AUT Journal of Mathematics and Computing

AUT J. Math. Comput., 5(1) (2024) 1-9
https://doi.org/10.22060/AJMC.2023.22015.1130

Original Article

Simulating mixture of sub-Gaussian spatial data

Seyedeh Somayeh Mousavi[a], Adel Mohammadpour*[a]

[a] *Department of Mathematics and Computer Science, Amirkabir University of Technology (Tehran Polytechnic), Iran*

ABSTRACT: Spatial datasets may contain extreme values and exhibit heavy tails. So, the Gaussianity assumption for the corresponding random field is not reasonable. A sub-Gaussian α-stable (SGαS) random field may be more suitable as a model for heavy-tailed spatial data. This paper focuses on geostatistical data and presents an algorithm for simulating SGαS random fields.

Review History:

Received:15 December 2022
Revised:03 April 2023
Accepted:04 April 2023
Available Online:01 January 2024

Keywords:

Simulation
Spatial data
Geostatistical data
SGαS random field

AMS Subject Classification (2010):

60E07; 62H11; 68U20

1. Introduction

Geostatistical (*point-reference*) *data* is one type of spatial data that consists of measurements taken at specific spatial locations in a dense study region. In application, it may be one or more response variables measured in each spatial location. Usually, a Gaussian random field (GRF) is chosen for modeling spatial data. However, spatial datasets only sometimes behave as the realization of GRFs. Many environmental, earth, ecological, physical, biological, social, financial, and other variables may include extreme values and exhibit heavy tails in their distributions, which are non-consistent with a Gaussian transformation such that the Gaussianity assumption is not reasonable, e.g., [3, 7, 11, 20, 27]. In such cases, sub-Gaussian α-stable (SGαS) distributions as a particular sub-class of heavy-tailed distributions can be appropriate for modeling spatial data with heavy tails [12, 18, 22, 26, 27, 28, 30].

Simulating spatial data is necessary for evaluating the efficiency and robustness of spatial statistical methods. There are a few packages in spatial statistics to generate spatial data with finite variances such as `R` packages `gstat`, `geoR`, and `ggplot2`. Also, some approaches have been introduced for generating GRFs. Brouste et al. [4] proposed a method to simulate GRFs and fractional Brownian random fields. Kolyukhin and Minakov [14] introduced a statistical method for modeling three-dimensional GRFs inside a unit ball and a numerical procedure for generating the corresponding random realizations. Lang and Potthoff [15] presented two algorithms using fast

*Corresponding author.
E-mail addresses: s.s.mousavi@aut.ac.ir, adel@aut.ac.ir

Fourier transforms for the fast generation of GRFs on a rectangular region of $\mathbb{R}^d$. Lang and Schwab [16] simulated Isotropic GRFs on the sphere by Cholesky decomposition and Karhunen–Loève expansions concerning the spherical harmonic functions and the angular power spectrum. Lantuéjoul et al. [17] proposed a spectral algorithm to simulate the isotropic GRFs on a sphere equipped with a geodesic metric. Stein [32] described an algorithm for simulating differentiable stationary Gaussian processes. It works by first simulating a filtered version of the stationary Gaussian process using circulant embedding and then recovering the original process from the filtered version.

Modeling the dependence of non-Gaussian spatial random fields (NGRFs), especially skewed fields, and simulating them is still a challenging question. The core of many techniques for simulating non-Gaussian processes is used from GRFs simulation methods, e.g. [2, 9, 10, 24]. In recent decades, these simulation techniques have improved, and some problems of the numerical simulation of NGRFs have been used. Bevilacqua et al. [1] considered a skew-Gaussian process and proposed a new regression model and dependence analysis of heavy-tailed spatial data with asymmetric marginal distributions. Gräler [8] introduced spatial vine copulas that include the extreme behavior of a spatial random field and suggested a sequential simulation algorithm proceeding along a random path based on the copulas. Mahmoudian [19] proposed a flexible NGRF model with a stochastic skew latent structure to analyze asymmetric heavy-tailed extreme spatial data. Montoya-Noguera et al. [21] proposed a novel method for simulating non-stationary NGRFs when just sparse data is available. Sakamoto and Ghanem [29] presented an algorithm for simulating multi-dimensional non-stationary NGRFs. Vio et al. [34] investigated the Lognormal, Gamma, and Beta random fields. Xu and Genton [35] modeled non-Gaussian spatial data with a new class of trans-Gaussian random fields named Tukey g-and-h (TGH) random fields. The proposed TGH random fields have skewed and/or heavy-tailed marginal distributions.

Also, a few studies have been conducted on heavy-tailed spatial processes assuming a SGαS distribution, some of which we mention. Kamo et al. [13] extended a statistical model of multichannel nonnegative matrix factorization to a time-variant SGαS distribution. They employed a multivariate complex generalized Gaussian distribution as a SGαS distribution by restricting parameter α. Li et al. [18] assessed the applicability and transferability of the Generalized SGαS model via a dataset of electrical resistivity collected in three deep boreholes in the northern Ordos Basin in China. By analyzing statistical scaling behaviors, frequency distributions, and model fittings of resistivity data, it was found that resistivity data from each borehole indeed could be characterized by the Generalized SGαS model. Riva et al. [27] generated signals having univariate α-stable. They considered SGαS random fields subordinated to the truncate fractional Brownian motion as well as fields having symmetric α-stable increments subordinated to the truncate fractional Gaussian noise. Riva et al. [26] proposed a generalized SGαS model for spatial variables and spatial or temporal increments possessing heavy-tailed distributions.

