introduced to represent spatial dependence, such as those using basis functions (e.g., radial basis functions, wavelet basis functions) to generate distance-preserved embeddings, thereby enabling distance information to be effectively integrated into machine learning models (Nychka et al., 2015; Mac Aodha et al., 2019; Mai et al., 2022; K. Chen et al., 2024; W. Chen et al., 2024; Su et al., 2024). Without requiring strict stationarity assumptions, repeated variogram modeling, and matrix inversions in kriging systems, these newly developed spatial covariates can offer a flexible complementary framework for spatial modeling (Kopczewska, 2022).

However, there is still a lack of comprehensive comparison of the capability of these methods for representing spatial dependence. In addition, although machine-learning solutions are increasingly explored alongside kriging for spatial estimation, relatively few studies have examined purely spatial dependence-driven stochastic simulation (Bai & Tahmasebi, 2022). Stochastic simulation can avoid smoothing effect inherent in weighted average-based estimators (e.g., kriging), and provide an arbitrary number of equally probable realizations to help quantify spatial uncertainty (Chiles & Delfiner, 2012). However, widely used geostatistical simulation techniques also involve practical constraints in some applications. One major drawback is the reliance on assumptions about the statistical distribution of the data, which may not accurately reflect the complexity and variability of natural phenomena (Mariethoz et al., 2010). Secondly, these methods may become computationally demanding for large and high-dimensional datasets because kriging-system solutions are repeatedly required during simulation (Cressie & Johannesson, 2008; Bai & Tahmasebi, 2021). A variety of methods have been employed to reduce the computational time of geostatistical simulation, such as transforming the spatial correlation structure into a more efficient framework that leverages sparse matrix algorithms (Lindgren et al., 2011), using parallel computing to conduct simultaneous calculations to reduce overall processing time (Tahmasebi et al., 2012), reducing the computational burden by approximating the covariance matrix with a low-rank matrix (Cressie & Johannesson, 2008), and others (e.g., Banerjee et al., 2008). However, few studies attempted to bridge the gap by leveraging machine learning to enhance stochastic simulation based on spatial dependence.

Artificial Neural Networks (ANNs) are a critical branch of machine learning, known for their flexible architectures that can be tailored to various applications (Rumelhart et al., 1986; Hochreiter & Schmidhuber, 1997; LeCun et al., 1998; Krizhevsky et al., 2012; Goodfellow et al., 2014; Raissi et al., 2019; Ho et al., 2020). Another useful property of ANNs is their ability to approximate complex functions, including with a single hidden layer under the universal approximation theorem (Hornik et al., 1989). Their adaptability, scalability, and parallel-processing capability make ANNs a powerful tool for modeling and solving complex problems across diverse fields. Bai & Tahmasebi (2022) proposed to substitute the extensive inverse operations in sequential Gaussian simulation (SGSim) with a physics-informed machine learning (PIML) model, in which the constraints defined by the unbiasedness and minimum variance conditions in ordinary kriging were incorporated into the empirical loss function to serve as physics-informed regularization terms. The model needs to be trained only once based on the set of observations and several realizations generated by SGSim, and can be then consistently used in future simulations. However, this PIML model is not completely independent from SGSim, in that several realizations require to be generated using SGSsim prior to training the model, and that the first several locations for each realization need also to be separately simulated using SGSim. Therefore, the PIML model by Bai & Tahmasebi (2022) still involves variogram modeling and matrix inversion operations for kriging systems; it also requires switching between two frameworks during simulation, which may increase workflow complexity.

In this study, we aim to (1) compare the capability of various methods for capturing spatial dependence, and (2) unearth the potential of establishing a purely artificial neural network-based model for conditional stochastic simulation (ANNSim) of 2D variables by leveraging only spatial dependence. We will use simulation cases to evaluate and benchmark the proposed approach.

2 Methodology

https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.

2.1 Spatial covariates construction

We summarized the following eight ways for encoding spatial dependence:

(1) Direct location encoder (DLE) (Xu et al., 2018). This simple approach directly adopts (normalized) coordinates to represent spatial dependence in spatial modeling (Cracknell and Reading, 2014). Studies show that this method often fails to capture the fine details of data distributions, and hence show limited prediction accuracy (Mai et al., 2022).

(2) Attribute values of neighboring observations (NO) (Sekulić et al., 2020). As described by Tobler's first law of geography, "everything is related to everything else, but near things are more related than distant things", it might be the most natural choice to explicitly include the nearby observations of the query location as spatial covariates $X = \{v_i, i = 1, 2, \cdots, n\}$ (Fig. 1). It is expected that there is a distance threshold, though it might vary with locations, beyond which the observations can make little or even negative contribution to the prediction.

(3) Euclidean buffer distance (EBD) (Hengl et al., 2018). This method uses the Euclidean distances from the query location to neighboring observations as spatial covariates, namely $X = \{d_i, i = 1, 2, \cdots, n\}$(Fig. 1). Studies indicate that the random forest model using such distance-based features as predictors can yield similar predictions to ordinary kriging, without requiring strict assumptions on data distribution type and stationarity. However, it suffers from exponentially growing computational intensity as the calibration data size increases. In addition, the predictive performance reveals high sensitivity to the input data quality (Hengl et al., 2018).

(4) Euclidean distance fields (EDF) (Behrens et al., 2018). This method is inspired by trend surface analysis, which uses the coordinates as explanatory variables and is widely used for detrending (Krumbein, 1959). Euclidean distance fields typically consist of seven independent fields. Five fields are generated by quantifying the distances between each query location and the four corners and the center of the bounding rectangle around the observations, while the other two represents the spatial coordinates along the horizontal and vertical axes, equivalently characterizing distances to the lower and left edges of the bounding rectangle. Behrens et al. (2018) pointed out that this method should result in high prediction accuracy when the spatial autocorrelation is strong.

(5) Moran eigenvector maps (MEM) (Getis & Griffith, 2002; Dray et al., 2006). The Moran eigenvector approach refers to the spatial patterns represented by maps of eigenvectors, which is derived from eigen decomposition of the diagonalized and double-centered spatial weight matrix. The spatial weight matrix is used to quantify the spatial dependence between different spatial units by using, such as, contiguity or distance-based approaches. The incorporation of eigenvectors as covariates can effectively filter out the spatial dependence structure present in the residuals, thereby leading to more accurate and reliable statistical results (Tiefelsdorf & Griffith, 2007).

(6) Spatial lag feature (SLF) (Li et al., 2017). The spatial lag feature is derived from the weighted average of observations within a specified neighborhood of the query location (Fig. 1). Different neighborhood definition can lead to different spatial lag features, enabling the representation of variability at multiple scales (Debarsy & LeSage, 2018). Li et al. (2017) proved that incorporating spatial lags as spatial covariates can significantly improve the predictive performance of spatially-aware models in comparison to non-spatial counterparts.

(7) Rotary position embedding (RPE) (Su et al., 2024). This method leverages rotation matrices to encode absolute positions such that incorporating relative position dependencies. It offers several key properties: flexibility in data size, coherent decaying inter-element dependency with increasing relative distances, and compatibility with linear attention mechanisms. This method can hence provide a distance-preserved representation that ensures nearby locations receive similar embeddings.

(8) Multi-resolution basis function (MRBF) (W. Chen et al., 2024). This method represents spatial dependence by embedding spatial coordinates through multi-resolution basis functions, such as compactly supported Wendland functions. It constructs an enriched feature space that can capture non-linear and non-stationary spatial patterns. Theoretically, it generalizes Kriging by linking to Gaussian processes under infinite network width and offers flexible covariance structures (W. Chen et al., 2024).

