ATLANTIS
PRESS

Generalized Skew Laplace Random Fields: Bayesian Spatial Prediction for Skew and Heavy Tailed Data

Mohammad Mehdi Saber[1,*], Alireza Nematollahi[2], Mohsen Mohammadzadeh[3]

[1]*Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran*
[2]*Department of Statistics, Shiraz University, Shiraz, Iran*
[3]*Department of Statistics, Tarbiat Modares University, Tehran, Iran*

ARTICLE INFO

ABSTRACT

Earlier works on spatial prediction issue often assume that the spatial data are realization of Gaussian random field. However, this assumption is not applicable to the skewed and kurtosis distributed data. The closed skew normal distribution has been used in these circumstances. As another alternative, we apply generalized skew Laplace distributions for defining a skew and heavy tailed random field for Bayesian prediction. Simulation study and a real problem are then applied to evaluate the performance of this model.

1. INTRODUCTION

Spatial prediction is an important subject in statistical analysis of environmental data which are often spatially correlated. In such cases it is supposed that data come from a Gaussian random field (RF). When spatial data are non-Gaussian but a Gaussian transformation for them exist, Oliveira *et al.* [1], Oliveira and Ecker [2], Oliveira [3] propose spatial prediction for transformed RFs. However, normalizing transformation is unknown in application and interpretation of the transformed data is also more difficult than original data [4,5]. When distribution of data has an appropriate number of similarities with normal distribution but is asymmetric, Kim and Mallick [6] used a skew Gaussian (SG) RF. Hosseini *et al.* [7,8] made inference on spatial generalized linear mixed models (SGLMMs) with SG latent variables. Because of some shortage in SG distribution such as nonclosing under conditioning, Dominguez-Molina *et al.* [9] defined multivariate closed skew normal (CSN) distribution. Karimi and Mohammadzadeh [10,11] and Karimi *et al.* [12] used CSN RF for Bayesian spatial prediction, Bayesian spatial regression and inversion of seismic data. Hosseini and Mohammadzadeh [13] have done Bayesian prediction for SGLMM with CSG latent variables. However, CSN random variables are not appropriate for modeling data with heavy tails. Since *t*-distribution is an applicable distribution to real data with heavy tails, Røislien and Omre [14] defined a *t*-distributed RF, but this RF could not be suitable for modeling the skew data. An appropriate choice for modeling skew and heavy tailed data which is well defined for RF and has some similar characteristics with Gaussian field is multivariate generalized asymmetric Laplace (GAL) distribution introduced by Kozubowski *et al.* [15]. In this paper, we have applied multivariate GAL distributions for defining the skew and heavy tail RFs. The Bayesian predictions are then achieved by using this RF. A simulation study is performed to check the validity of the model and the proposed prediction method is used on an environmental data set.

A review on the multivariate GAL distribution is presented in Section 2. As in Section 3, the GAL RF is defined, the Bayesian spatial prediction has been considered by GAL RF in Section 4. A simulation study is discussed in Section 5 followed by application of the GAL RF to environmental real data in Section 6.

*Corresponding author. Email: mmsaber@eghlid.ac.ir

2. MULTIVARIATE GAL RANDOM VARIABLE

In this section, we review the definition and basic setup of the multivariate GAL distribution introduced by [15].

Definition 2.1. Multivariate GAL law. A continuous p-dimensional random vector X has a GAL distribution, denoted by $X \sim GAL_p(\mu, \Sigma, q)$, if its characteristic function at $t \in R^p$ is

$$\phi(t) = \left(\frac{1}{1 + \frac{1}{2}t^T \Sigma t - i\, t^T \mu}\right)^q \tag{1}$$

where the location $\mu \in R^p$, nonnegative definite dispersion $p \times p$ matrix Σ and extension $q > 0$ are its parameters, and T denotes the matrix transpose. When $q = 1$, we deal with asymmetric Laplace (AL) distribution, denoted by $X \sim AL_p(\mu, \Sigma)$. By [15], "*The significance of AL distribution is partially due to the fact that these arise rather naturally as the only distributional limits for (appropriately normalized) random sums*

$$X_1 + ... + X_{N_g} \tag{2}$$

of independent and identically random vector X_i with finite second moments as g converges to zero, where the integer-valued random variable N_g is independent of $\{X_i\}$ and has a Geometric distribution with mean $\frac{1}{g}$. Sums similar to (2) *frequently appear in many applied problems of fields such as Biology, Economics, Insurance mathematics, Reliability and other fields (Kalashnikov* [16]*; Klebanov et al.* [17]*). Thus, AL distributions have a wide variety of applications (Kotz et al.* [18]*). The AL distributions play an analogous role among the heavy tailed Geometric stable laws approximating sums* (2) *without the restriction of finite second moment* [17]*. From amongst the stable laws, Gaussian distributions have finite moments of all orders, and their theory is elegant and straight forward. However, in spite of finiteness of moments, their tails are substantially longer than those of the Gaussian laws. This, coupled with the fact that they allow for asymmetry, makes them more flexible in modeling heavy tailed and asymmetrical data.*"

If matrix Σ is positive-definite, the distribution is truly p-dimensional and has a probability density function of the following form:

$$f_X(x) = \frac{2e^{\mu^T \Sigma^{-1} x}}{(2\pi)^{\frac{p}{2}} \Gamma(q) |\Sigma|^{\frac{1}{2}}} \left(\frac{Q(x)}{\Psi(\Sigma, \mu)}\right)^{q - \frac{p}{2}} K_{q - \frac{p}{2}}(Q(x, \Sigma)\Psi(\mu, \Sigma)), \tag{3}$$

where $K_u(u)$ is the modified Bessel function or Bessel function of type 3 with index u, $Q(x, \Sigma) = \sqrt{x^T \Sigma^{-1} x}$ and $\Psi(\mu, \Sigma) = \sqrt{2 + \mu^T \Sigma^{-1}\mu}$.