In this paper, we are interested in simulating geostatistical data from K disjoint classes of the stationary Gaussian and SGαS random fields. In the next section, we express the needed definitions and the used concepts. Then, we explain a simulation algorithm of geostatistical data as a realization of the stationary SGαS random field in Section 3 and give two examples in Section 4. The paper is concluded in Section 5.

2. Spatial Data from K Classes of GRFs

Mathematically, the spatial data $\boldsymbol{z} = (z(\boldsymbol{s}_1), \ldots, z(\boldsymbol{s}_n))'$ in the n known spatial locations $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ on a study region $\mathbb{D} \subseteq \mathbb{R}^d;\ d \geq 2$, are interpreted as a realization of the real-valued spatial random field $\left\{Z(\boldsymbol{s}):\ \boldsymbol{s} \in \mathbb{D} \subset \mathbb{R}^d\right\}$. Consider a spatial GRF with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. We write the joint density function of an observed spatial data $\boldsymbol{z}$ as:

$$f(\boldsymbol{z};\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}}|\boldsymbol{\Sigma}|^{\frac{1}{2}}} e^{-\frac{1}{2}(\boldsymbol{z}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\boldsymbol{z}-\boldsymbol{\mu})}. \tag{1}$$

Now, let $\boldsymbol{Z}_k$, $k = 1, \ldots, K$, be a random vector with dimension n_k of a stationary spatial GRF $\left\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{D} \subset \mathbb{R}^d\right\}$ at known $n = \sum_{k=1}^K n_k$ spatial locations. In other words, we can write:

$$\boldsymbol{Z} = (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_K)' = \left(Z(\boldsymbol{s}_1^1), \ldots, Z(\boldsymbol{s}_{n_1}^1), \ldots, Z(\boldsymbol{s}_1^K), \ldots, Z(\boldsymbol{s}_{n_K}^K)\right)' \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{2}$$

where $Z(\boldsymbol{s}_i^k)$ is the i^{th} observation in the k^{th} class; $k = 1, \ldots, K$, $i = 1, \ldots, n_k$, n_k is the number of observations in the k^{th} class, and $1 < n_k < n = \sum_{k=1}^K n_k$.

Correlated Classes

In this case, there is no restriction on the covariance matrix $\boldsymbol{\Sigma}$. However, to reduce the overlapping between classes, we can consider different mean vectors in the classes. We assume that the vector $\boldsymbol{\mu}_{n\times 1} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)'$, where $\boldsymbol{\mu}_k$ is the mean vector of k^{th} class with size n_k and $n = \sum_{k=1}^K n_k$, in density function (1).

Independent Classes

A particular case can be considered for assigning the spatial data into K classes so that the differences between classes are related to the means vectors and the covariance matrices of classes can be simplified to a block diagonal. Hence, we set $\boldsymbol{\mu}_{n\times 1} = (\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K)'$ same as before and the covariance matrix as blocked form $\boldsymbol{\Sigma} = \oplus_{k=1}^{K}\boldsymbol{\Sigma}_k = \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K)$, where $\boldsymbol{\Sigma}_k$ is $n_k \times n_k$ covariance matrix of k^{th} class. Therefore, the joint density function of all K classes of the spatial data $\boldsymbol{z} = (\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K)$ is written as

$$\begin{aligned}
f(\boldsymbol{z}_1, \ldots, \boldsymbol{z}_K;\ \ \boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_K, \boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K) &= \left((2\pi)^{\frac{n_1+n_2+\cdots+n_K}{2}}\left(|\boldsymbol{\Sigma}_1|\ |\boldsymbol{\Sigma}_2|\cdots|\boldsymbol{\Sigma}_K|\right)^{\frac{1}{2}}\right)^{-1} \\
&\times \exp\left\{-\tfrac{1}{2}\begin{pmatrix} \boldsymbol{z}_1-\boldsymbol{\mu}_1 \\ \boldsymbol{z}_2-\boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{z}_K-\boldsymbol{\mu}_K \end{pmatrix}' \begin{bmatrix} \boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{\Sigma}_K^{-1} \end{bmatrix} \begin{pmatrix} \boldsymbol{z}_1-\boldsymbol{\mu}_1 \\ \boldsymbol{z}_2-\boldsymbol{\mu}_2 \\ \vdots \\ \boldsymbol{z}_K-\boldsymbol{\mu}_K \end{pmatrix}\right\} \\
&= \textstyle\prod_{k=1}^{K}(2\pi)^{-\frac{n_k}{2}}|\boldsymbol{\Sigma}_k|^{-\frac{1}{2}}\ \exp\left\{-\frac{1}{2}(\boldsymbol{z}_k-\boldsymbol{\mu}_k)'\boldsymbol{\Sigma}_k^{-1}(\boldsymbol{z}_k-\boldsymbol{\mu}_k)\right\} \\
&= \textstyle\prod_{k=1}^{K} f(\boldsymbol{z}_k;\ \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k).
\end{aligned} \tag{3}$$