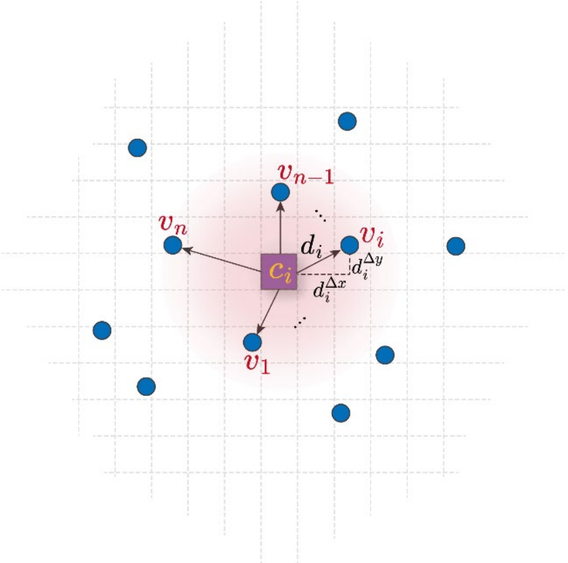


Figure 1: A schematic diagram of constructing spatial covariates. The region of interest has been discretized into a grid $G = \{c_i, i = 1, 2, \cdots, N_G\}$. The attribute for the n nearest neighbors of the unknown cell c_i are denoted v_i, and the corresponding Euclidean distances are d_i.

https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.

The dataset used for evaluating the above methods is constructed in such a way that each of the observations is iteratively taken as the response variable $y_i = z(\boldsymbol{u}_\alpha)$, and the predictors $\mathbf{x}_i = (x_{i1}, x_{i2}, \cdots, x_{in})^{\mathrm{T}}$ for y_i are generated using the presented method above. Therefore, the size of the dataset $D = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ is equal to the sample size N.

2.2 Data-driven stochastic simulation based on spatial dependence

A neural network serves as the core component of the proposed framework. It is designed to map spatial covariates $f: R^n \rightarrow R^2$ and estimate the parameters of the underlying conditional Gaussian distribution. The estimated parameters are then integrated into a sequential simulation framework, which generates multiple equiprobable realizations by successively sampling univariate conditional distributions along a random visiting path (Deutsch and Journel, 1992; Goovaerts, 1997).

2.2.1. Neural network architecture

The aforementioned methods for constructing spatial covariates generally fall into two groups: those based on neighbor attributes (e.g., NO, SLF) and those based on their relative locations (e.g., EBD, RPE). While effective in many settings, these approaches may be less expressive for some complex spatial patterns, such as non-linear dependencies or anisotropy, because they rely on predefined structures. Recent advancements in deep learning, particularly neural processes, offer a powerful and flexible data-driven framework to synergistically learn a comprehensive representation of spatial dependence by adaptively integrating both neighbor attributes and their relative locations.

We borrow the basic idea of neural processes (Garnelo et al., 2018) and structure the model (Fig. 2) into two main components: the first learns a compact representation of spatial dependence from local conditioning data, and the second predicts the conditional distribution at the query location. The model input, consisting of neighbor attributes and their relative locations, synergistically combines the idea of the two categories of spatial covariates to build a comprehensive representation of spatial dependence. The core of the spatial dependence representation component is a transformer architecture, which uses a multi-head attention mechanism to adaptively weigh the influence of each neighbor and capture spatial dependencies. The resulting aggregated spatial context vector is then fed into a final decoder, which is a multi-layer perceptron (MLP). The MLP, consisting of several fully-connected layers and rectified linear unit (ReLU) activation functions, maps the learned representation to parameters of the desired conditional distribution. Skip connections are incorporated to improve gradient flow and preserve information across deep layers. The model outputs the mean and variance of the underlying Gaussian distribution. Throughout the network, ReLU activation functions introduce non-linearity, while dropout layers are employed to mitigate overfitting.

180 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 for capturing spatial dependence.

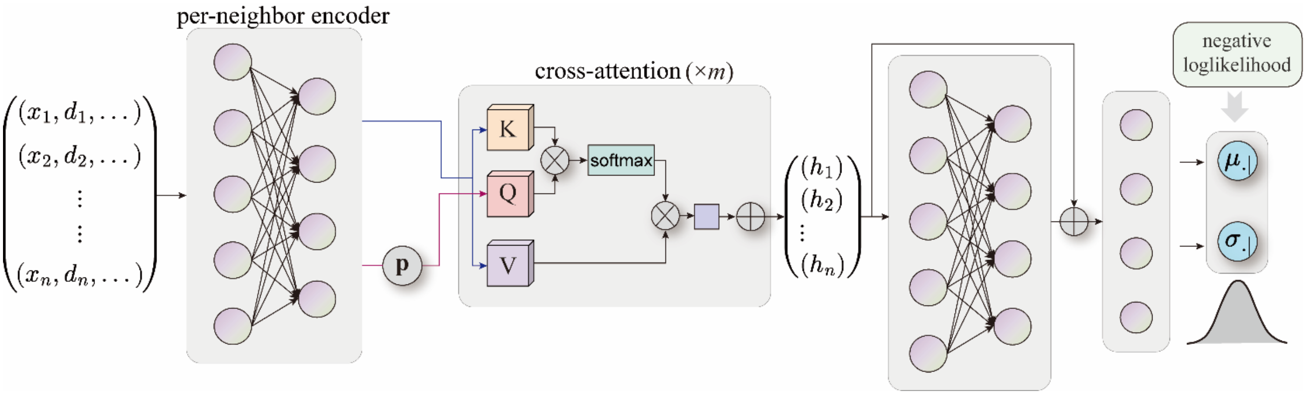


Figure 2: Illustration of the input, output, and the architecture of the proposed model.

185

2.2.2. Loss function

We used the negative log-likelihood to define the loss function for the conditional distribution of the query location (Fig. 2):

$$\mathcal{L}[\boldsymbol{\theta};\boldsymbol{\mathcal{D}}] = -\frac{1}{N}\sum_{i=1}^{N} \log p(y_i = z_i|\mathbf{x}_i;\boldsymbol{\theta}), \tag{1}$$

where $\boldsymbol{\mathcal{D}} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ represents the dataset, $\boldsymbol{\theta} = f(\boldsymbol{w};\boldsymbol{\mathcal{D}})$ represents the parameters of the underlying distribution, $\boldsymbol{w}$ 190 represents the model parameters. In the current implementation, we parameterize the local conditional distribution as Gaussian ($\mathcal{N}(\mu(\mathbf{x}), \sigma^2(\mathbf{x}))$, $\boldsymbol{\theta} = (\mu, \sigma)$), i.e., the network outputs a mean and a variance at each location. This choice is mainly made for likelihood tractability and stable optimization. The negative log-likelihood can then be written as:

$$\mathcal{L}[\boldsymbol{\theta};\boldsymbol{\mathcal{D}}] = \sum_{i=1}^{N}\left[\frac{1}{2}\log\left(2\pi\sigma^2(\mathbf{x}_i)\right) + \frac{\left(y_i - \mu(\mathbf{x}_i)\right)^2}{2\sigma^2(\mathbf{x}_i)}\right], \tag{2}$$

This formulation penalizes the model according to how well the predicted conditional-distribution parameters ($\mu(\mathbf{x})$ and 195 $\sigma^2(\mathbf{x})$) fit the observed data y under the selected output family (Gaussian in this study). Note that this model did not account for the unbiasedness and minimum variance conditions used in ordinary kriging. Nevertheless, if we replace the output $\mu(\mathbf{x})$ with n weights $\lambda_j, j = 1,2,\cdots,n$, and substitute $\mu(\mathbf{x})$ in the loss function (Eq. 2) with $\hat{\mu}(\mathbf{x}_i) = \boldsymbol{\lambda}^{\mathrm{T}}\mathbf{x}_i = \sum_{j=1}^{n}\lambda_j \cdot x_{ij}$, then the constraints as defined in kriging can also be incorporated. The forward pass computes the distribution parameters based on the given input data through linear weighted summation and nonlinear transformation. In the backward pass, the error 200 defined by the negative log-likelihood function is propagated backward through the gradient to update the model parameters.

https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.

In addition, regularization term can be applied to encourage the model to prioritize predictions with spatial correlations to decay with increasing separation distances.

2.2.3. Overall procedure of the data-driven stochastic simulation model

The proposed approach mainly consists of the following steps:

(1) For the current Gaussian-output implementation, transform the data to be approximately normally distributed. Here, normal score transformation is used as a practical preprocessing step to improve optimization stability and parameter calibration in the Gaussian-output setting. For correlated multivariate data, decorrelation, such as using min/max autocorrelation factors (Desbarats & Dimitrakopoulos, 2000), spatial multivariate morphing transformation (Avalos & Ortiz, 2023), should be performed prior to normal transformation. For geochemical data that is not only multivariate but also subjected to closure effect, transformations like flow anamorphosis (Van Den Boogaart et al., 2017) should be employed.

