Generalized Asymmetric Laplace random fields: Existence and Application

M. M. Saber[1], A. R. Nematollahi[2], and M. 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*

*Abstract: S*ome specific random fields have been studied by many researchers whose finite-dimensional marginal distributions are multivariate closed skew-normal or multivariate extended skew-t, in time and spatial domains. In this paper, a necessary and sufficient condition is provided for applicability of such random field in spatial interpolation, based on the marginal distributions. Two deficiencies of the random fields generated by some well-known multivariate distributions are pointed out and in contrast, a suitable skew and heavy tailed random field is proposed. The efficiency of the proposed random field is illustrated through the interpolation of a real data.

Keywords: Random field, Spatial autocorrelation, Multivariate Skew distributions.

1. Introduction

In recent years, random fields (RFs) have been successfully applied, as a statistical model, to the analysis of biological sequences, text and image processing, as well as many areas of computer vision and artificial intelligence. In many situations the availability of a RF is potentially of great importance particularly during defining the joint distribution for any arbitrary number of variables. For instance, a common approach for interpolation in time or spatial domain, is to compute conditional expectation of field variable in interpolated site given observed data. The joint distribution of variables in observed sites and any arbitrary sites is thus needed. In this case a RF usually is defined in terms of the finite-dimensional distributions based on a multivariate distribution. Recently some RFs have been defined by multivariate Skew-Normal (SN) distribution (Azzalini, 1985) [5], multivariate Closed Skew-Normal (CSN) distribution (González-Farías *et al*, 2004) [10], multivariate Skew-t distribution and its general form, multivariate Extended Skew-t (EST) distribution (Arellano-Valle and Genton, 2010) [3]. These RFs have been studied by some authors such as Kim and Mallick (2004) [12] and Karimi and Mohammadzadeh (2011). In this study, we show that the RFs generated by some versions of multivariate SN, CSN and EST distributions have two weakness in application to the spatial interpolation. These two deficiencies are in concern to the compatibility and considering proper spatial autocorrelation structure (PSAS) which will be explained in Section 3. Our findings show that definition of a RF based on multivariate CSN and EST distributions as studied by the corresponding researchers, have these two weaknesses. Although a suitable modification in these RFs may remove the second problem, but the first problem cannot be solved anymore. In the sequel, we define a new skew and heavy tailed RF based on the multivariate generalized asymmetric Laplace (GAL) distribution introduced by Kozubowski et al (2013) [15], which do not have these two deficiencies mentioned in the above RFs. This RF do not have the first weakness at all and have less weakness than the CSN and EST RFs regarded to the second deficiency. Moreover, it provides an enormous flexibility in modeling heavy-tailed and skewed data.

The paper is organized as follows. In Section 2, the multivariate CSN and EST distributions are reviewed. The necessary and sufficient condition for compatibility of a RF is discussed in Section 3. Then, we show that some versions of these distributions do not satisfy this condition. Also, we show these RFs, which have been defined and studied by some researchers, do not consider PSAS property. Section 4 is devoted to definition of GAL RF and study of its profits, and its application to a real environmental data set is included in Section 5.

2. Multivariate closed skew normal and extended skew-t distributions

In this section, the multivariate CSN and EST distributions are first reviewed with emphasis on characteristics of the joint and marginal distributions.

Recall that, a p-dimensional continuous random vector $\boldsymbol{X} = (X_1, \dots, X_p)'$ has a closed skew normal (CSN) distribution, denoted by $\boldsymbol{X} \sim CSN_{p,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{\Delta})$, if its density function is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\Phi_q(0;\xi,\Delta+\Gamma\ \Sigma\Gamma')} \varphi_p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})\ \Phi_q(\boldsymbol{\Gamma}(\boldsymbol{x}-\boldsymbol{\mu});\boldsymbol{\xi},\boldsymbol{\Delta}),\ \boldsymbol{x} \in R^p, \tag{2.1}$$

where $\boldsymbol{\Gamma} \in R^{q \times p}$ is the shape parameter, $\varphi_p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the density function of p-dimensional normal distribution with mean $\boldsymbol{\mu} \in R^p$, positive definite dispersion matrix $\boldsymbol{\Sigma}$, and $\Phi_q(\boldsymbol{x};\boldsymbol{\xi},\boldsymbol{\Delta})$ denotes the q-dimensional normal cumulative distribution function with mean $\boldsymbol{\xi} \in R^p$ and dispersion matrix $\boldsymbol{\Delta}$. Consider the partition $\boldsymbol{X}' = (\boldsymbol{X}_1', \boldsymbol{X}_2')$ with $\dim(\boldsymbol{X}_1) = p_1$, $\dim(\boldsymbol{X}_2) = p_2$, $p_1 + p_2 = p$ and the corresponding partition of the parameters $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma})$. Then the marginal distribution of CSN is given by

$$\boldsymbol{X}_1 \sim CSN_{p_1,q}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\Gamma}_1^*, \boldsymbol{\xi}, \boldsymbol{\Delta}_1^*), \tag{2.2}$$

where $\boldsymbol{\Sigma}_{22.1} = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$, $\boldsymbol{\Gamma}_1^* = \boldsymbol{\Gamma}_1 + \boldsymbol{\Gamma}_2 \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Delta}_1^* = \boldsymbol{\Delta} + \boldsymbol{\Gamma}_2 \boldsymbol{\Sigma}_{22.1}\boldsymbol{\Gamma}_2'$. See González-Farías et al. (2004) [10] for more details.