2.1. Covariance matrix

Several situations based on the covariance matrix structure and type of spatial dependencies could be considered. Assuming the random field to be stationary and isotropic, the covariance matrix is displayed as $\boldsymbol{\Sigma} = [C(\|\boldsymbol{h}_{ij}\|)]$, where $C(\cdot)$ is the covariance function, $\|\boldsymbol{h}_{ij}\| = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ indicates the spatial lag between locations i and j, and $i, j = 1, \ldots, n$. Therefore, the covariance matrices in the case of the independent classes are as $\boldsymbol{\Sigma}_k = \left[C_k(\|\boldsymbol{h}_{ij}^k\|)\right]$; where $C_k(\cdot)$ is the covariance function of k-th class, $\|\boldsymbol{h}_{ij}^k\| = \|\boldsymbol{s}_i^k - \boldsymbol{s}_j^k\|$, $i, j = 1, \ldots, n_k$, and $i \neq j$, $k = 1, \ldots, K$. Also, the covariance functions have parametric forms that are selected from among various isotropic valid models (in isotropic spatial correlation structure), and the most common of them are:

- Exponential Model: $C_{\sigma^2,\theta}(\|\boldsymbol{h}\|) = \sigma^2 e^{-\frac{\|\boldsymbol{h}\|}{\theta}}$
- Gaussian Model: $C_{\sigma^2,\theta}(\|\boldsymbol{h}\|) = \sigma^2 e^{-(\frac{\|\boldsymbol{h}\|}{\theta})^2}$
- Spherical Model: $C_{\sigma^2,\theta}(\|\boldsymbol{h}\|) = \sigma^2\left[1 - \frac{3}{2}\left(\frac{\|\boldsymbol{h}\|}{\theta}\right) + \frac{1}{2}\left(\frac{\|\boldsymbol{h}\|}{\theta}\right)^3\right] I(\|\boldsymbol{h}\| \leq \theta)$
- Matérn Model: $C_{\sigma^2,\theta,\lambda}(\|\boldsymbol{h}\|) = \frac{\sigma^2}{2^{\lambda-1}\Gamma(\lambda)}\left(\frac{\|\boldsymbol{h}\|}{\theta}\right)^{\lambda}\kappa_\lambda\left(\frac{\|\boldsymbol{h}\|}{\theta}\right)$,

where σ^2, $\theta > 0$, and $\lambda > 0$ are the variance, spatial range, and smoothness parameters, respectively. $\Gamma(\cdot)$ is the gamma function and $\kappa_\lambda(\cdot)$ denotes the modified Bessel function of the second kind of order λ [6].

3. Spatial Data from K Classes of SGαS Random Fields

The spatial datasets do not always behave as the realization of a GRF. Their distributions may exhibit heavy tails, which are inconsistent with a transformed Gaussian distribution. Therefore, SGαS distributions can be a more appropriate choice. Hence, we devote this section to some needed concepts about SGαS random fields and mixture distributions.

Definition 3.1. *(α-Stable Random Variable, (Nolan [23])). A random variable X is called α-stable if its characteristic function has the form of*

$$\varphi_X(t) = E(\exp(itX)) = \begin{cases} \exp\left\{-\gamma^\alpha|t|^\alpha\left(1 - i\beta\,\mathrm{sign}(t)\tan\left(\frac{\pi\alpha}{2}\right)\right) + i\delta t\right\}, & \alpha \neq 1 \\ \exp\left\{-\gamma|t|\left(1 + i\beta\frac{2}{\pi}\,\mathrm{sign}(t)\log|t|\right) + i\delta t\right\}, & \alpha = 1 \end{cases} \tag{4}$$

with tail index $\alpha \in (0, 2]$, skewness $\beta \in [-1, 1]$, scale $\gamma \in (0, \infty)$ and shift $\delta \in \mathbb{R}$. Since these four parameters determine the characteristic function of an α-stable random variable, we denote stable distributions by $S_\alpha(\gamma, \beta, \delta)$.

Definition 3.2. *(SGαS Random Field, (Spodarev et al. [31])). Let $X \sim S_{\frac{\alpha}{2}}\left(\left(\cos\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}, 1, 0\right)$ denotes a positive stable and $\boldsymbol{Z} = (Z(\boldsymbol{s}_1), \ldots, Z(\boldsymbol{s}_n))' \sim N_n(\boldsymbol{0}, \boldsymbol{\Sigma})$ be independent of X. A stationary spatial random field $\{Y(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{D} \subseteq \mathbb{R}^d\}$ is called SGαS, if for each $n \geq 1$ the n-dimensional random vector*

$$\boldsymbol{Y} = (Y(\boldsymbol{s}_1), \ldots, Y(\boldsymbol{s}_n))' \overset{D}{=} X^{\frac{1}{2}}\boldsymbol{Z} + \boldsymbol{\delta} \tag{5}$$

is a spatial SGαS random vector with a shift vector $\boldsymbol{\delta}$ and dispersion matrix $\boldsymbol{\Sigma}$.