(2) Prepare the spatial covariates and construct the training dataset from the observations. The dataset is then partitioned into training and validation subsets according to a predefined ratio for model fitting and performance evaluation.

(3) Build the model and train it by the training dataset. Regularization schemes, such as early stopping, dropout and batch normalization, can be optionally used to avoid model overfitting.

(4) Generating realizations via the trained model should follow these steps:

a) Define a random path visiting each node of the simulation grid once;

b) For each node, construct the spatial covariates based on the original data and previously simulated values;

c) Obtain the parameters of the underlying distribution, and draw a value from the distribution $\mathcal{N}\left(\mu(\mathbf{x}_i), \sigma^2(\mathbf{x}_i)\right)$ and add it to the dataset;

d) Proceed to the next node along the random path and repeat the steps (b)-(c) until all nodes have been visited.

(5) Transform the simulated realizations back to the original space.

2.3 Model evaluation

2.3.1. Model accuracy

Model accuracy is evaluated by comparing the model output (i.e., the conditional mean) with the observations in the test dataset using mean absolute error (MAE), root mean squared error (RMSE) and the coefficient of determination (R^2). The model with higher R^2 values and lower MAE and RMSE scores indicates higher accuracy in capturing spatial dependence.

2.3.2. Reproduction of lower-order statistics

The ability to reproduce lower-order statistics, including histograms and variograms, was examined to check whether the model effectively captures the variability within the data. The close agreement between the simulated and observed histograms and variograms demonstrates the model's effectiveness in capturing the data variability.

2.3.3. Time complexity

The time complexity of the proposed model was evaluated by measuring the computational time required for stochastic simulation. It should be noted that the trained model can generate new realizations when new data are added; by comparison, geostatistical workflows often require re-estimating variograms and re-running simulation with the updated data.

3 Experimental setup

Two-dimensional random fields were generated to compare the model performance of different methods for encoding spatial dependence, and further explore the potential of the proposed model in generating realizations comparable to SGSim. The size of the simulation domain is set to 100 × 100 with a uniform cell size of 1 × 1, and an omnidirectional exponential covariance model with range 30 and nugget 0 was adopted here. The expectation and marginal variance of the Gaussian field were set to 0 and 1, respectively. A range of randomly sampled points, namely [400, 800, 1200, 1600, 2000], was used to evaluate the effect of sampling density. The constructed dataset was split into a training and testing subset at a ratio of 8:2. In addition, to explore the applicability of the model for non-stationary scenario, a deterministic trend generated by a nonlinear function of the coordinates (e.g., Ba et al., 2012): $t(\boldsymbol{u}) = \sin\{30(\bar{u}-0.9)^4\}\cos\{2(\bar{u}-0.9)\} + (\bar{u}-0.9)/2$, where $\boldsymbol{u} = \left(u_x, u_y\right)^{\mathrm{T}} \in \mathbb{R}^{\mathbf{2}}$ and $\bar{u} = \left(u_x + u_y\right)/2$, was added to the Gaussian field. Moreover, a random field characterized by heavy-tail distribution is constructed to investigate the model performance in non-Gaussian scenarios.

The transformer-based encoder is configured with 2 dense layers and a 2-head attention mechanism. The model uses an embedding dimension of 16 and a model dimension of 32, followed by a multi-layer perceptron decoder consisting of fully-connected layers for conditional mean and log-variance prediction. For training, the model uses the Adam optimization algorithm with a learning rate of 0.0002 and a batch size of 32. For the method of rotary position embedding, the dimension embedded in the feature space is set to 128.

All the model components are coded via Python (Wang et al., 2026). The computational platform is a personal laptop with an intel i5-1260P 2.10 GHz central processing unit, and RAM 16G.

4 Results and discussions

4.1 Model performance for encoding spatial dependence

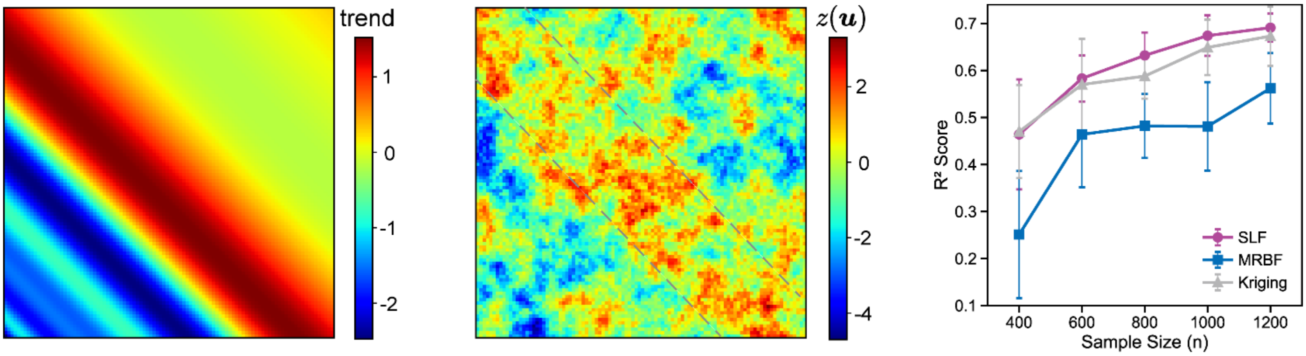
The model performance on the test dataset (Table 1) suggest that the performance varies with the method used to construct spatial covariates, and the uncertainty and variation patterns also differ with increasing sampling density. The methods DLE, NO, MEM, SLF, RPE and MRBF generally captured spatial dependence in our experiments, as indicated by improved model performance with increasing sampling density. Moreover, the evaluation metrics exhibit largely consistent variation patterns across these methods. The method of RPE is sensitive to the sampling density, provided that a dramatic performance gain was observed from the case $n = 400$ to $n = 2000$. For MRBF, the model performance increases from $n = 400$ to $n = 1600$, but declines at $n = 2000$, suggesting that there might be a trade-off between the complexity of the underlying field and the number of levels of basis functions required to represent spatial dependence. Generally, a higher performance can be obtained by encoding the spatial coordinates into a higher dimensional feature space, as evidenced by the performance metrics of the methods MEM and MRBF compared with DLE. The model performance using neighboring observations and spatial lag features as spatial covariates are generally high, indicating they are favorable in capturing spatial dependence under the Gaussian assumption. By contrast, because the spatial lag feature is constructed from neighboring observations, it may provide a relatively stable representation of spatial dependence compared with NO. Among all methods, SLF achieved the highest performance scores for $n = 400$ and ranked second in other cases, indicating performance that is close to ordinary kriging under the Gaussian setting. The method of EBD used alone cannot well capture the spatial dependence, as suggested by the general poor performance for the wide range of sampling densities. It might be due to the limited number of neighboring observations accounted for by the current model (Hengl et al., 2018). For the EDF method, model performance improves slightly with increasing sample size, indicating a limited capacity to capture spatial dependence.