An extended version of the skew-t distribution is introduced and studied by Arellano-Valle and Genton (2010) [3], see also Adcock (2010) [1]. A continuous p-dimensional random vector $\boldsymbol{X} = (X_1, \dots, X_p)'$ has a multivariate extended skew-t (EST) distribution, denoted by $\boldsymbol{X} \sim EST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda}, \nu, \tau)$, if its density function is given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{T_1\left(\frac{\tau}{\sqrt{1+\lambda'\overline{\Sigma}\lambda}};\nu\right)} t_p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu)T_1((\boldsymbol{\lambda}'\boldsymbol{z}+\tau)(\frac{\nu+p}{\nu+Q(\boldsymbol{z})})^{\frac{1}{2}};\nu+p), \boldsymbol{x} \in R^p, \tag{2.3}$$

where $\overline{\boldsymbol{\Sigma}} = \boldsymbol{\omega}^{-1}\boldsymbol{\Sigma}\ \boldsymbol{\omega}^{-1}$ (the correlation matrix) , $\boldsymbol{z} = \boldsymbol{\omega}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$, $Q(\boldsymbol{z}) = \boldsymbol{z}'\overline{\boldsymbol{\Sigma}}^{-1}\boldsymbol{z}$ and $\boldsymbol{\omega} = diag(\boldsymbol{\Sigma})^{\frac{1}{2}}$, $\boldsymbol{\lambda} \in R^p$ is the shape parameter, $\tau \in R$ is the extension parameter,

$$t_p(\boldsymbol{x};\boldsymbol{\mu},\boldsymbol{\Sigma},\nu) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{|\boldsymbol{\Sigma}|^{\frac{1}{2}}(\nu\pi)^{\frac{p}{2}}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{Q(\boldsymbol{z})}{\nu}\right)^{-\left(\frac{\nu+p}{2}\right)}, \tag{2.4}$$

denotes the density function of usual p-dimensional Student's t distribution with location parameter $\boldsymbol{\mu} \in R^p$, positive definite dispersion matrix $\boldsymbol{\Sigma}$, with p × p scale and correlation matrices $\boldsymbol{\omega}$ and $\overline{\boldsymbol{\Sigma}}$, respectively, and degrees of freedom $\nu > 0$, and $T_1(x;\nu)$ denotes the univariate standard Student's t cumulative distribution function with degrees of freedom $\nu > 0$. The marginal distribution of EST is also given by

$$X_1 \sim EST_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, \boldsymbol{\lambda}_1^*, \nu, \tau_1^*),\tag{2.5}$$

where $\boldsymbol{\lambda}_1^* = \frac{\lambda_1 + \overline{\Sigma}_{11}^{-1}\overline{\Sigma}_{12}\lambda_2}{\sqrt{1+\lambda_2'\tilde{\Sigma}_{22.1}\lambda_2}}, \tau_1^* = \frac{\tau}{\sqrt{1+\lambda_2'\tilde{\Sigma}_{22.1}\lambda_2}}$ and $\tilde{\boldsymbol{\Sigma}}_{22.1} = \overline{\boldsymbol{\Sigma}}_{22} - \overline{\boldsymbol{\Sigma}}_{21}\overline{\boldsymbol{\Sigma}}_{11}^{-1}\overline{\boldsymbol{\Sigma}}_{12}$. See Arellano-Valle and Genton (2010) [3] for more details.

3. Main results

The Kolmogorov extension theorem guarantees that a suitably "consistent" collection of finite-dimensional distributions will define a RF. Let $\nu_{s_1 \dots s_k}(.)$ be a system that satisfy the two following conditions (well known as consistency conditions):

- (Permutation) For all permutations π of $\{1, 2, \dots, k\}$ and measurable sets $B_i \subseteq R$,
$$\nu_{s_{\pi(1)} \dots s_{\pi(k)}}(B_{\pi(1)} \times \dots \times B_{\pi(k)}) = \nu_{s_1 \dots s_k}(B_1 \times \dots \times B_k),\tag{3.1}$$
- (Projection) For all measurable sets $B_i \subseteq R, m \in N$,

$$\nu_{s_1 \dots s_k, s_{k+1}, \dots, s_{k+m}}(B_1 \times \dots \times B_k \times R \times \dots \times R) = \nu_{s_1 \dots s_k}(B_1 \times \dots \times B_k).\tag{3.2}$$

Then, there exists a probability measure space (Ω, F, P) and a RF $\mathcal{Z} = \{Z(\boldsymbol{s}), \boldsymbol{s} \in \boldsymbol{D} \subseteq R^d\}$ for which, $\nu_{s_1 \dots s_k}$ is the finite dimensional distribution (FDD) of RF $\mathcal{Z}$, in the sense that

$$\nu_{s_1 \dots s_k}(\boldsymbol{B}) = P[(Z(\boldsymbol{s}_1), \dots, Z(\boldsymbol{s}_k)) \in \boldsymbol{B}], \quad \boldsymbol{B} \subseteq R^k.\tag{3.3}$$

Here, $\boldsymbol{D}$ is an arbitrary set in R^d. However, it is known that the direct check of the Kolmogorov conditions seems practically impossible. In the case that the system $\nu_{s_1 \dots s_k}(.)$ corresponds to a multivariate function $f_{s_1, \dots, s_k}$ with respect to a σ-finite measure μ in the sense that $\nu_{s_1 \dots s_k}(\boldsymbol{B}) = \int_{\boldsymbol{B}} f_{s_1, \dots, s_k}(z_1, \dots, z_k) \, \mu(dz_1 \dots dz_k)$, an easier alternative is to check the similar conditions with (3.1) and (3.2) for this function. To this end, the following definition is represented.

Definition 3.1 (Compatibility) A multivariate function f is said to be compatible with respect to measure μ, if it is exchangeable in its components, i.e.
$$f_{s_1, \dots, s_k}(z_1, \dots, z_k) = f_{s_{\pi(1)}, \dots, s_{\pi(k)}}(z_{\pi(1)}, \dots, z_{\pi(k)}),\tag{3.4}$$

and if all the marginal functions can be coincided with the ones whose non-common dimensions are integrated out, i.e.
$$\int_{R^m} f_{s_1, \dots, s_k, s_{k+1}, \dots, s_{k+m}}(z_1, \dots, z_{k+m}) \, \mu(dz_{k+1} \dots dz_{k+m}) = f_{s_1, \dots, s_k}(z_1, \dots, z_k).\tag{3.5}$$

The following proposition shows that compatibility is a necessary and sufficient condition to satisfy the consistency conditions in the Kolmogorov existence theorem.