We have the following representation for multivariate GAL random variable

$$X = \mu G + \sqrt{G} N, \tag{4}$$

where G has a standard Gamma distribution with shape parameter q and $N \sim N_p(0, \Sigma)$, showing that GAL distributions are location-scale mixtures of normal distributions. By Equation (4), this distribution has some similarities with normal distribution. Therefore, it is appropriate to modeling data which has some similar characteristics with Gaussian data but are skew and heavy tailed. The stochastic representation (4) leads to many further properties of GAL random vectors, including moments, marginal, conditional distributions and linear transformations (Kozubowski *et al.* [15]). Here, we only review some of useful results for our work.

If $X \sim GAL_p(\mu, \Sigma, q)$ then

$$E(X) = \mu q \tag{5}$$

$$Cov(X) = q\left(\Sigma + \mu \mu^T\right) \tag{6}$$

If A be a real matrix $l \times p$. Then,

$$AX \sim GAL_l\left(A\mu, A\Sigma A^T, q\right) \tag{7}$$

Also, let $X \sim GAL_p(\mu, \Sigma, q)$ and consider the partition $X^T = \left(X_1^T, X_2^T\right)$ with $\dim(X_1) = p_1$, $\dim(X_2) = p_2 = p - p_1$ and the corresponding partition of the parameters (μ, Σ). Then

$$X_1 \sim GAL_{p_1}\left(\mu_1, \Sigma_{11}, q\right) \tag{8}$$

If q = 1, the conditional distribution of X_2 given $X_1 = x_1$ is the generalized p_2 dimensional hyperbolic distribution $H_{p_2}(\lambda, \alpha, \beta, \delta, m, \Delta)$. By Kotz *et al.* [18], for conditional mean and variance we have the following results:

$$E\left(X_2|X_1 = x_1\right) = \Sigma_{21}\Sigma_{11}^{-1}x_1 + \left(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1\right)\frac{Q(x_1, \Sigma_{11})}{\Psi(\mu_1, \Sigma_{11})}R_{1-\frac{p_1}{2}}\left(\Psi(\mu_1, \Sigma_{11})Q(x_1, \Sigma_{11})\right) \tag{9}$$

$$Var\left(X_2|X_1 = x_1\right) = \frac{Q(x_1, \Sigma_{11})}{\Psi(\mu_1, \Sigma_{11})}\left(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}\right)R_{1-\frac{p_1}{2}}\left(\Psi(\mu_1, \Sigma_{11})Q(x_1, \Sigma_{11})\right)$$
$$+ \left(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1\right)\left(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1\right)^T\left(\frac{Q(x_1, \Sigma_{11})}{\Psi(\mu_1, \Sigma_{11})}\right)^2 G\left(x_1, \mu_1, \Sigma_{11}, p_1\right) \tag{10}$$

where $R_\zeta(x) = \frac{K_{\zeta+1}(x)}{K_\zeta(x)}$ and

$$G\left(x_1, \mu_1, \Sigma_{11}, p_1\right) = R_{1-\frac{p_1}{2}}\left(\Psi(\mu_1, \Sigma_{11})Q(x_1, \Sigma_{11})\right)R_{2-\frac{p_1}{2}}\left(\Psi(\mu_1, \Sigma_{11})Q(x_1, \Sigma_{11})\right) - \left(R_{1-\frac{p_1}{2}}\left(\Psi(\mu_1, \Sigma_{11})Q(x_1, \Sigma_{11})\right)\right)^2$$

3. SPATIAL MODEL

First of all, we use the following definition for GAL RF.

Definition 3.1. A RF $\mathcal{Z} = \left\{Z(s) : s \in D \subseteq R^d\right\}$ is termed a GAL RF if $Z = (Z(s_1), \ldots, Z(s_n)) \sim GAL_n\left(\mu, \frac{1}{q}C - \mu\mu^T, q\right)$ for all configurations $(s_1, \ldots, s_n) \in D \times \ldots \times D$ and all $n \in \mathbb{N}$.

The parameters in this RF has been determined

Let $Z(s_1), \ldots, Z(s_n)$ be the observations from a GAL RF $\left\{Z(s) : s \in D \subseteq R^d\right\}$ at n locations $(s_1, \ldots, s_n)$. For predicting $Z(s_0)$ at new location s_0, based on the observations $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))$, we define $\mathbf{Z}^* = \left(Z(s_0), \mathbf{Z}^T\right)^T$. Then we have $\mathbf{Z}^* \sim GAL_{n+1}\left(F^*\beta, \Sigma^*, q\right)$ where $\Sigma^* = \frac{C^*}{q} - F^*\beta\beta^T F^{*T}, f(s_0) = \left(f_1(s_0), \ldots, f_r(s_0)\right)^T, F^* = \left(f(s_0), F^T\right)^T, F = [f_j(s_i)]_{n \times r}$ are known regression functions (covariates), β is the regression coefficients, $C^* = \begin{pmatrix} C_{00} & c^T \\ c & C \end{pmatrix}$ is spatial covariance matrix for all observations and prediction, $c = (C_{0i})_{n \times 1}$, $C = Var(\mathbf{Z})$ and $C_{ij} = Cov\left(Z(s_i), Z(s_j)\right)$.

Note 3.1. In some works such as Karimi and Mohammadzadeh [10] that have used CSN models, they used C^* as scale matrix of $\mathbf{Z}^*$. This choice does not consider spatial structure properly, since $Var(\mathbf{Z}^*) \neq C^*$. Also, there is no another choice for parameters in CSN model which results in $Var(\mathbf{Z}^*) = C^*$. Here for GAL model, we consider $\frac{C^*}{q} - F^*\beta\beta^T F^{*T}$ as the scale matrix of $\mathbf{Z}^*$ that by Equation (6) results in $Var(\mathbf{Z}^*) = C^*$. In Section 5 we will show the efficiency of choosing $\frac{C^*}{q} - F^*\beta\beta^T F^{*T}$ over C^* through a simulation study, too.