Table 1 Model performance metrics for different methods representing spatial dependence.

		n=400	n=800	n=1200	n=1600	n=2000
DLE	MAE	0.647 ± 0.074	0.588 ± 0.043	0.539 ± 0.030	0.525 ± 0.024	0.511 ± 0.022
	RMSE	0.804 ± 0.093	0.738 ± 0.053	0.675 ± 0.036	0.658 ± 0.028	0.642 ± 0.025
	R^2	0.325 ± 0.122	0.454 ± 0.071	0.545 ± 0.058	0.564 ± 0.037	0.581 ± 0.036
NO	MAE	0.493 ± 0.048	0.431 ± 0.028	0.382 ± 0.015	0.353 ± 0.015	0.336 ± 0.011
	RMSE	0.624 ± 0.053	0.547 ± 0.033	0.485 ± 0.019	0.446 ± 0.017	0.428 ± 0.015
	R^2	0.591 ± 0.076	0.700 ± 0.037	0.765 ± 0.028	0.800 ± 0.016	0.814 ± 0.014
EBD	MAE	0.790 ± 0.053	0.803 ± 0.051	0.802 ± 0.048	0.803 ± 0.035	0.799 ± 0.021
	RMSE	0.988 ± 0.072	1.006 ± 0.051	1.005 ± 0.054	1.000 ± 0.033	0.996 ± 0.027
	R^2	-0.015 ± 0.075	-0.008 ± 0.045	0.000 ± 0.033	-0.004 ± 0.032	-0.005 ± 0.027
EDF	MAE	0.778 ± 0.060	0.776 ± 0.058	0.778 ± 0.047	0.774 ± 0.045	0.766 ± 0.027
	RMSE	0.967 ± 0.081	0.971 ± 0.065	0.972 ± 0.054	0.964 ± 0.044	0.957 ± 0.035
	R^2	0.030 ± 0.065	0.060 ± 0.068	0.064 ± 0.044	0.066 ± 0.043	0.070 ± 0.050
MEM	MAE	0.596 ± 0.072	0.512 ± 0.035	0.481 ± 0.027	0.465 ± 0.028	0.453 ± 0.021
	RMSE	0.773 ± 0.090	0.676 ± 0.051	0.646 ± 0.046	0.626 ± 0.043	0.619 ± 0.038
	R^2	0.370 ± 0.151	0.540 ± 0.070	0.584 ± 0.056	0.604 ± 0.054	0.611 ± 0.044
SLF	MAE	***0.465*** ± 0.049	**0.404** ± 0.029	**0.358** ± 0.018	**0.328** ± 0.014	**0.308** ± 0.014
	RMSE	***0.591*** ± 0.054	**0.520** ± 0.032	**0.464** ± 0.022	**0.428** ± 0.014	**0.407** ± 0.016
	R^2	***0.634*** ± 0.056	**0.730** ± 0.032	**0.785** ± 0.027	**0.816** ± 0.014	**0.832** ± 0.016
RPE	MAE	0.839 ± 0.079	0.708 ± 0.043	0.616 ± 0.044	0.547 ± 0.026	0.505 ± 0.022
	RMSE	1.073 ± 0.095	0.921 ± 0.054	0.815 ± 0.059	0.733 ± 0.035	0.673 ± 0.039
	R^2	-0.202 ± 0.169	0.150 ± 0.101	0.340 ± 0.080	0.457 ± 0.056	0.540 ± 0.046
MRBF	MAE	0.571 ± 0.045	0.501 ± 0.038	0.447 ± 0.022	0.408 ± 0.018	0.413 ± 0.019
	RMSE	0.718 ± 0.055	0.652 ± 0.055	0.587 ± 0.035	0.533 ± 0.026	0.552 ± 0.028
	R^2	0.459 ± 0.090	0.573 ± 0.069	0.657 ± 0.046	0.713 ± 0.033	0.691 ± 0.029
OK	MAE	**0.466** ± 0.040	***0.386*** ± 0.027	***0.337*** ± 0.013	***0.314*** ± 0.014	***0.297*** ± 0.011
	RMSE	**0.595** ± 0.048	***0.497*** ± 0.031	***0.437*** ± 0.016	***0.406*** ± 0.016	***0.387*** ± 0.013
	R^2	**0.627** ± 0.071	***0.752*** ± 0.033	***0.809*** ± 0.023	***0.834*** ± 0.016	***0.847*** ± 0.014

Note: 30 experiments were performed to derive robust metrics.

Geostatistical theory argues that if the target and neighboring positions are not jointly Gaussian, the conditional mean under Gaussian assumptions remains the best linear unbiased prediction; however, nonlinear predictors may perform better in some settings. We firstly evaluated the model performance for the scenario where the underlying spatial process is non-stationary. Only the methods SLF, MRBF and Kriging are investigated here, among which the first one represents the spatial attribute covariates and the second one represents the spatial distance covariates (Fig. 3). It can be observed that for non-stationary field, the model performance also increases with the sampling density since more observations available in the neighborhood can help capture hidden patterns of the data. The reason for the performance of the method MRBF staying at a plateau when the sampling density ranges from n =600 to n =1000 consists in that it uses the same number of levels (hence basis functions) in this size range, which is determined by $H = 1 + \lceil \log_2(\sqrt{n}/10) \rceil$ (Nychka et al., 2015). The model performance decreased relative to the case where the field is stationary. More importantly, baseline kriging showed a larger performance decline than the other tested methods, suggesting that data-driven methods may offer complementary robustness when data depart from kriging assumptions. The model performance for the methods SLF, MRBF and kriging is also compared in the situation where the random field distributed with a heavy tail (such as a multifractal distribution), a very common situation in earth sciences. A field with t-distribution with a degree of freedom being 1.0 was used in this study (Fig. 4). It can be observed that model performance declined and did not increase with sampling density as in the previous setting. This suggests that spatial dependence cannot be captured solely by obtaining more observations in the local neighborhood, if the local variability is high. The variability of the model performance metrics is especially high for kriging. By contrast, SLF showed slightly better and more stable performance in this setting; however, kriging may still outperform SLF in some cases because metric variability is relatively large.

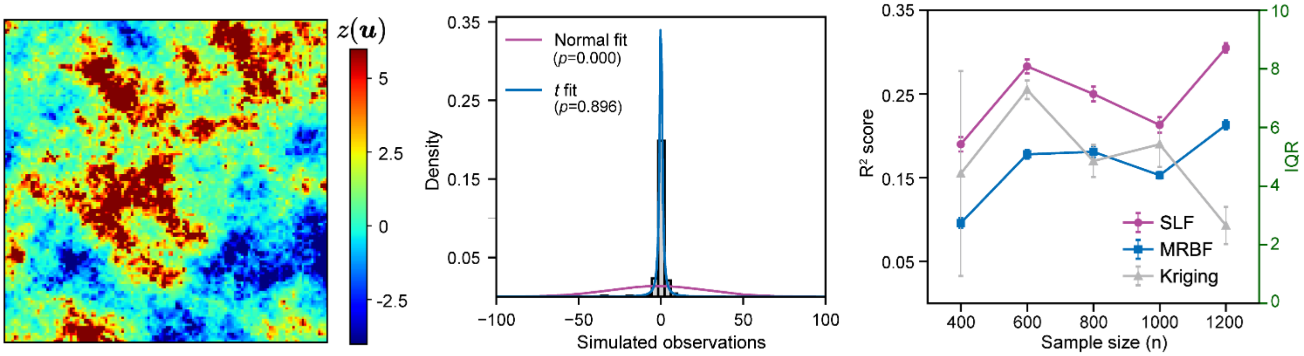


(a) (b) (c)

Figure 3: The model performance against sampling density for non-stationary field. (a) The trend term, (b) the non-stationary random field generated by adding the trend term to the Gaussian field, and (c) model performance (R2 score) for the methods SLF, MRBF and Kriging.

https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.

These results suggest that the data-driven model may be less sensitive to non-stationarity and non-Gaussianity than kriging in the tested scenarios. However, the observed improvement on approximately Gaussian data indicates that normalization and detrending are beneficial preprocessing choices for the current formulation (note that it is not mandatory assumptions of the overall framework). The above results reveal two distinct, yet effective, strategies for spatial dependence modeling. The first leverages the attributes of neighboring data, in which SLF shows strong performance. The second, exemplified by MEM and MRBF, demonstrates the utility of encoding spatial coordinates into a higher-dimensional feature space. This latter finding is particularly salient, as it underscores the power of projecting location information into a rich embedding space, a core mechanism utilized by advanced generative frameworks, such as neural processes, to synergistically learn from both spatial context and feature attributes.