Notice that if $\boldsymbol{Z}$ is from a stationary and isotropic GRF, then the resulting SGαS random vector $\boldsymbol{Y}$ is strictly stationary and isotropic. Also, the dependence structure of a SGαS random field inherits from the underlying GRF [31]. Therefore, assuming a stationary SGαS random field, the dispersion matrix at all spatial locations is displayed as $\boldsymbol{\Sigma}$.

Definition 3.3. *(Mixture Distribution, (Casella and Berger [5], page 165)) A random variable is said to have a mixture distribution if it depends on a quantity that also has a distribution.*

For example, finite, countable, or infinite Gaussian mixtures are examples of mixture distributions when, e.g., the location parameter has a distribution with finite, countable, or infinite supports. The proposed model is a finite mixture but cannot be written in an additive form or convex combinations of some density functions. In this paper, we propose an algorithm for simulating a mixture of SGαS random fields using Definition 3.3. This algorithm consists of several steps and we describe the details of each step in the following.

Algorithm: Generating Spatial Data From a mixture of SGαS Distribution with K Classes

Inputs: n: total number of observations
K: number of classes
α: value of tail index
$\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$: spatial location points
$C_k(\|\boldsymbol{h}\|);\ k = 1, \ldots, K$: covariance functions of K classes
$\sigma_k^2,\ \theta_k,\ \lambda_k;\ k = 1, \ldots, K$: parameters of covariance functions
$p_k; k = 1, \ldots, K$: parameter of each class, $0 \leq p_k \leq 1$, $\sum_{k=1}^K p_k = 1$
$\boldsymbol{\delta}_k;\ k = 1, \ldots, K$: components of location vector $\boldsymbol{\delta}_{n\times 1} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K)'$ defined in (5)

1: Generate $(n_1, \ldots, n_K)$ from Multinomial distribution with parameters $(n, p_1, \ldots, p_K)$; $\sum_{k=1}^K n_k = n$
2: Select K subsets from $\{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n\}$ as $\{\boldsymbol{s}_1^1 \ldots, \boldsymbol{s}_{n_1}^1\}, \ldots, \{\boldsymbol{s}_1^K, \ldots, \boldsymbol{s}_{n_K}^K\}$
3: Compute the Euclidean distances between any two locations or spatial lags $\|\boldsymbol{h}_{ij}^k\| = \|\boldsymbol{s}_i^k - \boldsymbol{s}_j^k\|$, $i, j = 1, \ldots, n_k$, $i \neq j$
4: Construct the $n_k \times n_k$ covariance matrices $\boldsymbol{\Sigma}_k = \left[C_k(\|\boldsymbol{h}_{ij}^k\|)\right]$; $k = 1, \ldots, K$, $i, j = 1, \ldots, n_k$
5: Create the block matrix $\boldsymbol{\Sigma}_{n\times n} = \operatorname{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_K)$
6: Take $\boldsymbol{\mu}_{n\times 1} = [\boldsymbol{0}]$
7: $\boldsymbol{z} \leftarrow$ Generate multivariate Gaussian distribution $N_n(\boldsymbol{\mu}_{n\times 1}, \boldsymbol{\Sigma}_{n\times n})$ with density function (3)
8: $x \leftarrow$ Generate positive stable distribution $S_{\frac{\alpha}{2}}\left(\left(\cos\frac{\pi\alpha}{4}\right)^{\frac{2}{\alpha}}, 1, 0\right)$, using the characteristic function (4)
9: Take $\boldsymbol{\delta}_{n\times 1} = (\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K)'$
10: $\boldsymbol{y} \leftarrow$ Compute $x^{\frac{1}{2}}\boldsymbol{z} + \boldsymbol{\delta}$ using (5)

Outputs: $\boldsymbol{y}$: A vector of SGαS spatial data from K classes.

Select K Subsets from n Spatial Locations

In this article, we use two methods for selecting spatial locations of each class in Step 2 of the algorithm: (1) non-random selection according to the neighborhood and adjacency of locations in the desired map, regular or irregular grid, and (2) random selection. In both methods, we must be careful in assigning the shift vector in Step 9 of the proposed algorithm. The vectors $\boldsymbol{\delta}_k$, $k = 1, \ldots, K$, should be allocated according to the spatial locations of classes.

Correlated Classes

As we explained a general case for determining classes in the previous section, it can also be used in the proposed algorithm. Therefore, the overlap between classes can be controlled using shift vectors $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_k$. The covariance matrix $\boldsymbol{\Sigma}$ has not any restrictions. In this case, components of $\boldsymbol{\Sigma}$ as inputs are σ^2, β, and λ, and Step 4 of the algorithm should be adjusted based on our desired dependencies and spacial dependencies.