Now, the best predictor of $Z(s_0)$ based on the square error loss function $L\left(\hat{Z}(s_0), Z(s_0)\right) = E\left[\left(\hat{Z}(s_0) - Z(s_0)\right)^2\right]$ is given by $E(Z(s_0)|\mathbf{Z})$, that can be computed by conditional distribution of $Z(s_0)|\mathbf{Z}$. In applications, the regression coefficients and spatial correlation structure are unknown and they have to be estimated. For parameter estimation we can only use the observation vector $\mathbf{Z}$ which has n-dimensional GAL distribution, $\mathbf{Z} \sim GAL_n\left(F\beta, \frac{C}{q} - F\beta\beta^T F^T, q\right)$. We assume a stationary GAL RF with stationary spatial covariance function $C(h) = \sigma^2\rho(h, \theta)$, where $\rho(., \theta)$ is a known correlation function, θ is spatial correlation parameter and σ^2 is variance of the RF. Let $\eta = \left(\beta^T, \sigma^2, \theta\right)^T$ and choose the priors $\beta \sim N_r(\beta_0, \Sigma_0), \sigma^2 \sim IG(\varphi, \zeta)$ and also suitable improper prior $\pi(\theta) = 1$, then the posterior density of η is given by

$$\pi(\eta|Z) \propto f(Z|\eta)\phi_r(\beta, \beta_0, \Sigma_0)\frac{\zeta^\varphi}{\Gamma(\varphi)}\left(\frac{1}{\sigma^2}\right)^{\varphi+1}e^{\frac{-\zeta}{2\sigma^2}}\pi(\theta). \tag{11}$$

This posterior density has a complicated form. Therefore we use Markov chain Monte Carlo (MCMC) method to generate sample from the posterior distribution of the parameters. To use Gibbs sampler, the derived full conditional distributions are given by

$$\pi(\beta|Z, \sigma^2, \theta) \propto f(Z|\eta)\phi_r(\beta, \beta_0, \Sigma_0) \tag{12}$$

$$\pi(\sigma^2|Z, \beta, \theta) \propto f(Z|\eta)\frac{\zeta^\varphi}{\Gamma(\varphi)}\left(\frac{1}{\sigma^2}\right)^{\varphi+1}e^{\frac{-\zeta}{2\sigma^2}} \tag{13}$$

$$\pi\left(\boldsymbol{\theta}|\mathbf{Z},\boldsymbol{\beta},\sigma^2\right) \propto f\left(\mathbf{Z}|\boldsymbol{\eta}\right) \tag{14}$$

These distributions do not have a closed form solution. For generating data from these densities, we use Metropolis–Hastings (MH) algorithms in Gibbs sampler. The proposal distribution $g_{\boldsymbol{\beta}}\left(\mathbf{y}\right)\ :\ N_r\left(\boldsymbol{\beta}, diag\left(b^2\right)\right)$ is used for (12), where b is a suitable number that controls efficiency in MH sampling. The proposal distributions for (13) and (14) are $g_{\sigma^2}\left(\mathbf{y}\right)\ :\ Gamma\left(\alpha_0, \frac{1}{\sigma^2}\right)$ and $g_{\theta}\left(\mathbf{y}\right) = 1$, respectively. After generating a random sample $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{N^*}$ with size N^* of the parameter $\boldsymbol{\eta} = \left(\boldsymbol{\beta}^T, \sigma^2, \theta\right)^T$, in the second stage if $q = 1$ we compute $E\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}_i\right)$ then

$$E\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}\right) = \frac{1}{N^*}\sum_{i=1}^{N^*} E\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}_i\right). \tag{15}$$

The conditional expectations in summation of Equation (15) is provided in following theorem.

Theorem 4.1. *Suppose* $\boldsymbol{Z}^* \sim GAL_{n+1}\left(\boldsymbol{F}^*\boldsymbol{\beta}, \boldsymbol{\Sigma}^*, 1\right)$, *then*

$$E\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}\right) = \boldsymbol{l}^T\boldsymbol{\Gamma}^{-1}\mathbf{Z} + \left(\boldsymbol{f}^T\left(\boldsymbol{s}_0\right) - \boldsymbol{l}^T\boldsymbol{\Gamma}^{-1}\boldsymbol{F}\right)\boldsymbol{\beta}\frac{Q(\mathbf{Z},\boldsymbol{\Gamma})}{\Psi(\boldsymbol{F}\boldsymbol{\beta},\boldsymbol{\Gamma})}R_{1-\frac{n}{2}}\left(\psi\left(\boldsymbol{F}\boldsymbol{\beta}, \boldsymbol{\Gamma}\right)Q\left(\mathbf{Z}, \boldsymbol{\Gamma}\right)\right)$$

where $\boldsymbol{\Gamma} = \mathbf{C} - \boldsymbol{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T$ *and* $\boldsymbol{l} = \boldsymbol{c} - \boldsymbol{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{f}^T\left(\boldsymbol{s}_0\right)$.