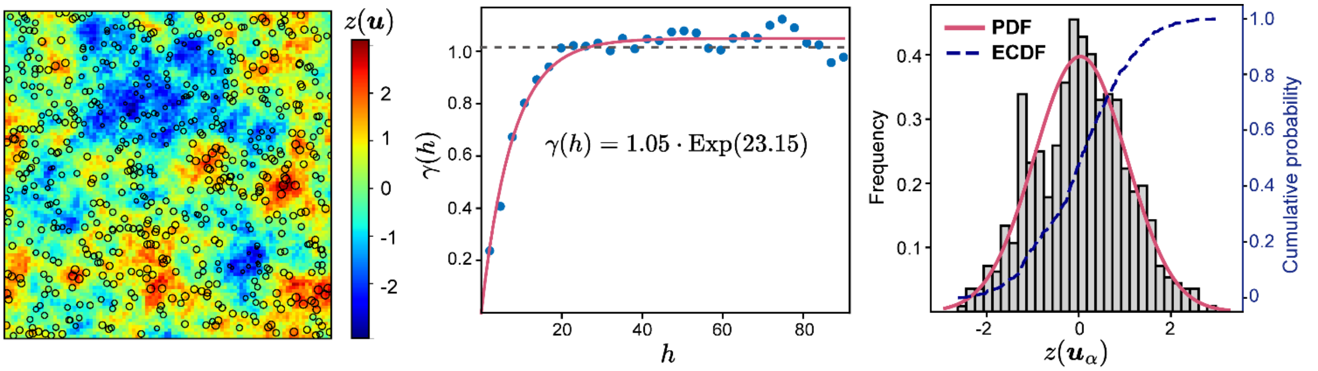
(a) (b) (c)

Figure 4: The model performance against sampling density for field with heavy-tail t-distribution. (a) The simulated random field, (b) histogram of the simulated observations and fit of normal and t distributions, and (c) model performance (R2 score) boxplots for the methods SLF, MRBF and Kriging. Note that the random field is displayed with an interquartile range (IQR)-based scale (data limits are determined by $\mathbf{Q_{0.25} - 1.5 \times IQR}$ and $\mathbf{Q_{0.75} - 1.5 \times IQR}$, respectively) to accommodate the extremely-large data limits of the generated heavy-tailed distribution. The Kolmogorov–Smirnov test statistics for normal and t distributions are also calculated and annotated in the figure.

4.2 Evaluation of the data-driven stochastic simulation method

A total number of 600 observations randomly drawn from a 100 × 100 Gaussian field generated by an exponential model was simulated (Fig. 5a). Experimental variograms were calculated and an omnidirectional exponential model was fitted based on the least square method (Fig. 5b). The nugget, sill and effective range for the fitted exponential model is 0, 1.05 and 23.15, respectively. The histogram of the simulated observations (Fig. 5c) indicates that it follows normal distribution (Kolmogorov-Smirnov Test: $p = 0.459 > 0.05$). The observations were then used to construct the dataset for calibrating and evaluating the proposed model. The negative log-likelihood loss for the model converged successfully for both the training and validation datasets (Fig. 6).

https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.



(a) (b) (c)

Figure 5: The simulated two-dimensional Gaussian field and observations. (a) The simulated Gaussian field and the spatial distribution of the observations, (b) and (c) are the variogram and histogram of the observations, respectively. The dashed line in (b) represents the marginal variance of the observations, and the dashed line in (c) represents empirical cumulative distribution function. Please refer to the text for more information.

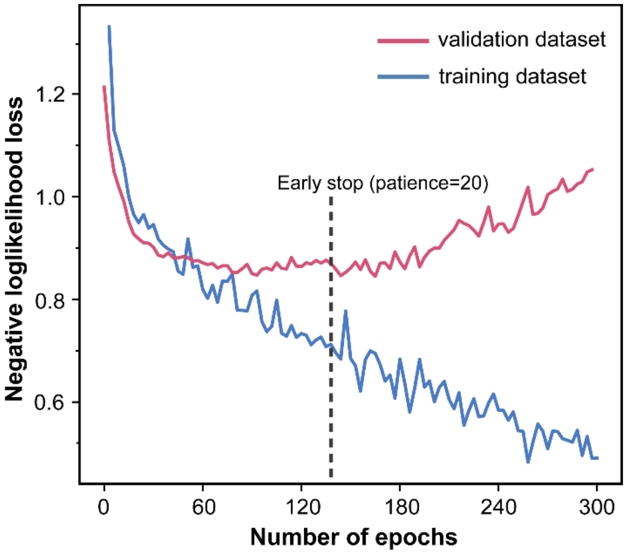
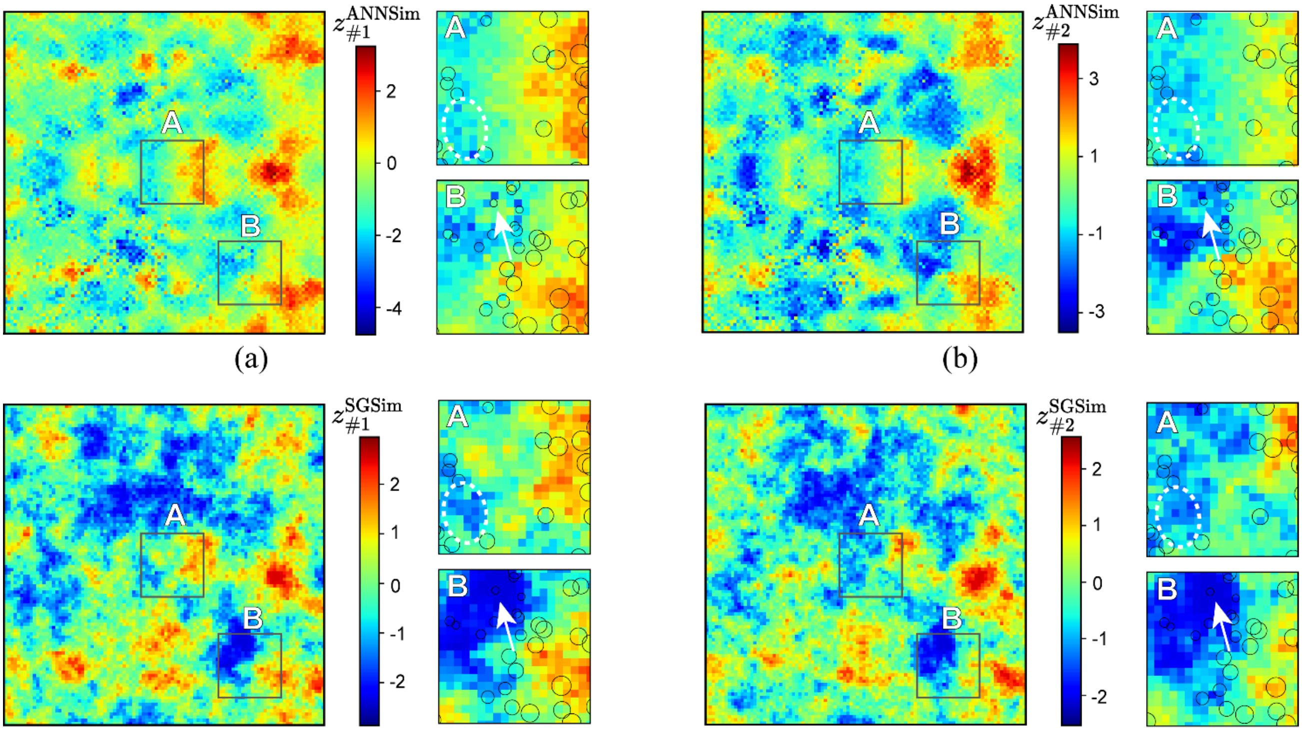


Figure 6: The training process of the ANNSim model. We used 20% of the dataset, namely 120 examples, as the validation dataset. The parameter 'patience' represents the number of consecutive epochs with no significant performance improvement, after which training will be forced to stop.

A total of 100 realizations were generated by the trained ANNSim model (Fig. 7). The same number of realizations were produced by SGSim for comparison. It can be observed that the large-scale variation patterns were successfully reproduced by the data-driven stochastic simulation model, given that there are approximately 10, 000 grid cells to simulate while only several hundreds of conditional data entries are available. The difference in local details between the realizations from the two types of methods is evident, as suggested by the enlarged subfigure, which depicts the spatial uncertainty at unsampled

© Author(s) 2026. CC BY 4.0 License.

locations. Compared with SGSim, the realizations generated by ANNSim shows slightly weak continuous patterns at small scales, as evidenced by the local region ‘A’ enclosed by the white dashed ellipse. In addition, the simulated values surrounding the conditional point, as pointed by the white arrow in the local region ‘B’, are consistently low for the realizations generated by SGSim; by contrast, there are some relatively high values in the simulated results of ANNSim. The presence of screening effect in solving kriging systems, the phenomenon that optimal linear prediction of a spatial process on an unobserved location mostly depends on nearby observations, forces the distribution of surrounding locations to be significantly affected by the conditional low data values in SGSim. In contrast, ANNSim may retain more long-range correlation information, which can lead to larger variability in local simulated patterns.