Proposition 3.1. Let system $\nu_{s_1 \dots s_k}$ corresponds to a multivariate positive function f with respect to a σ-finite measure μ ($\nu_{s_1 \dots s_k}(\boldsymbol{B}) = \int_{\boldsymbol{B}} f_{s_1, \dots, s_k}(z_1, \dots, z_k) \, \mu(dz_1 \dots dz_k)$). Then the compatibility of f is a necessary and sufficient condition for which $\nu_{s_1 \dots s_k}$ satisfy consistency conditions in Kolmogorov theorem.

Proof: Since ν corresponds to function f with respect to measure μ, we conclude that
$$\nu_{s_1 \dots s_k}(B_1 \times \dots \times B_k) = \int_{B_1 \times \dots \times B_k} f_{s_1, \dots, s_k}(z_1, \dots, z_k) \, \mu(dz_1 \dots dz_k),$$

and so

$$v_{s_{\pi(1)} \dots s_{\pi(k)}}\left(B_{\pi(1)} \times \dots \times B_{\pi(k)}\right) = \int_{B_{\pi(1)} \times \dots \times B_{\pi(k)}} f_{s_{\pi(1)}, \dots, s_{\pi(k)}}\left(z_{\pi(1)}, \dots, z_{\pi(k)}\right) \mu(dz_{\pi(1)} \dots dz_{\pi(k)})$$

$$= \int_{B_1 \times \dots \times B_k} f_{s_{\pi(1)}, \dots, s_{\pi(k)}}\left(z_{\pi(1)}, \dots, z_{\pi(k)}\right) \mu(dz_1 \dots dz_k).$$

The second equation comes from Fubini's Theorem. Now, the condition (3.1) holds iff (3.4) holds.

Therefore

$$v_{s_1 \dots s_k s_{k+1}, \dots, s_{k+m}}(B_1 \times \dots \times B_k \times R \times \dots \times R) = \int_{B_1 \times \dots \times B_k \times R \times \dots \times R} A \ \mu(dz_1 \dots dz_{k+m})$$

$$= \int_{B_1 \times \dots \times B_k} \left\{ \int_{R \times \dots \times R} A \ \mu(dz_{k+1} \dots dz_{k+m}) \right\} \mu(dz_1 \dots dz_k),$$

where $A = f_{s_1 \dots s_k, s_{k+1}, \dots, s_{k+m}}(z_1, \dots, z_{k+m})$.

Again the second equation is derived by Fubini's Theorem. Therefore, the condition (3.2) is satisfied iff (3.5) is right.

Notice that positivity of f and being σ-finite for measure μ are both essential in Proposition 3.1 and cannot be removed. However, positivity of f can be replaced by integrability of f with respect to measure μ i.e. $\int_{R^{k+m}} f_{s_1 \dots s_k, s_{k+1}, \dots, s_{k+m}}(z_1, \dots, z_{k+m}) \ \mu(dz_1 \dots dz_{k+m}) < \infty$.

Now, we apply Proposition 3.1 to the functions (2.1) and (2.3) in order to show that these two multivariate functions cannot be used for defining an applicable RF to spatial statistics problems. Before stating our original theorem, we explain (in the following remark) a circumstance which both sides of Equation (3.5) may is used separately. The end of this remark is portraying an example which explains why an applicable RF has to be consistent or more clearly, how compatibility of a multivariate function generating a RF will be necessary and essential.

Remark 1. When a random field is defined, we often need to have precise expression of any marginals. For instance in all works which are referenced in the introduction (Kim and Mallick (2004) and Karimi and Mohammadzadeh (2011)), the authors need to have marginal distributions of any arbitrary points in defined random fields. In fact, for spatial prediction studied in these works, it is supposed that the marginal distributions of any arbitrary points with size n is an n-dimensional multivariate CSN distribution where its parameters are completely specified by these n point which do not depend to any other point. For example, consider a CSN random field and let a researcher has 2 observations in locations $\{s_1, s_2\}$ and it is desired to predict his variable in unobserved location s_3. Without miss of generality, suppose that $q = 1$. It is also assumed that the joint distribution of the variables in these 3 locations is

$$(X_{s_1}, X_{s_2}, X_{s_3})^T \sim CSN_{3,1}(\boldsymbol{\mu}_{123}, \boldsymbol{\Sigma}_{123}, \boldsymbol{\Gamma}_{123}, \xi, \Delta) \tag{3.6}$$

where $\boldsymbol{\mu}_{123} = (\mu_1, \mu_2, \mu_3)^T$, $\boldsymbol{\Sigma}_{123} = [\sigma_{ij}]_{i,j=1}^3$ and $\boldsymbol{\Gamma}_{123} = (\Gamma_1, \Gamma_2, \Gamma_3)$. Assume that (3.6) comes from this fact that a CSN random field is considered. From this joint distribution the researcher computes conditional expectation $E(X_{s_3} | X_{s_1}, X_{s_2})$ as the best predictor in location s_3. Now, let another researcher has another observation in location s_4 and therefore should suppose

$$(X_{s_1}, X_{s_2}, X_{s_3}, X_{s_4})^T \sim CSN_{4,1}(\boldsymbol{\mu}_{1234}, \boldsymbol{\Sigma}_{1234}, \boldsymbol{\Gamma}_{1234}, \xi, \Delta). \tag{3.7}$$

Also let this researcher decides to predict variable only by two observations in locations $\{s_1, s_2\}$. Second researcher has two choices for distribution of $(X_{s_1}, X_{s_2}, X_{s_3})^T$. One choice is Equation (3.6) similar with first researcher and another choice is computing distribution of $(X_{s_1}, X_{s_2}, X_{s_3})^T$ from Equation (3.7) and Equation (2.2) which are right-hand side and left-hand side of (3.5), respectively. These two choices have to lead to a unique distribution for $(X_{s_1}, X_{s_2}, X_{s_3})^T$.

The following theorem shows that these two choices given in Remark 1 lead to the various results whenever we need to model skewness using a RF produced based upon functions (2.1) and (2.3).

Theorem 3.1. The multivariate functions (2.1) and (2.3) are compatible if and only if $\Gamma = 0$ and $\lambda = 0$, respectively.