Proof. Let $\boldsymbol{Z}^* = \left(Z\left(\boldsymbol{s}_0\right), \mathbf{Z}^T\right)^T \sim GAL_{n+1}\left(\boldsymbol{F}^*\boldsymbol{\beta}, \boldsymbol{\Sigma}^*, q\right)$. We use Equation (9) in order to compute $E\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}\right)$. From $\boldsymbol{F}^*\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{f}^T\left(\boldsymbol{s}_0\right) \\ \mathbf{F} \end{pmatrix}\boldsymbol{\beta}$ it can be easily concluded that $\boldsymbol{\mu}_2 = \boldsymbol{f}^T\left(\boldsymbol{s}_0\right)\boldsymbol{\beta}$ and $\boldsymbol{\mu}_1 = \mathbf{F}\boldsymbol{\beta}$. We only need to find suitable matrices $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{11}$. So we have to partition the matrix $\boldsymbol{\Sigma}^*$.

$$\boldsymbol{\Sigma}^* = \frac{\mathbf{C}^*}{q} - \mathbf{F}^*\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^{*T} = \begin{pmatrix} C_{00} & \boldsymbol{c}^T \\ \boldsymbol{c} & \boldsymbol{C} \end{pmatrix} - \begin{pmatrix} \boldsymbol{f}^T\left(s_0\right) \\ \mathbf{F} \end{pmatrix}\boldsymbol{\beta}\boldsymbol{\beta}^T\left(\boldsymbol{f}\left(s_0\right), \mathbf{F}^T\right)$$

$$= \begin{pmatrix} C_{00} - \boldsymbol{f}^T\left(s_0\right)\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{f}\left(s_0\right) & \boldsymbol{c}^T - \boldsymbol{f}^T\left(s_0\right)\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T \\ \boldsymbol{c} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{f}\left(s_0\right) & \boldsymbol{C} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T \end{pmatrix}$$

Therefore, $\boldsymbol{\Sigma}_{12} = \boldsymbol{c} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{f}\left(s_0\right)$ and $\boldsymbol{\Sigma}_{11} = \boldsymbol{C} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T$ that complete proof.

For computing variance of prediction, $Var\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}\right)$, we use Equation (10) with the same replacements of $\boldsymbol{\mu}_1 = \mathbf{F}\boldsymbol{\beta}$, $\boldsymbol{\mu}_2 = \boldsymbol{f}^T\left(\boldsymbol{s}_0\right)\boldsymbol{\beta}$, $\boldsymbol{\Sigma}_{12} = \boldsymbol{c} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{f}\left(s_0\right)$, $\boldsymbol{\Sigma}_{11} = \boldsymbol{C} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T$. There is no necessity for prediction in one location at a fix stage. By using Theorem 4.1, we can predict in more than two locations. If $q \neq 1$ we generate samples $Z_1, \dots, Z_M$ from $Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}_i$ for all $i = 1, \dots, N^*$ for generated parameters $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{N^*}$, respectively. Since for $q \neq 1$, $Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}_i$ does not have a closed form, we use MH algorithm in order to generate a sample, then

$$E\left(Z\left(\mathbf{s}_0\right)|\mathbf{Z}\right) = \frac{1}{N^*}\sum_{i=1}^{N^*} E\left(Z\left(\boldsymbol{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}_i\right) = \frac{1}{N^*}\sum_{i=1}^{N^*}\frac{1}{N_i}\sum_{j=1}^{N_i} Z_{ij} \tag{16}$$

where Z_{ij} is the jth sample generated from distribution of $Z\left(\mathbf{s}_0\right)|\mathbf{Z}, \boldsymbol{\eta}_i$. Here, we use proposal distribution $g_{Z_0}\left(\mathbf{y}\right)\ :\ N\left(Z_0, b_1^2\right)$ in MH algorithm.

4. SIMULATION STUDY

In order to study the performance of GAL model, a simulation study is performed with calculations done in R. To generate, say 50 realizations, from $Z\left(.\right)$, we used a stationary GAL RF with exponential covariance function $C\left(|\boldsymbol{h}|\right) = \sigma^2\exp\left(-|\boldsymbol{h}|/\theta\right)$ on a regular lattice 500×500 with parameters $\boldsymbol{\beta} = (4, 7), \sigma^2 = 1, \theta = 4$. Other parameters are $\beta_0 = (5, 6), \boldsymbol{\Sigma}_0 = diag\left(2\right), \zeta = 2, \varphi = 1, b = 2$ and $\alpha_0 = 2$. The histogram and P-P plot of simulated data given in Figure 1, shows that data are skewed and not Gaussian. The parameters are firstly estimated by using Gibbs sampler with 5000 iteration. Parameter estimates are $\hat{\boldsymbol{\beta}} = (3.84, 7.32), \hat{\sigma}^2 = 1.76$ and $\hat{\theta} = 4.11$. Plots of convergence for mean of simulated parameters given in Figure 2, show that the Bayes estimators have converged to the real parameters. In the next step, the sample was divided at random into an estimation set of size n and a validation set of size m. We choose $m = 10$ and in this 10 random selected locations (test set), prediction is done by the other 40 locations (training data set).

Comparing the real and estimated values at these locations given in Table 1, shows plausible results. The used sample size N^* in (19) is 60. From Table 1, we conclude that the estimation for q $= 1$ and $\boldsymbol{\Sigma}^*$ is more precise than q $\neq 1$ and $\mathbf{C}^*$. To compare the performance of four used methods in Table 1, the leave-one-out method is used. In this method, we choose $m = 1$ and repeat this process n times for $n - 1$ remaining observations. The prediction mean of square errors (PREMS) are computed by