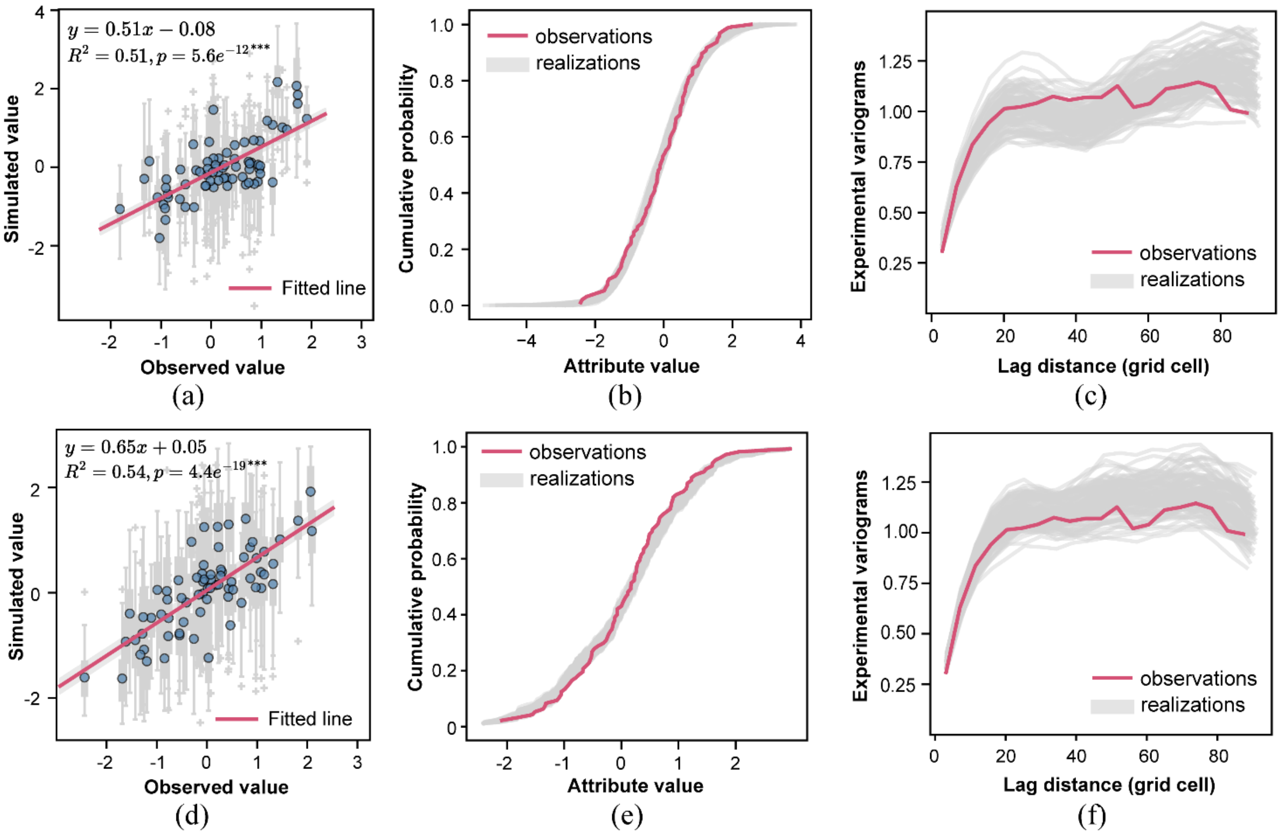


(a) (b)

(c) (d)

Figure 7: Visual comparison of the realizations produced by ANNSim and SGSim. (a) and (b) are two realizations generated by ANNSim, (c) and (d) are two realizations generated by SGSim. Note that the realizations were randomly selected from the ensemble of realizations. Two local regions, denoted by “A” and “B”, were selected and enlarged for ease of comparison of simulation details. Note that the circle symbols in the enlarged figures represent conditional training data.

The model accuracy, evaluated by comparing the mean of realizations with random hold-out references in the test dataset, showed that both ANNSim and SGSim provide effective estimates of unknown values, with SGSim giving slightly lower error for conditional-mean prediction in this experiment (Fig. 8a and 8d). Both ANNSim and SGSim reproduce the observed histograms well. However, ANNSim generates a noticeably wider data range than SGSim (Figs. 7 and 8b), indicating greater variability, particularly in the tails. This enhanced dispersion likely stems from the nonlinear mapping inherent in neural networks, while the overall distributional pattern remains preserved. The variograms (Fig. 8c) show that both methods capture the spatial structure of the observations on average. Nevertheless, ANNSim realizations display slightly stronger ergodic fluctuations than those from SGSim.



(a) (b) (c)

(d) (e) (f)

Figure 8: Evaluation of the realizations produced by ANNSim and SGSim. (a) Model accuracy for ANNSim, (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 capability of reproducing the histogram and variogram of the observations, respectively for SGSim. The reference variogram model and empirical cumulative distribution for the observations are also plotted for comparison.

4.3 Time complexity

To assess computational efficiency, both ANNSim and SGSim were benchmarked under two experimental settings. The first experiment varied the number of spatial neighbors involved in local conditioning, while the second altered the number of realizations generated. All tests were performed on Gaussian random fields with identical grid dimensions (100 × 100), and an exponential covariance structure. For ANNSim, the number of hidden dimensions, attention heads, and latent

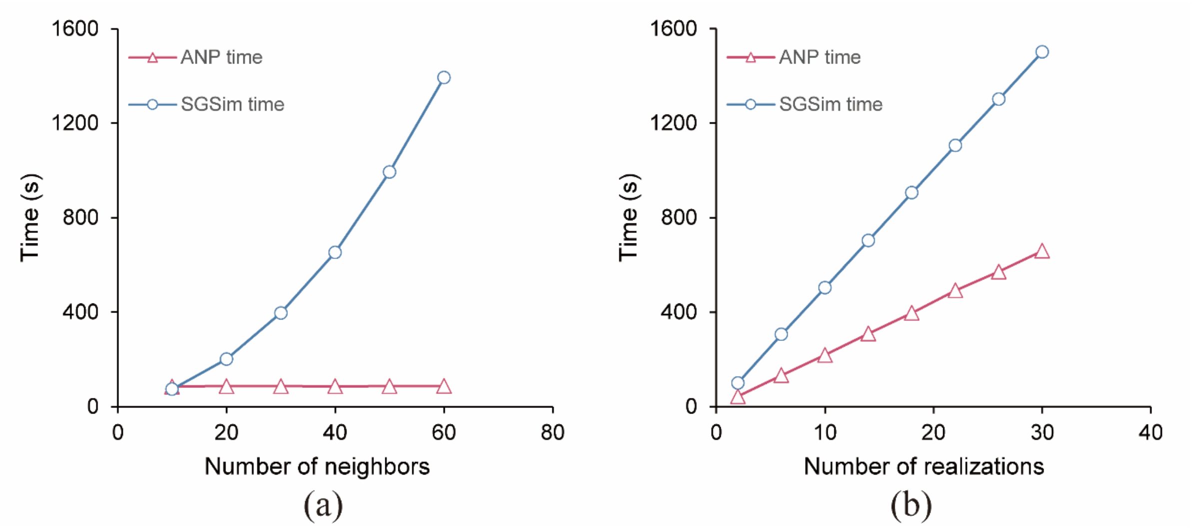
© Author(s) 2026. CC BY 4.0 License.

representation layers were kept constant, ensuring that timing differences arose solely from neighborhood size and realization number.

In the first experiment, the number of neighbors was systematically increased from 10 to 60 while keeping the number of realizations fixed. The results show that the computational time of ANNSim remained nearly constant across different neighborhood sizes, suggesting that the attention-based encoding efficiently handles neighbor information without explicit pairwise distance recalculations during simulation. In contrast, SGSim showed a faster and nonlinear increase in computational cost as the number of neighbors increased under this benchmark. This behavior stems from its sequential conditioning scheme, in which each simulated node requires kriging-based estimation using all selected neighbors, making its complexity scale approximately with $\mathcal{O}\left(n_{neighbors}^2\right)$ due to repeated covariance matrix operations.