4. Simulation Examples

4.1. Gaussian Random Field

The purpose of this example is to create a set of geostatistical data having 441 spatial locations with one recorded response at each location ($n = 441$) from three distinct distributions or classes ($K = 3$) of a GRF. We created spatial locations on a 21×21 regular grid at the square $[0, 20] \times [0, 20]$. First, the sizes of classes were generated from a Multinomial distribution with equal percentages $p_1 = p_2 = p_3 = \frac{1}{3}$ that resulted in $n_1 = 154$, $n_2 = 140$, and $n_3 = 147$.

We want to simulate a sample of multivariate Gaussian distribution with three classes with $\boldsymbol{\mu}_1 = -2\,\mathbf{1}_{n_1\times 1}$, $\boldsymbol{\mu}_2 = \mathbf{0}_{n_1\times 1}$, $\boldsymbol{\mu}_3 = 2\,\mathbf{1}_{n_2\times 1}$, and Matérn covariance function as spatial correlation structure. We considered four cases to determine the locations of each class and the covariance matrix:

1. Non-random selection of the locations of each class according to the adjacency and $\boldsymbol{\Sigma}$ as a blocked form,
2. Similar to case 1, but the covariance matrix is correlated,
3. Random selection of the locations of each class and $\boldsymbol{\Sigma}$ has a blocked form, and
4. Similar to case 3, but the covariance matrix is correlated.

In cases 1 and 3, parameter values were set to $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 1$, $\beta_1 = \beta_2 = \beta_3 = \frac{1}{7\sqrt{3}}$, and $\lambda_1 = \lambda_2 = \lambda_3 = 2$. In cases 2 and 4, parameter values were set to $\sigma^2 = 1$, $\beta = \frac{1}{7\sqrt{3}}$, and $\lambda = 2$. We used the known package `MASS` [25] of the `R` software for this simulation. The perspective and contour plots of the generated spatial data from three classes of the mentioned Gaussian distribution have shown in Figure 1.

4.2. SGαS Random Field

In this example, we used the setting of the Gaussian random field example. The perspective and contour plots of the spatial data from three classes of a SGαS distribution with $\alpha = 0.75$, the different shift vectors, and the dispersion matrices have shown in Figure 2. In this example $n_1 = 151$, $n_2 = 149$, and $n_3 = 141$ were selected. We set the location vectors of three classes as follows: $\boldsymbol{\delta}_1 = 22\,\mathbf{1}_{n_1\times 1}$, $\boldsymbol{\delta}_2 = 28\,\mathbf{1}_{n_2\times 1}$, and $\boldsymbol{\delta}_3 = 33\,\mathbf{1}_{n_3\times 1}$. Also, the parameter values in cases 2 and 4 were set $\sigma^2 = 2$, $\beta = \frac{1}{\sqrt{3}}$, and $\lambda = \frac{1}{2}$. We used the `R` package `stable` [33] for this simulation.

5. Conclusion

Some natural phenomena include extreme values and exhibit heavy tails in which no Gaussian transformation can be found. An SGαS random field may be a more appropriate choice in such cases. This paper presented a comprehensible algorithm for simulating geostatistical data from a mixture of SGαS distributions with various α at the determined spatial locations. We assumed the differences between classes are related to both the shift vectors and the dispersion matrices of classes. Then, we investigated four different methods by considering two cases for choosing spatial locations of each class as randomly and non-randomly, and two cases for covariance matrix $\boldsymbol{\Sigma}$ as blocked form and correlated form. These four cases were also used for generating geostatistical data having a Gaussian distribution.

Acknowledgment

The authors thank the associate editor and anonymous referees for their comments that helped to improve the manuscript.