$$PREMS\left(\hat{Z}_l\right) = \frac{\sum_{i\neq l}\left(\hat{Z}_l - Z_i\right)^2}{n}, \tag{17}$$
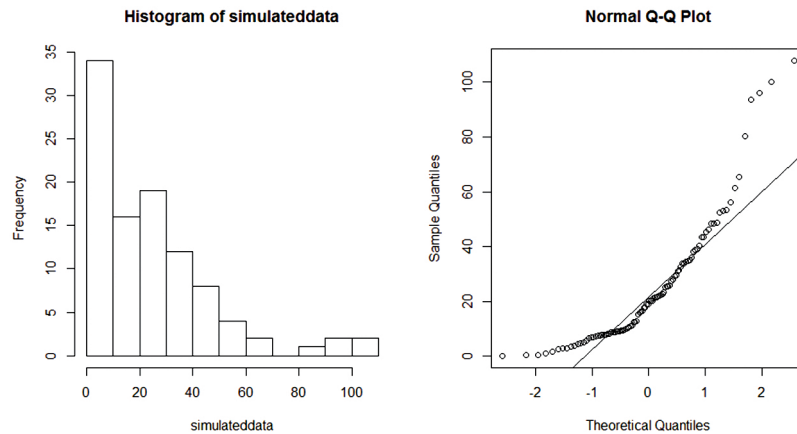
Figure 1 | Histogram and normal Q-Q plot for simulated data show basic similarities with generalized asymmetric Laplace (GAL) distribution with respect to skewness, heavy tail and non-Gaussian.
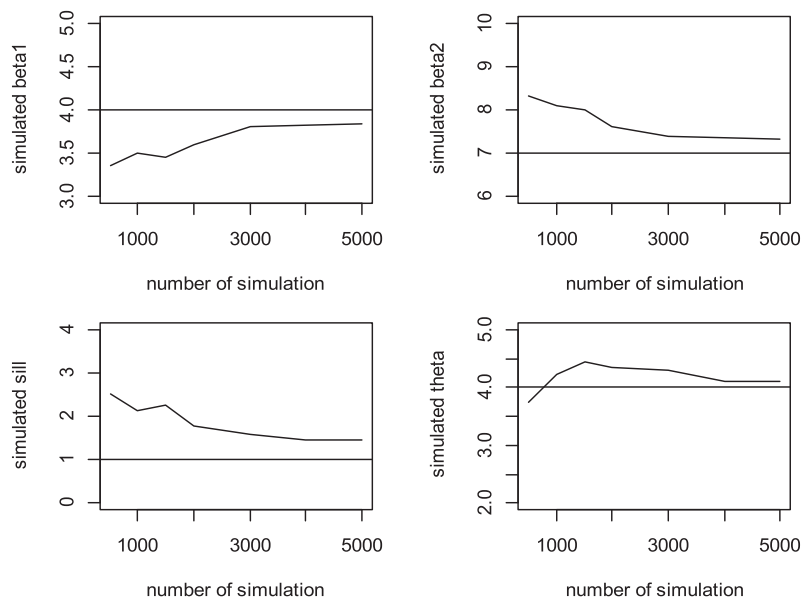

Figure 2 | Plot of convergence for parameters.

and are illustrated in Table 2, for all 4 methods. From the results in Table 2, we observe that the best results are derived by applying $\mathbf{\Sigma}^*$ for $q = 1$. As we expected, the estimation in GAL model for $q = 1$ has less error than GAL model for $q \neq 1$, since for $q = 1$, the conditional expectation $E\left(Z(s_0)|\mathbf{Z}, \boldsymbol{\eta}\right)$ has closed form but for $q \neq 1$ this conditional expectation has to be computed by MCMC algorithms. Also, this comes from Table 2 that more precise predictions are given by using $\mathbf{\Sigma}^*$ instead of $\mathbf{C}^*$. As is mentioned in Note 1, $\mathbf{\Sigma}^*$ properly considers the spatial covariance structure while $\mathbf{C}^*$ does not completely consider this structure.

In order to have a knowledge about sensitivity of prediction with respect to the estimated parameters, we did predicted Z by assuming parameters, $\boldsymbol{\beta}_e = (3, 9)$, $\sigma_e^2 = 2.5$ and $\theta_e = 2.5$ which have remarkable difference with real parameters. Results show that this prediction has a very weak sensitivity to the value of estimated parameters. However, when the difference in assumed parameters and their real values is increasing, the required sample size N^* in (16) has to increase in order to have a good prediction. Sensitivity of prediction on the values of q is very negligible, somewhat conditional expectation in Theorem 4.1 is a very appropriate approximation for the case $q \neq 1$.

5. APPLICATION TO REAL DATA

In this section, we analyze a real data set of 45 metals in 811 locations at a region near to Darab city of Iran. The histogram and Normal Q-Q plot of all metals show the skewness, heavy tail and non-Gaussian behavior of almost all data. However due to limited space we do not

Table 1 Real values in 10 locations and their predictions (prediction error) for different models.

Real Value	q = 1		Real Value	q ≠ 1	
	Σ^*	C^*		Σ^*	C^*
30.54	30.50 (1.07)	30.59 (1.13)	183.33	183.34 (0.71)	183.36 (1.40)
14.40	14.25 (0.72)	14.60 (0.96)	34.11	33.99 (1.90)	33.97 (1.81)
15.30	15.34 (0.11)	16.10 (0.74)	16.14	16.30 (2.13)	15.8 (2.30)
57.81	57.80 (1.12)	57.42 (1.06)	17.52	17.91 (1.17)	17.86 (0.90)
65.02	64.98 (0.36)	65.70 (0.60)	34.08	33.92 (0.04)	35.10 (0.11)
19.60	19.54 (0.85)	20.50 (0.76)	9.47	9.63 (1.86)	9.94 (1.64)
4.41	4.23 (0.21)	4.90(0.50)	84.23	84.29 (2.09)	84.72 (1.83)
5.13	5.18 (1.00)	5.63 (1.27)	15.16	15.63 (0.14)	15.87 (0.27)
58.71	58.46 (.06)	59.40 (0.49)	66.74	66.71 (1.91)	67.13 (2.13)
3.71	3.77 (0.13)	3.80 (0.09)	64.98	65.42 (2.47)	65.48 (2.58)

Table 2 The prediction mean of square errors (PREMS) in simulated generalized asymmetric Laplace (GAL) model for different models.