In the second experiment, the number of realizations increased from 2 to 30 while the neighborhood size was fixed. Both models showed nearly linear growth in computational time with respect to the number of realizations, consistent with the expectation that each realization represents an independent simulation pass. Nevertheless, ANNSim maintained lower absolute computation time in this benchmark, indicating potential scalability benefits for larger simulation tasks.

Also, note that the training time required by ANNSim is related to the model architecture and hyperparameters, where more model parameters involved will generally induce more training time. However, in our setup, training time was typically lower than simulation time and had limited impact on total runtime.



(a) (b)

Figure 9: The relation of time complexity with the neighborhood size for SGSim and ANNSim (a), and the computational time for ANNSim when different number of realizations are generated (b).

© Author(s) 2026. CC BY 4.0 License.

4.4 Implications and limitations

Considering that a geostatistical observation can be formulated as a generalized additive model $Z(\boldsymbol{u}) = \boldsymbol{x}(\boldsymbol{u})^T\boldsymbol{\beta} + s(\boldsymbol{u}) + \varepsilon(\boldsymbol{u})$, where $s(\boldsymbol{u})$ denotes a spatially dependent and zero-mean random process, this study provided implications about how to incorporate spatial dependence into data-driven modeling using artificial neural networks. However, it is still an open topic on what is the optimal and generalizable way to represent spatial dependence in data-driven models, given that more and more location encoding techniques are being emerging in recent years.

The current study suggests that the proposed data-driven framework can be considered as a complementary option to geostatistical approaches for spatial prediction or simulation. Although it did not outperform geostatistical methods in all cases, the current data-driven formulation provides an additional framework that can be less sensitive to stationarity and distributional misspecification in the tested scenarios, while still using a Gaussian conditional output head. In addition, it is scalable in that the model input can be easily extended to incorporate secondary covariates, and the model output can be adjusted to model different types of distributions. The loss function can be also tailored to account for different effects of the spatial model. For heavy-tailed distributions that are hard to normalize, future extensions may replace the current Gaussian output head (mean/variance) with alternative likelihoods (e.g., Student-*t*, mixture density, or quantile-based objectives), together with corresponding customized calibration losses. More importantly, the model can provide an estimation of the non-stationary covariance structure, whose parameters can serve as useful criteria for understanding the underlying spatial process, in addition to benefit spatial modeling.

Nevertheless, due to the data-driven nature of the proposed model, it may show poor performance when only a small number of observations are available while the underlying spatial patterns are complicated. In particular, it is impracticable to use such data-driven model for unconditional simulation.

5 Conclusions

This study presents a data-driven framework (ANNSim) for stochastic geostatistical simulation. By leveraging an autoencoder architecture inspired by neural processes, ANNSim is designed to implicitly learn complex spatial dependence structures directly from neighboring data. The findings offer several significant contributions to the field:

Firstly, the comparative analysis of spatial dependence representations yielded a key insight: effective spatial modeling benefits from two distinct strategies. The strong performance of attribute-based methods (e.g., SLF) confirms the importance of neighboring values. Simultaneously, the success of coordinate-based methods (e.g., MEM, MRBF) underscores the

© Author(s) 2026. CC BY 4.0 License.

critical power of location embedding. This dual success validates ANNSim's architectural approach, which synergistically integrates both neighbor attributes and their relative locations to build a comprehensive spatial representation.

Secondly, the data-driven method is flexible and shows less sensitivity to the non-Gaussianity and non-stationarity characteristic of complex geological datasets. Compared with workflows such as SGSim that are typically implemented under Gaussian and stationarity assumptions, ANNSim learns spatial representations directly from data; in this study, however, the final conditional sampler is parameterized with a Gaussian output head. This representation-learning design may better accommodate complex local statistical variations without imposing a single global stationarity structure.

Thirdly, ANNSim appears to be a computationally efficient option that can complement SGSim. The model can successfully reproduce the target marginal distribution and spatial continuity. The time-complexity analysis indicated that, in our implementation, ANNSim runtime was nearly constant with respect to neighbor count, whereas SGSim increased approximately polynomially. This makes ANNSim exceptionally scalable for problems requiring large neighborhoods to capture long-range dependencies. While both methods scaled linearly with the number of realizations, ANNSim showed lower CPU time per realization in our experiments, supporting its practical use.

Code and data availability

The current version of the code for ANNSim model is archived on Zenodo under: https://doi.org/10.5281/zenodo.18834446 (Wang et al., 2026). The code is released under the MIT licence. This study is based on synthetic data generated by simulation, which can be accessed in the same repository as code. No third-party or proprietary datasets were used.

Author contributions

J. Wang: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, validation, and writing of the original draft. R. Zuo: review and supervision. D. Huang: review, methodology and visualization. M. Liu: review, code development and visualization.

Competing interests

The authors declare that they have no competing interests.

Disclaimer

The standard publisher’s disclaimer will be applied by Copernicus Publications during the production stage.

Acknowledgements

The authors acknowledged the editor and reviewers' constructive comments that help improve the study.

Financial support

This research benefited from the financial support from the financial support from the "DeepEarth Probe and Mineral Resources Exploration - National Science and Technology Major Project" (Project No. 2025ZD1007703) and the National Natural Science Foundation of China (No. 42572390).

References

Ahn, S., Ryu, D.-W., Lee, S., 2020. A machine learning-based approach for spatial estimation using the spatial features of coordinate information. ISPRS International Journal of Geo-Information, 9(10), 587.

Avalos, S., Ortiz, J. M., 2023. Spatial multivariate morphing transformation. Mathematical Geosciences, 55(6), 735–771.

Ba, S. Joseph, V. R., 2012. Composite gaussian process models for emulating expensive functions. The Annals of Applied Statistics, 6(4):1838–1860.

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

Bai, T., Tahmasebi, P., 2022. Sequential gaussian simulation for geosystems modeling: A machine learning approach. Geoscience Frontiers, 13(1), 101258.

Banerjee, S., Gelfand, A. E., Finley, A. O., Sang, H., 2008. Gaussian predictive process models for large spatial data sets. Journal of the Royal Statistical Society Series B: Statistical Methodology, 70(4), 825–848.

Behrens, T., Schmidt, K., Viscarra Rossel, R. A., Gries, P., Scholten, T., MacMillan, R. A., 2018. Spatial modelling with euclidean distance fields and machine learning. European Journal of Soil Science, 69(5), 757–770.

Bergen, K. J., Johnson, P. A., De Hoop, M. V., Beroza, G. C., 2019. Machine learning for data-driven discovery in solid earth geoscience. Science, 363(6433), eaau0323.

Chen, K., Liu, E., Deng, M., Tan, X., Wang, J., Shi, Y., Wang, Z., 2024. DKNN: Deep kriging neural network for interpretable geospatial interpolation. International Journal of Geographical Information Science, 38(8), 1486–1530.

Chen, W., Li, Y., Reich, B. J., Sun, Y., 2024. DeepKriging: Spatially dependent deep neural networks for spatial prediction. Statistica Sinica. 34(1), 291-311.

Chiles, J.-P., Delfiner, P., 2012. Geostatistics: Modeling spatial uncertainty (Vol. 713). John Wiley & Sons.

Cressie, N., Johannesson, G., 2008. Fixed rank kriging for very large spatial data sets. Journal of the Royal Statistical Society Series B: Statistical Methodology, 70(1), 209–226.

https://doi.org/10.5194/egusphere-2026-1126
Preprint. Discussion started: 17 June 2026
© Author(s) 2026. CC BY 4.0 License.

Debarsy, N., LeSage, J., 2018. Flexible dependence modeling using convex combinations of different types of connectivity structures. Regional Science and Urban Economics, 69, 48–68.

Desbarats, A., Dimitrakopoulos, R., 2000. Geostatistical simulation of regionalized pore-size distributions using min/max autocorrelation factors. Mathematical Geology, 32, 919–942.

Deutsch, C. V., Journel, A. G., 1992. Geostatistical software library and user's guide. New York, 119(147), 578.