Proof. A simple computation shows that condition (3.4) is satisfied for both functions (2.1) and (2.3). First check condition (3.5) for the function (2.1). Choose $f_{s_1 \dots s_k, s_{k+1}, \dots, s_{k+m}}$ be equal with function (2.1) for $X \sim CSN_{k+m,q}(\mu, \Sigma, \Gamma, \xi, \Delta)$ and consider the partition $X' = (X'_1, X'_2)$ with $\dim(X_1) = k$, $\dim(X_2) = m$ and the corresponding partition of the parameters (μ, Σ, Γ). Also by Note 1, $f_{s_1 \dots s_k}$ has to be equal with function (2.1) for $X_1 \sim CSN_{k,q}(\mu_1, \Sigma_{11}, \Gamma_1, \xi, \Delta)$. By these assumptions, the left-hand side of (3.5) is function (2.1) for $X_1^* \sim CSN_{k,q}(\mu_1, \Sigma_{11}, \Gamma_1^*, \xi, \Delta_1^*)$ where, $\Gamma_1^* = \Gamma_1 + \Gamma_2 \Sigma_{21} \Sigma_{11}^{-1}$, $\Delta_1^* = \Delta + \Gamma_2 \Sigma_{22.1} \Gamma_2'$, $\Sigma_{22.1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$, see (2.2). Therefore (3.5) is satisfied iff $\Gamma_1^* = \Gamma_1$ and $\Delta_1^* = \Delta$ iff $\Gamma_1 + \Gamma_2 \Sigma_{21} \Sigma_{11}^{-1} = \Gamma_1$ and $\Delta + \Gamma_2 \Sigma_{22.1} \Gamma_2' = \Delta$ iff $\Gamma_2 \Sigma_{21} \Sigma_{11}^{-1} = 0$ and $\Gamma_2 \Sigma_{22.1} \Gamma_2' = 0$ ($\Sigma_{22.1}$ is positive definite) iff $\Gamma_2 = 0$. This proves that (3.5) holds for all possible choices of indexes iff $\Gamma = 0$.

The proof about function (2.3) is the same as (2.1). Let $f_{s_1 \dots s_k, s_{k+1}, \dots, s_{k+m}}$ is the same as expression in (2.3) for $X \sim EST_{k+m}(\mu, \Sigma, \lambda, \nu, \tau)$ and consider the partition $X^T = (X_1^T, X_2^T)$ with $\dim(X_1) = k$, $\dim(X_2) = m$ and the corresponding partition of the parameters (μ, Σ, λ). Also by Remark 1, $f_{s_1 \dots s_k}$ has to be equal with function (2.3) for $X_1 \sim EST_k(\mu_1, \Sigma_{11}, \lambda_1, \nu, \tau)$. Also, the left-hand side of (3.5) is function (2.3) for $X_1^* \sim EST_k(\mu_1, \Sigma_{11}, \lambda_1^*, \nu, \tau_1^*)$, where $\lambda_1^* = \frac{\lambda_1 + \bar{\Sigma}_{11}^{-1} \bar{\Sigma}_{12} \lambda_2}{\sqrt{1 + \lambda_2' \tilde{\Sigma}_{22.1} \lambda_2}}$, $\tau_1^* = \frac{\tau}{\sqrt{1 + \lambda_2' \tilde{\Sigma}_{22.1} \lambda_2}}$, $\tilde{\Sigma}_{22.1} = \bar{\Sigma}_{22} - \bar{\Sigma}_{21} \bar{\Sigma}_{11}^{-1} \bar{\Sigma}_{12}$, see (2.5). Therefore (3.5) holds iff $\lambda_1^* = \lambda_1$ and $\tau_1^* = \tau$ iff $\frac{\lambda_1 + \bar{\Sigma}_{11}^{-1} \bar{\Sigma}_{12} \lambda_2}{\sqrt{1 + \lambda_2' \tilde{\Sigma}_{22.1} \lambda_2}} = \lambda_1$ and $\frac{\tau}{\sqrt{1 + \lambda_2' \tilde{\Sigma}_{22.1} \lambda_2}} = \tau$ iff $\lambda_2' \tilde{\Sigma}_{22.1} \lambda_2 = 0$ and $\bar{\Sigma}_{11}^{-1} \bar{\Sigma}_{12} \lambda_2 = 0$ ($\tilde{\Sigma}_{22.1}$ is positive definite) iff $\lambda_2 = 0$. This proves that (3.5) holds for all possible choices of indexes iff $\lambda = 0$.

As we mentioned in introduction, the defined CSN and EST Rfs have another problem in spatial statistics matters especially in Geostatistics. This problem is highlighted in some spatial prediction activities referred in Karimi and Mohammadzadeh (2011) and Karimi et al (2010). In fact, an inappropriate choice for parameters of these RFs have been caused they do not consider PSAS. We explain this characteristic and its deficiency for some RFs in Remark 2 as follows.

Remark 2. Let $Z(s_1), \dots, Z(s_n)$ be the observations from a stationary and isotropic RF $\{Z(s): s \in D \subseteq R^d\}$ at n locations $(s_1, \dots, s_n)$. In Geostatistics literature, it is assumed that there exist a correlation between corresponding variables based on their distance named by spatial autocorrelation. In the other word, $Cov\left(Z(s_i), Z(s_j)\right) = C(|s_i - s_j|)$ where $|h|$ is the norm of vector h and C is a real value function. In a matrix form, $Var(Z) = C$ where $Z = (Z(s_1), \dots, Z(s_n))$ and $C = [C_{i,j}]_{i,j=1}^n$ with notation $C_{i,j} = C(|s_i - s_j|)$ for simplicity. For instance,

when we deal with a Gaussian RF then $\boldsymbol{Z} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $Var(\boldsymbol{Z}) = \boldsymbol{\Sigma}$. Therefore, we choose $\boldsymbol{\Sigma} = \boldsymbol{C}$ in order to have spatial correlation between variables. Another example is t RF which $\boldsymbol{Z} \sim \mathrm{t}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, d)$ ($d \geq 3$) and $Var(\boldsymbol{Z}) = \frac{d}{d-2}\boldsymbol{\Sigma}$ that results in appropriate choice for reaching to PSAS is $\boldsymbol{\Sigma} = \frac{d-2}{d}\boldsymbol{C}$. These examples show that the parameters of a defined RF for applying to problems that have spatial autocorrelation, should be exactly $Var(\boldsymbol{Z}) = \boldsymbol{C}$.