Model	q = 1		q ≠ 1	
Method	Σ^*	C^*	Σ^*	C^*
PREMS $(\hat{Z})$	0.374	0.560	0.547	0.810
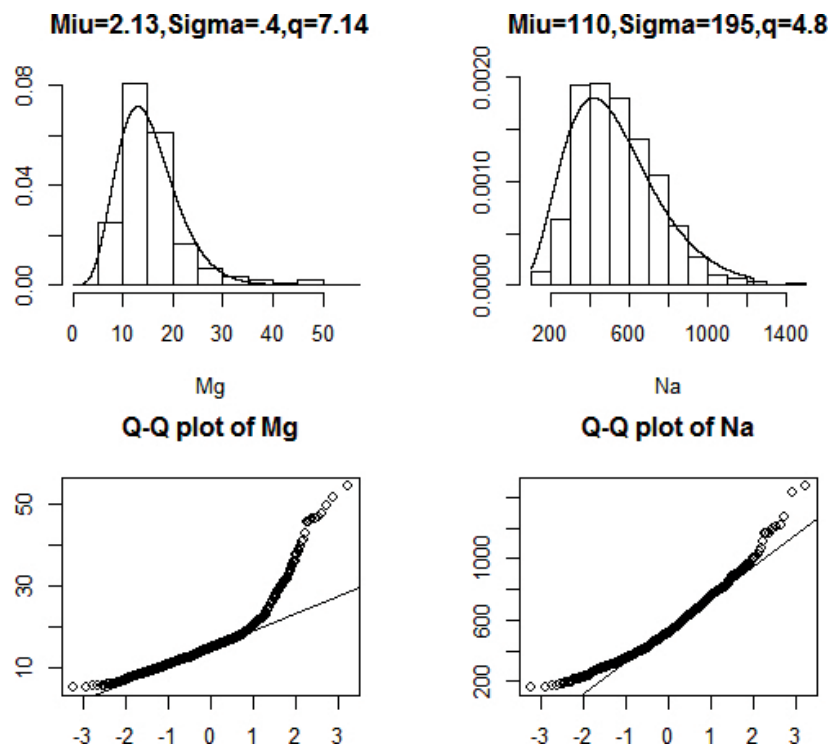
Figure 3 Histogram and normal Q-Q plot for metals Na and Mg.

include all diagrams. The Histogram and Q-Q plot of two metals Sodium (Na) and Magnesium (Mg) have been shown in Figures 3. These two metals are selected because of their similarities to the GAL distribution. Figure 3, shows that GAL density functions has a good fitness to these data, where the parameters are estimated by using the maximum likelihood estimation method. Normal Q-Q plot of data shows that data are non-Gaussian. Also, the small *p*-values of Kolmogorov–Smirnov test and Shapiro–Wilk test for both two data set confirm that data are not Gaussian. In remainder of this section, we only consider Mg data. The scatter plots of data with respect to its coordinates given in Figure 4, show existence of some outliers in data, which we have ignored them for the time being. The scatter plots of 102 remained data show no trend in data.

We use 100 observations as the training set and 2 observations as the test set. Although we have used spatial covariance function throughout the paper, the variogram function is used for selecting a valid model. This comes from the fact that the variogram is more precise in model selection [19]. The empirical variogram of data is plotted in Figure 5.

The isotropic exponential model $\gamma\left(\|h\|\right) = \tau^2 + \sigma^2\left(1 - \exp\left\{-\frac{\|h\|}{\theta}\right\}\right)$ has been fitted to the empirical variogram. The variogram parameters are estimated by using package geoR. The estimates $\sigma^2 = 3, \theta = 12521, \tau^2 = 4.5, p = 7.5$ are used as starting values in MH algorithm. The Gibbs sampler with 5000 iterations in order to estimate parameters is also applied. Final estimated parameters by MCMC method are $\hat{\beta} = (7.2, 7.5), \widehat{\sigma^2} = 3.67$ and $\hat{\theta} = 11719$. In order to compare performance of GAL model with a full Gaussian model, we predicted in a test location by two methods. Table 3 shows surprising results of this comparison.
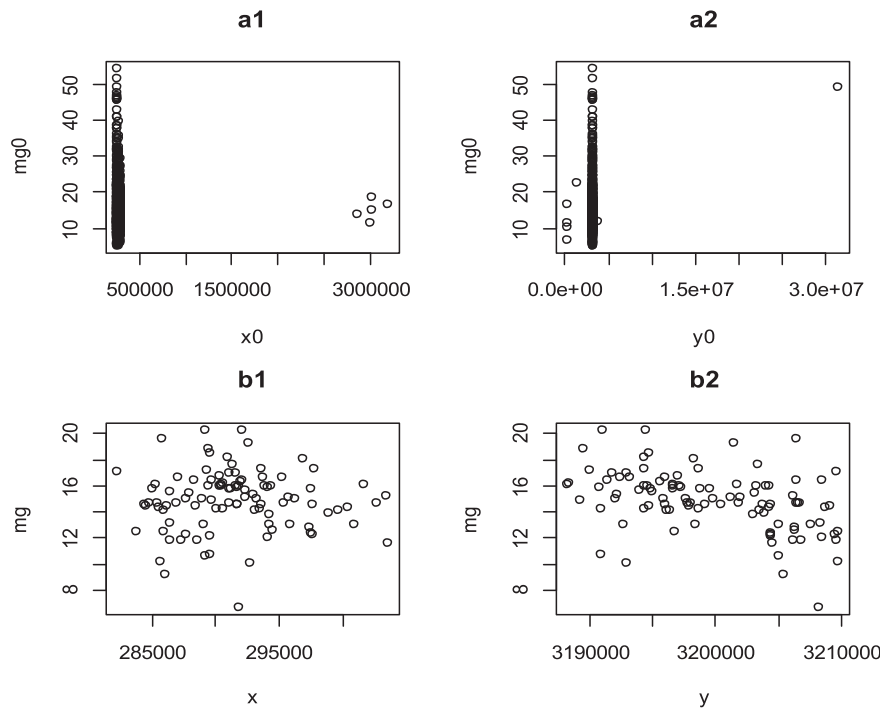