Dray, S., Legendre, P., Peres-Neto, P. R., 2006. Spatial modelling: A comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecological Modelling, 196(3–4), 483–493.

Garajeh, M. K., Malakyar, F., Weng, Q., Feizizadeh, B., Blaschke, T., Lakes, T., 2021. An automated deep learning convolutional neural network algorithm applied for soil salinity distribution mapping in lake urmia, iran. Science of the Total Environment, 778, 146253.

Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende, D. J., Eslami, S. M. A., Teh, Y. W., 2018. Neural processes (No. arXiv:1807.01622). arXiv.

Georganos, S., Grippa, T., Niang Gadiaga, A., Linard, C., Lennert, M., Vanhuysse, S., Mboga, N., Wolff, E., Kalogirou, S., 2021. Geographical random forests: A spatial extension of the random forest algorithm to address spatial heterogeneity in remote sensing and population modelling. Geocarto International, 36(2), 121–136.

Getis, A., Griffith, D. A., 2002. Comparative spatial filtering in regression analysis. Geographical Analysis, 34(2), 130–140.

Goodchild, M. F., 2004. The validity and usefulness of laws in geographic information science and geography. Annals of the Association of American Geographers, 94(2), 300–303.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Advances in Neural Information Processing Systems, 27.

Goovaerts, P., 1997. Geostatistics for natural resources evaluation. Oxford University Press, USA.

Griffith, D. A., Peres-Neto, P. R., 2006. Spatial modeling in ecology: The flexibility of eigenfunction spatial analyses. Ecology, 87(10), 2603–2613.

Hengl, T., Nussbaum, M., Wright, M. N., Heuvelink, G. B. M., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. PeerJ, 6, e5518.

Ho, J., Jain, A., Abbeel, P., 2020. Denoising diffusion probabilistic models. Advances in Neural Information Processing Systems, 33, 6840–6851.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Computation, 9(8), 1735–1780.

Hornik, K., Stinchcombe, M., White, H., 1989. Multilayer feedforward networks are universal approximators. Neural Networks, 2(5), 359–366.

Kim, H.-J., Mawuntu, K. B. A., Park, T.-W., Kim, H.-S., Park, J.-Y., Jeong, Y.-S., 2023. Spatial autocorrelation incorporated machine learning model for geotechnical subsurface modeling. Applied Sciences, 13(7), 4497.

Kirkwood, C., Cave, M., Beamish, D., Grebby, S., Ferreira, A., 2016. A machine learning approach to geochemical mapping. Journal of Geochemical Exploration, 167, 49–61.

Kirkwood, C., Economou, T., Pugeault, N., Odbert, H., 2022. Bayesian deep learning for spatial interpolation in the presence of auxiliary information. Mathematical Geosciences, 54(3), 507–531.

Kopczewska, K., 2022. Spatial machine learning: New opportunities for regional science. The Annals of Regional Science, 68(3), 713–755.

Krizhevsky, A., Sutskever, I., Hinton, G. E., 2012. Imagenet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25.

Krumbein, W. C., 1959. Trend surface analysis of contour-type maps with irregular control-point spacing. Journal of Geophysical Research, 64(7), 823–834.

Lary, D. J., Alavi, A. H., Gandomi, A. H., Walker, A. L., 2016. Machine learning in geosciences and remote sensing. Geoscience Frontiers, 7(1), 3–10.

LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278–2324.

Li, T., Shen, H., Yuan, Q., Zhang, L., 2020. Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM2. 5. ISPRS Journal of Photogrammetry and Remote Sensing, 167, 178–188.

Li, T., Shen, H., Yuan, Q., Zhang, X., Zhang, L., 2017. Estimating ground-level PM 2.5 by fusing satellite and station observations: A geo-intelligent deep learning approach. Geophysical Research Letters, 44(23).

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach. Journal of the Royal Statistical Society Series B: Statistical Methodology, 73(4), 423–498.

Liu, X., Kounadi, O., Zurita-Milla, R., 2022. Incorporating spatial autocorrelation in machine learning models using spatial lag and eigenvector spatial filtering features. ISPRS International Journal of Geo-Information, 11(4), 242.

Mai, G., Janowicz, K., Hu, Y., Gao, S., Yan, B., Zhu, R., Cai, L., Lao, N., 2022. A review of location encoding for GeoAI: Methods and applications. International Journal of Geographical Information Science, 36(4), 639–673.

Mariethoz, G., Renard, P., Straubhaar, J., 2010. The direct sampling method to perform multiple-point geostatistical simulations. Water Resources Research, 46(11), 2008WR007621.

Meyer, H., 2019. Importance of spatial predictor variable selection in machine learning applications – moving from data reproduction to spatial prediction.

Mole, D., Fiorentini, M. L., Thebaud, N., Cassidy, K. F., McCuaig, T. C., Kirkland, C. L., Romano, S. S., Doublier, M. P., Belousova, E. A., Barnes, S. J., Miller, J., 2014. Archean komatiite volcanism controlled by the evolution of early continents. Proceedings of the National Academy of Sciences, 111(28), 10083–10088.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., Sain, S., 2015. A multiresolution gaussian process model for the analysis of large spatial datasets. Journal of Computational and Graphical Statistics. https://www.tandfonline.com/doi/abs/10.1080/10618600.2014.914946

© Author(s) 2026. CC BY 4.0 License.

Raissi, M., Perdikaris, P., Karniadakis, G. E., 2019. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat., 2019. Deep learning and process understanding for data-driven Earth system science. Nature, 566(7743), 195–204.

Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1986. Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

Sekulić, A., Kilibarda, M., Heuvelink, G. B., Nikolić, M., Bajat, B., 2020. Random forest spatial interpolation. Remote Sensing, 12(10), 1687.

Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y., 2024. RoFormer: Enhanced transformer with rotary position embedding. Neurocomputing, 568, 127063.

Szatmári, G., Pásztor, L., Heuvelink, G. B. M., 2021. Estimating soil organic carbon stock change at multiple scales using machine learning and multivariate geostatistics. Geoderma, 403, 115356.

Tahmasebi, P., Sahimi, M., Mariethoz, G., Hezarkhani, A., 2012. Accelerating geostatistical simulations using graphics processing units (GPU). Computers & Geosciences, 46, 51–59.

Talebi, H., Peeters, L. J. M., Otto, A., Tolosana-Delgado, R., 2022. A truly spatial random forests algorithm for geoscience data analysis and modelling. Mathematical Geosciences, 54(1), 1–22.

Tiefelsdorf, M., Griffith, D. A., 2007. Semiparametric filtering of spatial autocorrelation: The eigenvector approach. Environment and Planning A, 39(5), 1193–1221.

Tobler, W. R., 1970. A computer movie simulating urban growth in the detroit region. Economic Geography, 46, 234–240.

Van Den Boogaart, K. G., Mueller, U., Tolosana-Delgado, R., 2017. An affine equivariant multivariate normal score transform for compositional data. Mathematical Geosciences, 49(2), 231–251.

Vance, T. C., Huang, T., Butler, K. A., 2024. Big data in Earth science: Emerging practice and promise. Science, 383(6688), eadh9607.

Wang, J., Zuo, R., Huang, D., Liu, M., 2026. The code of the neural process-based 2D stochastic simulation model (v1.0). Zenodo. https://doi.org/10.5281/zenodo.18834446.

Wikle, C. K., Zammit-Mangion, A., 2022. Statistical deep learning for spatial and spatio-temporal data (No. arXiv:2206.02218). arXiv. http://arxiv.org/abs/2206.02218

Wilford, J., De Caritat, P., Bui, E., 2016. Predictive geochemical mapping using environmental correlation. Applied Geochemistry, 66, 275–288.

Zhao, W., Ma, J., Liu, Q., Dou, L., Qu, Y., Shi, H., Sun, Y., Chen, H., Tian, Y., Wu, F., 2023. Accurate prediction of soil heavy metal pollution using an improved machine learning method: A case study in the pearl river delta, China. Environmental Science & Technology, 57(46), 17751–17761.

© Author(s) 2026. CC BY 4.0 License.

605 Zhu, A., Lu, G., Liu, J., Qin, C., Zhou, C., 2018. Spatial prediction based on third law of geography. Annals of GIS, 24(4), 225–240.