In the following theorem, it is shown that pre-defined RFs based upon the *CSN* and *EST* distributions consider PSAS if and only if they are symmetric.

Theorem 3.2. The random fields generated by multivariate *CSN* and *EST* distributions given by (2.1) and (2.3) consider PSAS if and only if $\boldsymbol{\Gamma} = \boldsymbol{0}$ and $\boldsymbol{\lambda} = \boldsymbol{0}$, respectively, provided that $\boldsymbol{\Sigma}$ is considered as the variance matrix.

Proof. Let $\mathrm{Z}(\boldsymbol{s}_1), \dots, \mathrm{Z}(\boldsymbol{s}_n)$ be the observations from a stationary and isotropic CSN RF $\{\mathrm{Z}(\boldsymbol{s}): \boldsymbol{s} \in \mathbf{D} \subseteq R^d\}$ at n locations $(\boldsymbol{s}_1, \dots, \boldsymbol{s}_n)$ and consider $\mathbf{Z} = (\mathrm{Z}(\boldsymbol{s}_1), \dots, \mathrm{Z}(\boldsymbol{s}_n))$. Therefore, $\boldsymbol{Z} \sim \mathrm{CSN}_{n,q}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\Gamma}, \boldsymbol{\xi}, \boldsymbol{\Delta})$ and

$$Var(\boldsymbol{Z}) = \boldsymbol{\Sigma} + \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\boldsymbol{\Lambda}\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\,\boldsymbol{\Psi}\,\boldsymbol{\Psi}^T\,\boldsymbol{\Gamma}\boldsymbol{\Sigma}, \tag{2.13}$$

where $\boldsymbol{\Lambda}$ and $\boldsymbol{\Psi}$ are complicated matrices based on $\Phi_q(\boldsymbol{0}; \boldsymbol{\xi}, \boldsymbol{\Delta} + \boldsymbol{\Gamma}\,\boldsymbol{\Sigma}\boldsymbol{\Gamma}')$. For more details about these matrices see González-Farías *et al* (2004). PSAS is considered if and only if $Var(\boldsymbol{Z}) = \boldsymbol{C}$, that leads to $\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\boldsymbol{\Lambda}\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\,\boldsymbol{\Psi}\,\boldsymbol{\Psi}^T\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} = \boldsymbol{C}$. However, by assumption $\boldsymbol{\Sigma} = \boldsymbol{C}$, so to preserve PSAS in pre-defined CSN RF, it is required to $\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\boldsymbol{\Lambda}\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\,\boldsymbol{\Psi}\,\boldsymbol{\Psi}^T\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$ or equivalently $\boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T(\boldsymbol{\Lambda} - \boldsymbol{\Psi}\,\boldsymbol{\Psi}^T)\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} = \boldsymbol{0}$. Last equation is satisfied iff $\boldsymbol{\Gamma}\boldsymbol{\Sigma} = \boldsymbol{0}$ since $\boldsymbol{\Lambda} - \boldsymbol{\Psi}\,\boldsymbol{\Psi}^T$ is positive definite. Again $\boldsymbol{\Gamma}\boldsymbol{\Sigma} = \boldsymbol{0}$ iff $\boldsymbol{\Gamma} = \boldsymbol{0}$ by positive definiteness of $\boldsymbol{\Sigma}$. The proof for pre-defined EST RF is exactly similar to the CSN case and so is omitted to sake of space.

In the referenced works such as Karimi and Mohammadzadeh (2011) where a CSN RF has been used, it is supposed that $\boldsymbol{\Sigma} = \boldsymbol{C}$ which does not lead to $Var(\boldsymbol{Z}) = \boldsymbol{C}$ and obviously spatial correlation is not correctly considered in their models.

Although we showed that functions (2.1) and (2.3), which had been already used by some authors as the density function of two well-known multivariate distribution CSN and EST, are invalid in order to define a RF, but this is not the only system of finite dimensional marginal distributions that can be taken into consideration, and it is just a particular case. Indeed, in Minozzo and Ferracuti (2012), according to Bagnato and Minozzo (2014), it is shown that a stationary spatial stochastic process having all its finite dimensional marginal distributions to be CSN.

Also, this was shown that pre-defined CSN and EST distributions cannot consider PSAS. However, the required equation $\boldsymbol{\Sigma} + \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\boldsymbol{\Lambda}\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} - \boldsymbol{\Sigma}\,\boldsymbol{\Gamma}^T\,\boldsymbol{\Psi}\,\boldsymbol{\Psi}^T\,\boldsymbol{\Gamma}\boldsymbol{\Sigma} = \boldsymbol{C}$ for achieving PSAS can be solved numerically for getting valid parameter. The right hand side of this equation is the only known part of it. Therefore, this equation has not unique solution for none of parameters. To cope with this drawback, we need to assume some parameters are known.

In some circumstances, imposing the equality $Var(\boldsymbol{Z}) = \boldsymbol{C}$, without any other constraints or specifications, would imply that a matrix $\boldsymbol{\Sigma}$ which has to be positive definite, which is clearly not possible. However, this remains an open problem that how we can consider PSAS without loss of positive definiteness of $\boldsymbol{\Sigma}$. Indeed for having positive definiteness of $\boldsymbol{\Sigma}$ (or PSAS) declining PSAS (or positive definiteness of $\boldsymbol{\Sigma}$) is basically wrong. A method is to change this matrix as small as possible by converting its negative eigenvalues to small and positive values. This may be done in library *psych* of R software by using the function *cor.smooth* (for details see Revelle, (2015)). This change causes loss of PSAS again. However, our simulation computations along with a real example in Section 5, show that efficiency of this method with respect to choose $Var(\boldsymbol{Z}) = \boldsymbol{\Sigma}$ in application.