Figure 4 | Panels a1 and a2 are scatter plots of Mg *versus* coordinates of locations and panels b1 and b2 are scatter plots of data with removed outliers.
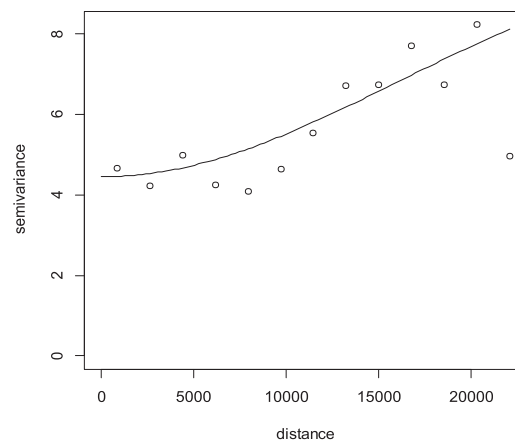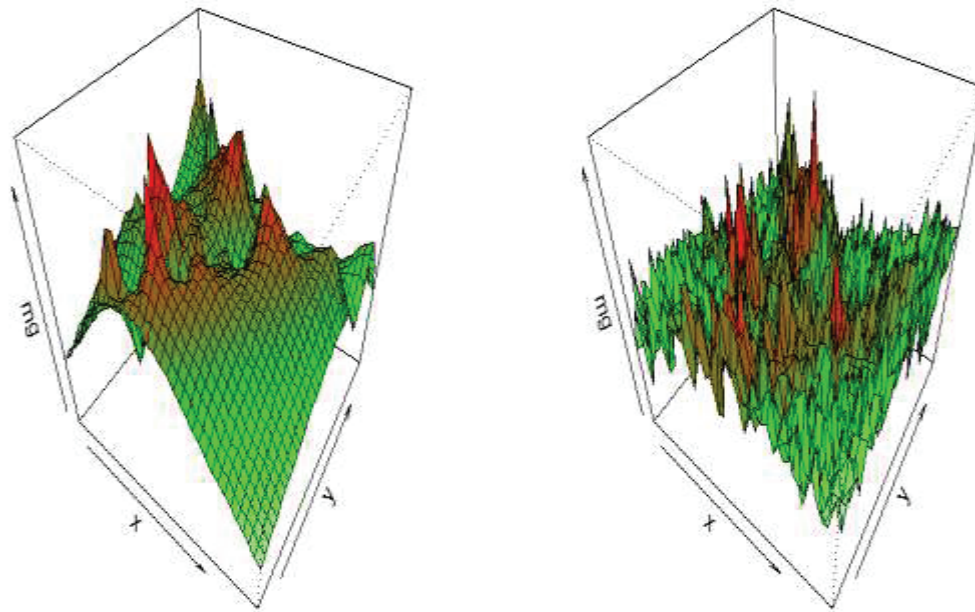

Figure 5 | Empirical variogram and fitted isotropic exponential model.

Table 3 │ Comparison between Kriging and generalized asymmetric Laplace (GAL).

| Real Value | Predicted Value (Standard Deviation) | | | |
| | Kriging | | GAL | |
	Σ^*	C^*	Σ^*	C^*
15.29	15.52, NaN	16.63 (1.86)	15.29 (1.4)	16.21 (3.4)
18.05	19.45, NaN	17.69 (1.87)	17.93 (1.3)	14.74 (2.8)

NaN, refers to not a number.

Figure 6 │ 3D plot of surface prediction based on (left).

This comparison shows the efficiency of GAL model with respect to Kriging. It is remarkable that using Σ^* insteadof C^* has more precise in GAL model. However, as expected, there is no benefit in using Σ^* instead of C^* for Gaussian model. The negative estimate of prediction variance in Gaussian model based on Σ^* shows that this model is not basically well defined. Prediction surfaces in Figure 6 show no smoothness for predictions based on C^*, while prediction based on Σ^*, is more smooth.

Contour graph of surface prediction is shown in Figure 7. Because of having an erratic shape of surface prediction based on C^*, the contour graph only has plotted for the left panel of Figure 6. From Figure 7, we can see spatial structure in predicted value of Mg in whole of region. This point comes from this fact that contour lines with near numbers are in a neighborhood. Surface of prediction error in Figure 8, shows less prediction error for inner locations. This result has a logical justification in statistical point of view. There is more information for prediction in an inner location than a marginal location by considering the number of observations that have basic rule in prediction for mentioned location.

6. DISCUSSION AND CONCLUDING REMARKS

In order to define a RF in terms of the multivariate skew distributions, we propose the multivariate GAL distribution which is shown to be an appropriate choice for modeling skew and heavy tailed data. A Bayesian approach was then used for spatial interpolation. The simulation study and the analysis of a real data set show the reliable performance of this model. The proposed RF can be also used for many spatial purposes such as spatial regression and spatial generalized mixed linear models which have been previously studied by SN RF and CSN RF (see, e.g., Karimi and Mohammadzadeh [11] and Hosseini and Mohammadzadeh [13]).

CONFLICTS OF INTEREST
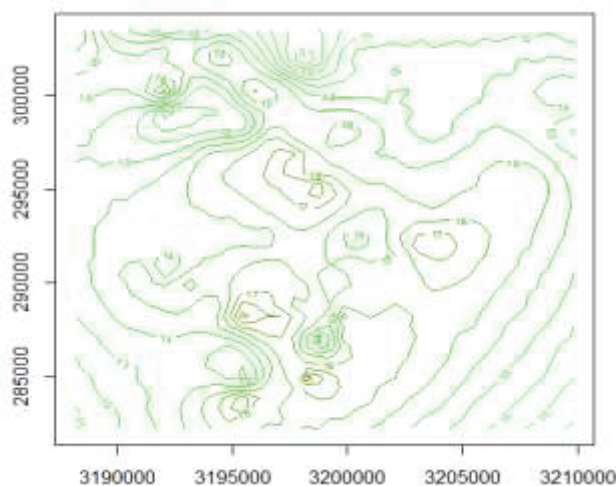
There is no conflicts of interest in this article.