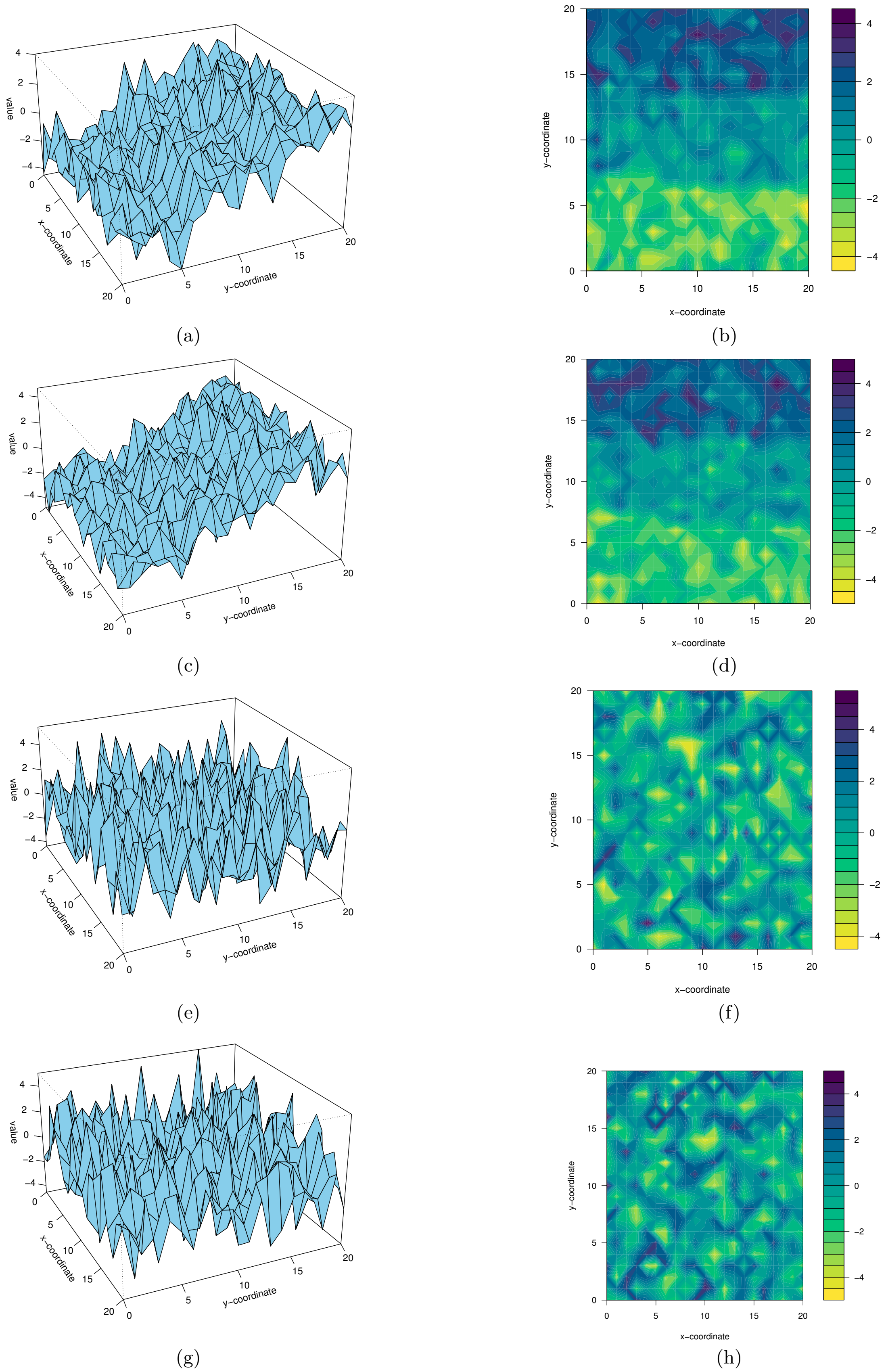

Figure 1: Perspective (left) and contour (right) plots of the spatial data from three classes of Gaussian distributions: (a,b) non-random locations and blocked form for $\mathbf{\Sigma}$, (c,d) non-random locations and correlated form for $\mathbf{\Sigma}$, (e,f) random locations and blocked form for $\mathbf{\Sigma}$, (g, h) random locations and correlated form for $\mathbf{\Sigma}$.