Until now, it is shown that the multivariate CSN and EST distributions both violate compatibility condition and cannot consider PSAS. Therefore, the CSN and EST RFs cannot be applied to a wide class of spatial statistics problems. According to Arellano-Valle and Genton (2010), "the EST distribution provides a very flexible class of statistical models. For $\tau = 0$, the multivariate skew-t distribution $ST_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \nu, \boldsymbol{\lambda})$ in the form adopted by Azzalini and Capitanio (2003) is derived. The multivariate skew normal $SN_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ distribution of Azzalini and Dalla Valle (1996) and the skew-Cauchy $SC_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\lambda})$ distribution of Arnold and Beaver (2000) arise when we further let $\nu \to \infty$ and $\nu = 1$, respectively." Therefore, any RF based on these skew multivariate distributions cannot be applied to spatial statistics problems, too. This problem is our motivation to define a well- defined skew RF which is both compatible and consider PSAS. In the next section, a skew RF is constructed by incorporating the multivariate GAL distribution.

4.   The Generalized skew Laplace random fields

In this section, a multivariate skew distribution is proposed for which the compatibility condition (given by (3.4) and (3.5)) is satisfied. Also, this RF consider PSAS.

Definition 4.1 (multivariate generalized asymmetric Laplace distribution, Kozubowski et al. (2013)) [15]. A continuous p-dimensional random vector $\boldsymbol{X}$ has a generalized asymmetric Laplace (GAL) distribution, denoted by $\boldsymbol{X} \sim GAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q)$, if its characteristic function is given by

$$\phi_X(\boldsymbol{t}) = \left(\frac{1}{1 + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\,\boldsymbol{t} - i\,\boldsymbol{t}^T\boldsymbol{\mu}}\right)^q, \qquad \boldsymbol{t} \in R^p, \tag{4.1}$$

where $\boldsymbol{\mu} \in R^p$ is the location parameter and $\boldsymbol{\Sigma}$ is the positive definite dispersion matrix. If matrix Σ is positive-definite, the distribution is truly *p*-dimensional and has a probability density function of the following form

$$f_X(x) = \frac{2\,e^{\mu^T \Sigma^{-1} x}}{(2\pi)^{\frac{p}{2}} \Gamma(q) |\Sigma|^{\frac{1}{2}}} \left(\frac{Q(x)}{\Psi(\Sigma,\mu)}\right)^{q-\frac{p}{2}} K_{q-\frac{p}{2}}(Q(x,\Sigma)\Psi(\mu,\Sigma)),\qquad(4.2)$$

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

Here, we review some useful results about GAL distribution which will be used to prove the existence of GAL RF.

- If $X \sim GAL_p(\mu,\Sigma,q)$ and A be a real matrix $l \times p$. Then,

$$AX \sim GAL_l(A\,\mu, A\,\Sigma\,A^T, q).\qquad(4.3)$$

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

$$X_1 \sim GAL_{p_1}(\mu_1, \Sigma_{11}, q).\qquad(4.4)$$

- By Kotz et al. (2001), the conditional mean finds the following form for the case of q = 1,

$$E(X_2|X_1 = x_1) = \Sigma_{21}\Sigma_{11}^{-1}x_1 + \left(\mu_2 - \Sigma_{21}\Sigma_{11}^{-1}\mu_1\right)\frac{Q(x_1,\Sigma_{11})}{\Psi(\mu_1,\Sigma_{11})} R_{1-\frac{p_1}{2}}\big(\Psi(\mu_1,\Sigma_{11})\,Q(x_1,\Sigma_{11})\big),\qquad(4.5)$$

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

- If $X \sim GAL_p(\mu,\Sigma,q)$ then $E(X) = \mu\,q$ and $Var(X) = q(\Sigma + \mu\,\mu^T)$. Also, we have the following representation for multivariate GAL random variable

$$X = \mu\,G + \sqrt{G}N,\qquad(4.6)$$

  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.

To check the existence of a GAL RF, it is sufficient to check the compatibility of the GAL distribution with respect to Lebesgue measure. For checking (3.4), we use (4.3) with $X^T = (Z(s_1),\dots,Z(s_k))$, $X_\pi^T = (Z(s_{\pi(1)}),\dots,Z(s_{\pi(k)}))$ and an appropriate permutation matrix A such that $AX = X_\pi$. It is clear that $f_X(x) = f_{X_\pi}(x_\pi)$ by using density function (4.2). For checking (3.5), use (4.4) and let $X^T = (Z(s_1),\dots,Z(s_k), Z(s_{k+1}),\dots,Z(s_{k+m}))$, $X_1^T = (Z(s_1),\dots,Z(s_k))$ and $X_2^T = (Z(s_{k+1}),\dots,Z(s_{k+m}))$. Before we define this RF, this is shown that this RF can consider PSAS. Let $Z(s_1),\dots,Z(s_n)$ be the observations from a stationary and isotropic GAL RF $\{Z(s): s \in D \subseteq R^d\}$ at n locations $(s_1,\dots,s_n)$. Then, $Z = (Z(s_1),\dots,Z(s_n)) \sim GAL_n(\mu,\Sigma,q)$ and $Var(Z) = q(\Sigma + \mu\,\mu^T)$. As we mentioned in Note 2, for having PSAS, this is required that $Var(Z) = C$ which leads to $C = q(\Sigma + \mu\,\mu^T)$. Therefore, choosing

$$\Sigma = \frac{1}{q}C - \mu\,\mu^T,\qquad(4.7)$$

results in defined RF use PSAS.

However, positive definiteness of Σ maybe not satisfied for some special choices of C and μ in (4.7). Therefore, a RF $\mathcal{Z} = \{Z(s): s \in D \subseteq R^d\}$ is termed a GAL RF if $Z = (Z(s_1),\dots,Z(s_n)) \sim$

$GAL_n(\boldsymbol{\mu}, \frac{1}{q}\boldsymbol{C} - \boldsymbol{\mu}\,\boldsymbol{\mu}^T, q)$ for all configurations $(\boldsymbol{s}_1, \dots, \boldsymbol{s}_n) \in \boldsymbol{D} \times \dots \times \boldsymbol{D}$ and all $n \in \mathrm{N}$. This RF is well suited for fitting to the skew and heavy tailed data. The ergodicity of this RF is indicated in the following remark.