Figure 7 | Contour plot of surface for predictions based on.
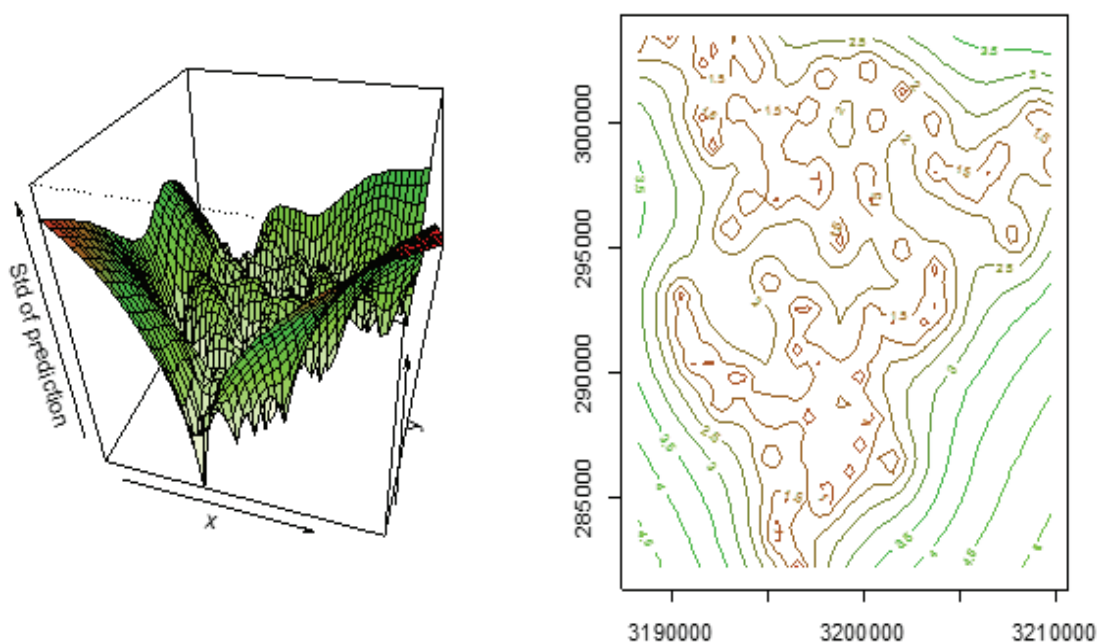

Figure 8 | 3D and contour plot of surface for prediction errors.

AUTHORS' CONTRIBUTIONS

All authors have effective role in Sections 1–4 and 7. However, two other sections which is simulation and application and are concerned with programming in R, is only work of the first author.

Funding Statement

ACKNOWLEDGMENTS

REFERENCES

1. V.D. Oliveira, B. Kedem, D. Short, J. Am. Stat. Assoc. 92 (1997), 1422–1433.
2. V.D. Oliveira, M.D. Ecker, Environmetrics. 13 (2002), 85–101.
3. V.D. Oliveira, Aust. N. Z. J. Stat. 45 (2003), 353–366.
4. A. Azzalini, A. Capitanio, J. R. Stat. Soc. B. 61 (1999), 579–602.
5. H.-M. Kim, B.K. Mallick, J. Stat. Plan Inference. 120 (2004), 85–101.
6. H.-M. Kim, B.K. Mallick, in: A.B. Lawson, D.G.T. Denison (Eds.), Spatial Cluster Modeling, Chapman and Hall/CRC, 2020, pp. 163–173.
7. F. Hosseini, J. Eidsvik, M. Mohammadzadeh, in: Symposium Workshop on Markov Chain-Monte Carlo, Mathematics Institute Warwick, Coventry, UK, 2009. https://warwick.ac.uk/fac/sci/maths/research/events/2008_2009/symposium/wks3/
8. F. Hosseini, J. Eidsvik, M. Mohammadzadeh, CSDA. 55 (2012), 1791–1806.
9. J.A. Dominguez-Molina, G. Gonzales-Farias, A.K. Gupta, in: M.G.Genton (Ed.), Skew-Elliptical Distribution and Their Applications: A Journey Beyond Normality, Chapman and Hall, Boca Rayton, FL, USA, 2004, pp. 25–42.
10. O. Karimi, M. Mohammadzadeh, Math. Geosci. 43 (2011), 565–582.
11. O. Karimi, M. Mohammadzadeh, Stat. Pap. 53 (2012), 205–218.
12. O. Karimi, H. Omre, M. Mohammadzadeh, Geophysics. 75 (2010), 1–11.
13. F. Hosseini, M. Mohammadzadeh, Aust. N. Z. J. Stat. 54 (2012), 43–62.
14. J. Røislien, H. Omre, Math. Geol. 38 (2006), 821–849.
15. T.J. Kozubowski, K. Podgórski, I. Rychlik, J. Multivar. Anal. 113 (2013), 59–72.
16. V. Kalashnikov, Geometric Sums: Bounds for Rare Events with Applications: Risk Analysis, Reliability, Queuing, Springer, Dordrecht, Netherlands, 2010.
17. L.B. Klebanov, T.J. Kozubowski, S.T. Rachev, Ill-Posed Problems in Probability and Stability of Random Sums, Nova Science Publishers, New York, NY, USA, 2006.
18. S. Kotz, T.J. Kozubowski, K. Podgórski, The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance, Birkhäuser, Boston, MA, USA, 2001.
19. N. Cressie, Statistics for Spatial Data, Wiley, New York, NY, USA, 1993.