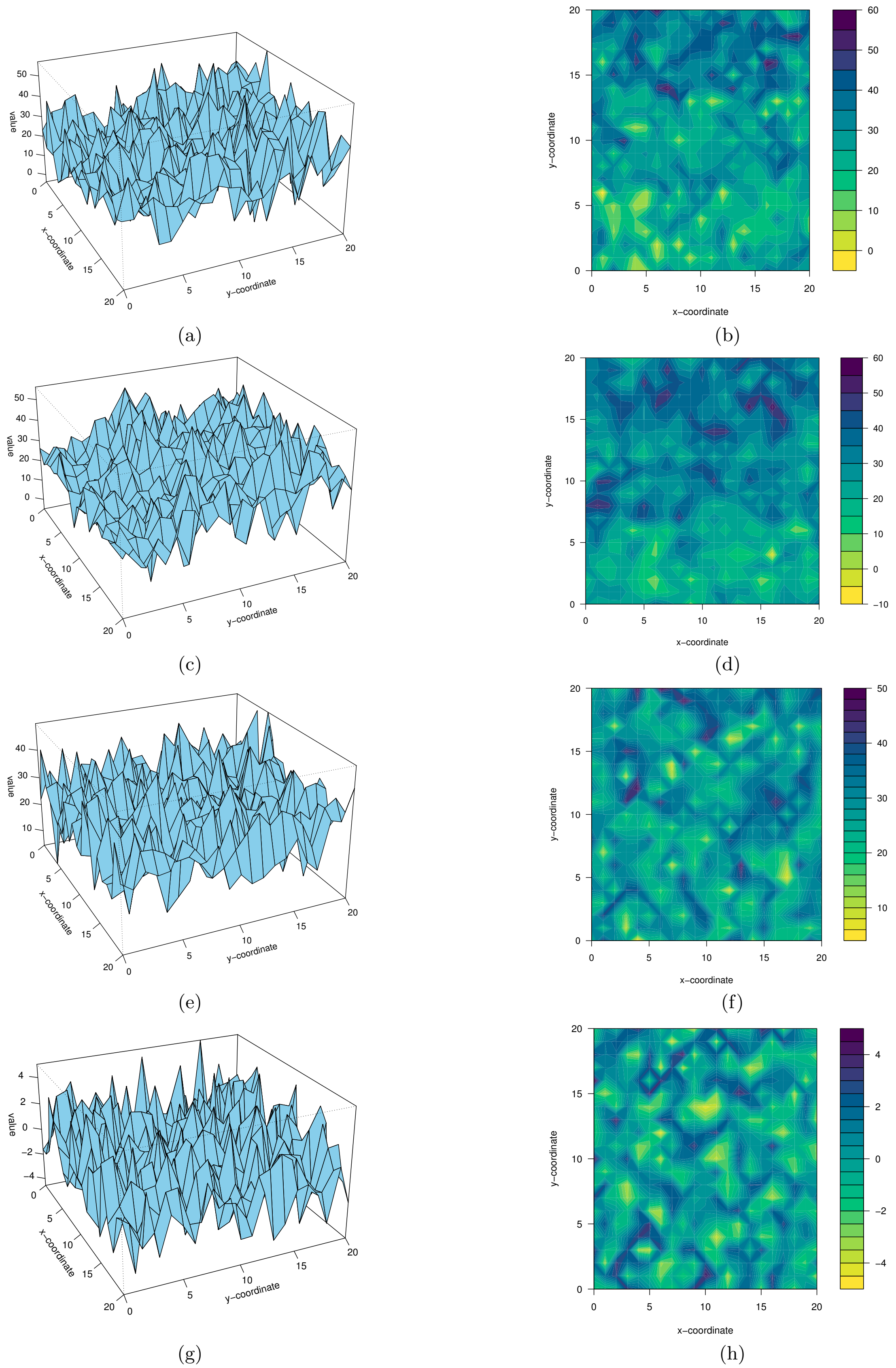

Figure 2: Perspective (left) and contour (right) plots of the spatial data from three classes of SGαS distributions with $\alpha = 0.75$: (a,b) non-random locations and blocked form for $\mathbf{\Sigma}$, (c,d) non-random locations and correlated form for $\mathbf{\Sigma}$, (e,f) random class locations and blocked form for $\mathbf{\Sigma}$, (g, h) random class locations and correlated form for $\mathbf{\Sigma}$.

References

[1] M. Bevilacqua, C. Caamaño-Carrillo, R. B. Arellano-Valle, and V. Morales-Oñate, *Non-Gaussian Geostatistical Modeling Using (Skew) t Processes*, Scandinavian Journal of Statistics, 48 (2021), pp. 212–245.

[2] P. Bocchini and G. Deodatis, *Critical Review and Latest Developments of a Class of Simulation Algorithms for Strongly Non-Gaussian Random Fields*, Probabilistic Engineering Mechanics, 23 (2008), pp. 393–407.

[3] G. Boffetta, A. Mazzino, and A. Vulpiani, *Twenty-Five Years of Multifractals in Fully Developed Turbulence: A Tribute to Giovanni Paladin*, Journal of Physics A: Mathematical and Theoretical, 41 (2008), p. 363001.

[4] A. Brouste, J. Istas, and S. Lambert-Lacroix, *On Fractional Gaussian Random Fields Simulations*, Journal of Statistical Software, 23 (2008), pp. 1–23.

[5] G. Casella and R. L. Berger, *Statistical Inference*, Thomson Learning, 2001.

[6] N. Cressie, *Statistics for Spatial Data*, John Wiley & Sons, 2015.

[7] V. Ganti, A. Singh, P. Passalacqua, and E. Foufoula-Georgiou, *Subordinated Brownian Motion Model for Sediment Transport*, Physical Review E, 80: 011111 (2009).

[8] B. Gräler, *Modelling Skewed Spatial Random Fields Through the Spatial Vine Copula*, Spatial Statistics, 10 (2014), pp. 87–102.

[9] M. Grigoriu, *Simulation of Stationary Process via a Sampling Theorem*, Journal of sound and vibration, 166 (1993), pp. 301–313.

[10] M. Grigoriu, *Applied Non-Gaussian Processes: Examples, Theory, Simulation, Linear Random Vibration, and MATLAB Solutions*, PTR Prentice Hall, 1995.

[11] A. Guadagnini, S. P. Neuman, T. Nan, M. Riva, and C. L. Winter, *Scalable Statistics of Correlated Random Variables and Extremes Applied to Deep Borehole Porosities*, Hydrology and Earth System Sciences, 19 (2015), pp. 729–745.

[12] A. Guadagnini, M. Riva, and S. P. Neuman, *Recent Advances in Scalable Non-Gaussian Geostatistics: The Generalized Sub-Gaussian Model*, Journal of hydrology, 562 (2018), pp. 685–691.

[13] K. Kamo, Y. Kubo, N. Takamune, D. Kitamura, H. Saruwatari, Y. Takahashi, and K. Kondo, *Joint-Diagonalizability-Constrained Multichannel Nonnegative Matrix Factorization Based on Multivariate Complex Sub-Gaussian Distribution*, 28th European Signal Processing Conference (EUSIPCO), (2021), pp. 890–894.