Remark 3. It is well-known that the stochastic processes must be ergodic as well as stationary in order to establish strong consistency of estimators given in the form of spatial means. It ensure that individual realizations of the process contains sufficient information to produce consistent estimates for the parameters, see e.g. Koopmans (1995). Indeed, ergodicity can be considered as a mixed property of stationarity and asymptotic independence. On the other hands, a RF without PSAS may not poses ergodicity properties.

A sufficient (but not necessary) condition for ergodicity of a stationary stochastic process $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \mathrm{D} \subseteq R^d\}$ is that $C_Z(\boldsymbol{h}) = Cov(Z(\boldsymbol{s}), Z(\boldsymbol{s} + \boldsymbol{h}))$ tends to zero as $||\boldsymbol{h}||$ tends to infinity, see Gaetan and Guyon (2010, Theorem B.2, p.257). Although the covariance function $C_Z(\boldsymbol{h})$ is chosen in such a way that this condition satisfies, lack of PSAS ($Var(\boldsymbol{Z}) \neq \boldsymbol{C}$) which means $cov(Z(\boldsymbol{s}), Z(\boldsymbol{s} + \boldsymbol{h})) \neq C(\boldsymbol{h})$ causes that we cannot conclude $\lim\limits_{||\boldsymbol{h}|| \to +\infty} Cov(Z(\boldsymbol{s}), Z(\boldsymbol{s} + \boldsymbol{h})) = 0$ from $\lim\limits_{||\boldsymbol{h}|| \to +\infty} C(\boldsymbol{h}) = 0$. Therefore, the defined CSN and EST RFs may not be ergodic whenever theirs ergodicity is concluded from $\lim\limits_{||\boldsymbol{h}|| \to +\infty} C(\boldsymbol{h}) = 0$.

For ergodicity of GAL RF $\{Z(\boldsymbol{s}) : \boldsymbol{s} \in \boldsymbol{D} \subseteq R^d\}$, first we need to this RF be stationary. Since $E(Z(\boldsymbol{s})) = q\mu_{\boldsymbol{s}}$, for having stationarty we need to consider vector $\boldsymbol{\mu}$ in (2.14) to be equal to $\boldsymbol{\mu} = (\mu, \dots, \mu)^T$. Therefore, by (4.6) and stationarity we have $Z(\boldsymbol{s}) = \mu Y + \sqrt{Y}\,X(\boldsymbol{s})$ where Y has a standard Gamma distribution with shape parameter q and is independent of $X(\boldsymbol{s})$ which is a zero mean Gaussian process with autocovariance function $C_X(\boldsymbol{h})$. By a straightforward computation we have $C_Z(\boldsymbol{h}) = q(\mu^2 + C_X(\boldsymbol{h}))$. Therefore, by choosing $C_X(\boldsymbol{h}) = \frac{G(\boldsymbol{h})}{q} - \mu^2$, where $G(\boldsymbol{h})$ is a valid covariance function with the property $\lim\limits_{||\boldsymbol{h}|| \to +\infty} G(\boldsymbol{h}) = 0$, one can easily shows that a stationary GAL RF by an appropriate covariance function is ergodic. Notice that matrix $\boldsymbol{\Sigma}$ in (4.7), as the covariance matrix of process $X(\boldsymbol{s})$, is built by $C_X(\boldsymbol{h})$ and matrix $\boldsymbol{C}$ in (4.7) is built by $G(\boldsymbol{h})$, too.

Some properties of the RF proposed are highlighted in the continuation of this section. In following, we illustrate the usefulness of the GAL RF. Figure 1 shows the (one dimensional) GAL density functions for different values of the location parameter μ. It is clear that the GAL distribution is skew to right, symmetric and skew to left, for positive, zero and negative values of μ, respectively. So the location parameter μ controls both location and skewness of the distribution. Figure 2 shows the GAL density functions for different values of σ^2.
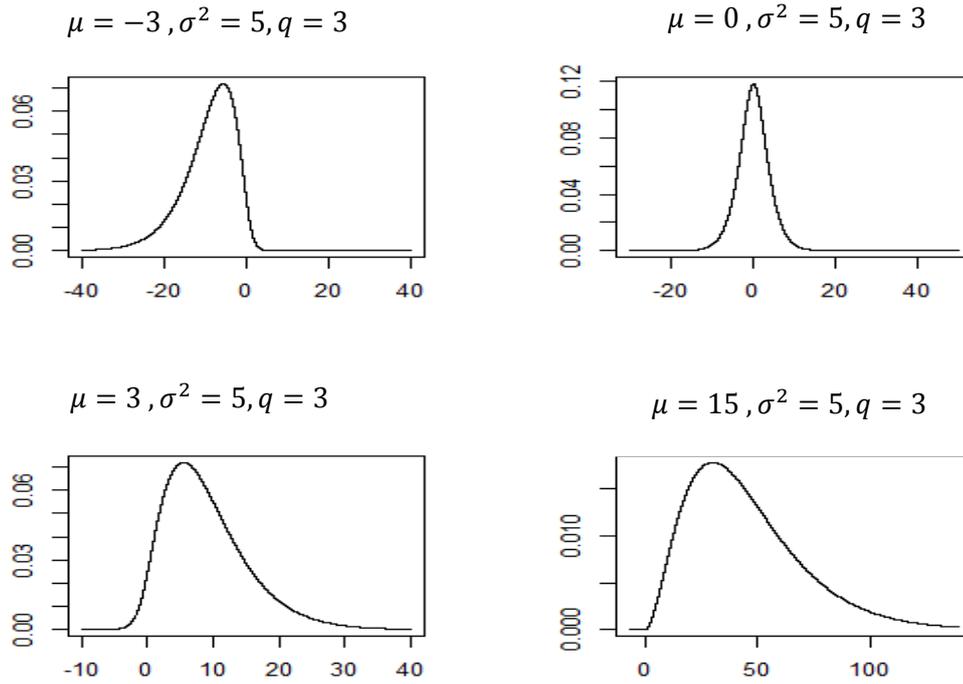
$\mu = -3, \sigma^2 = 5, q = 3$

$\mu = 0, \sigma^2 = 5, q = 3$

$\mu = 3, \sigma^2 = 5, q = 3$

$\mu = 15, \sigma^2 = 5, q = 3$

Fig 1. Plots of GAL density function for different location parameters.