[14] D. Kolyukhin and A. Minakov, *Simulation of Gaussian Random Field in a Ball*, Monte Carlo Methods and Applications, 28 (2022), pp. 85–95.

[15] A. Lang and J. Potthoff, *Fast Simulation of Gaussian Random Fields*, Monte Carlo Methods and Applications, 17 (2011), pp. 195–214.

[16] A. Lang and C. Schwab, *Isotropic Gaussian Random Fields on the Sphere: Regularity, Fast Simulation and Stochastic Partial Differential Equations*, The Annals of Applied Probability, 25 (2015), pp. 3047–3094.

[17] C. Lantuéjoul, X. Freulon, and D. Renard, *Spectral Simulation of Isotropic Gaussian Random Fields on a Sphere*, Mathematical Geosciences, 51 (2019), pp. 999–1020.

[18] K. Li, J. Wu, T. Nan, X. Zeng, L. Yin, and J. Zhang, *Analysis of Heterogeneity in a Sedimentary Aquifer Using Generalized Sub-Gaussian Model Based on Logging Resistivity*, Stochastic Environmental Research and Risk Assessment, (2022), pp. 1–17.

[19] B. Mahmoudian, *A Skewed and Heavy-Tailed Latent Random Field Model for Spatial Extremes*, Journal of Computational and Graphical Statistics, 26 (2017), pp. 658–670.

[20] M. M. Meerschaert, T. J. Kozubowski, F. J. Molz, and S. Lu, *Fractional Laplace Model for Hydraulic Conductivity*, Geophysical Research Letters, 31 (2004).

[21] S. Montoya-Noguera, T. Zhao, Y. Hu, Y. Wang, and K.-K. Phoon, *Simulation of Non-Stationary Non-Gaussian Random Fields From Sparse Measurements Using Bayesian Compressive Sampling and Karhunen-Loève Expansion*, Structural Safety, 79 (2019), pp. 66–79.

[22] G. R. Naik and W. Wang, *Audio Analysis of Statistically Instantaneous Signals With Mixed Gaussian Probability Distributions*, International Journal of Electronics, 99 (2012), pp. 1333–1350.

[23] J. P. Nolan, *Univariate Stable Distributions*, Springer, 2020.

[24] V. A. Ogorodnikov and S. M. Prigarin, *Numerical Modelling of Random Processes and Fields: Algorithms and Applications*, Vsp, 1996.

[25] B. Ripley, B. Venables, D. M. Bates, K. Hornik, A. Gebhardt, and D. Firth, ***MASS**: Support Functions and Datasets for Venables and Ripley's MASS*, (2022). R Package Version 7.3-58.1.

[26] ——, *New Scaling Model for Variables and Increments with Heavy-Tailed Distributions*, Water Resources Research, 51 (2015), pp. 4623–4634.

[27] M. Riva, S. P. Neuman, and A. Guadagnini, *Sub-Gaussian Model of Processes with Heavy-Tailed Distributions Applied to Air Permeabilities of Fractured Tuff*, Stochastic Environmental Research and Risk Assessment, 27 (2013), pp. 195–207.

[28] M. Riva, M. Panzeri, A. Guadagnini, and S. P. Neuman, *Simulation and Analysis of Scalable Non-Gaussian Statistically Anisotropic Random Functions*, Journal of Hydrology, 531 (2015), pp. 88–95.

[29] S. Sakamoto and R. Ghanem, *Simulation of Multi-Dimensional Non-Gaussian Non-Stationary Random Fields*, Probabilistic Engineering Mechanics, 17 (2002), pp. 167–176.

[30] M. Siena, A. Guadagnini, A. Bouissonnié, P. Ackerer, D. Daval, and M. Riva, *Generalized Sub-Gaussian Processes: Theory and Application to Hydrogeological and Geochemical Data*, Water Resources Research, 56 (2020), p. e2020WR027436.

[31] E. Spodarev, E. Shmileva, and S. Roth, *Extrapolation of Stationary Random Fields*, in Stochastic Geometry, Spatial Statistics and Random Fields, Springer, 2015, pp. 321–368.

[32] M. L. Stein, *Simulation of Gaussian Random Fields With One Derivative*, Journal of Computational and Graphical Statistics, 21 (2012), pp. 155–173.

[33] B. Swihart, J. Lindsey, and P. Lambert, ***stable**: Probability Functions and Generalized Regression Models for Stable Distributions*, (2022). R Package Version 1.1.6.

[34] R. Vio, P. Andreani, and W. Wamsteker, *Numerical Simulation of Non-Gaussian Random Fields with Prescribed Correlation Structure*, Publications of the Astronomical Society of the Pacific, 113 (2001), p. 1009.

[35] G. Xu and M. G. Genton, *Tukey g-and-h Random Fields*, Journal of the American Statistical Association, 112 (2017), pp. 1236–1249.

Please cite this article using:

Seyedeh Somayeh Mousavi, Adel Mohammadpour, Simulating mixture of sub-Gaussian spatial data, AUT J. Math. Comput., 5(1) (2024) 1-9
https://doi.org/10.22060/AJMC.2023.22015.1130