$\mu = 0, \sigma^2 = .5, q = 1$

$\mu = 0, \sigma^2 = 1, q = 1$

$\mu = 0, \sigma^2 = 5, q = 1$
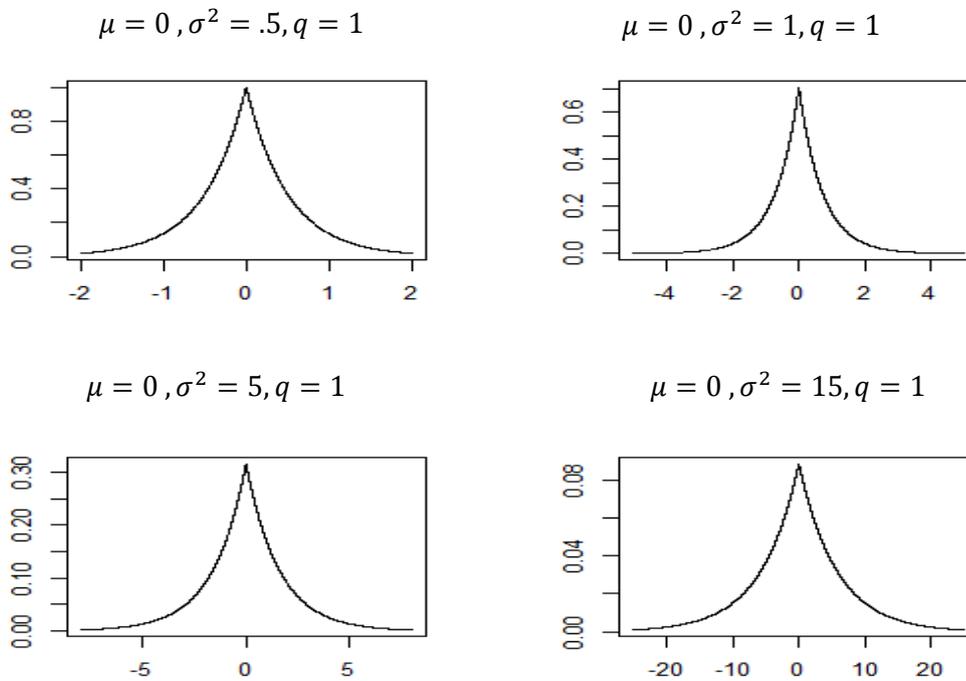
$\mu = 0, \sigma^2 = 15, q = 1$

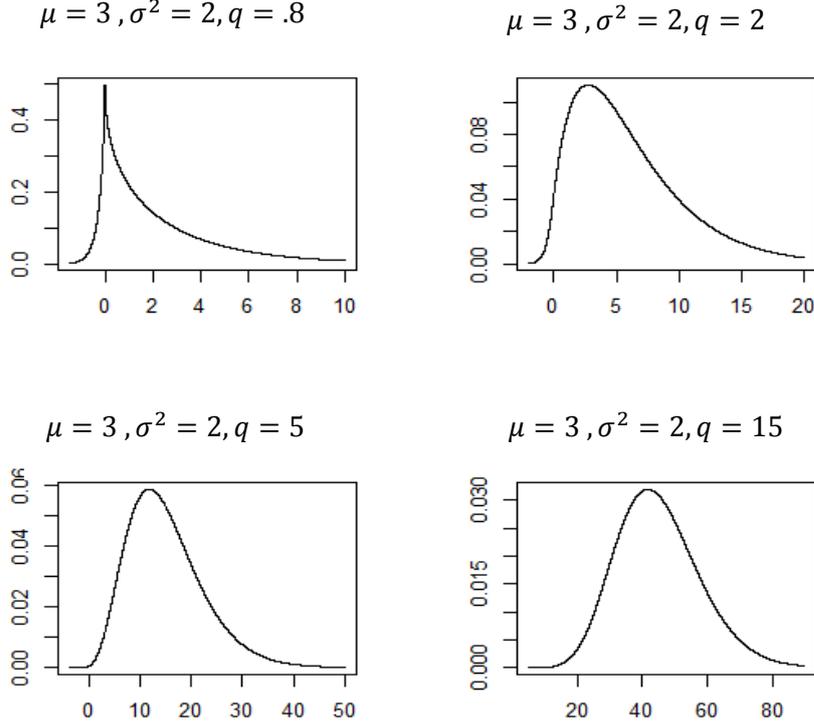Fig 2. Plots of GAL density function for different variance parameters.

Fig 3. Plots of GAL density function for different extension parameters.

It is seen that the large variance leads to the small kurtosis and vice versa. The graphs of the GAL density functions for different values of q is plotted in Figure 3, where the effect of increasing q is demonstrated. We see that with increasing q, both skewness and kurtosis are decreasing.

5. Application to spatial prediction

Let $Z(s_1), \ldots, Z(s_n)$ be the observations from a GAL RF $\{Z(s): s \in D \subseteq R^d\}$ at n locations $(s_1, \ldots, s_n)$. For predicting $Z(s_0)$ at new location s_0, based on the observations $Z = (Z(s_1), \ldots, Z(s_n))$, we define $Z^* = (Z(s_0), Z^T)^T$. Then we have $Z^* \sim GAL_{n+1}(F^*\beta, \Sigma^*, q)$ where $\Sigma^* = \frac{C^*}{q} - F^*\beta \beta^T F^{*T}$, $f(s_0) = (f_1(s_0), \ldots, f_r(s_0))^T$, $F^* = (f(s_0), F^T)^T$, $F = [f_j(s_i)]_{n \times r}$ 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, $c = (C_{0i})_{n \times 1}$, $C = Var(Z)$ and $C_{ij} = Cov(Z(s_i), Z(s_j))$. Now, the best predictor of $Z(s_0)$ based on the square error loss function is given by $E(Z(s_0)|Z)$, that can be computed by conditional distribution of $Z(s_0)|Z$. 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 random field. In application both θ and σ^2 can be estimated in stage of variogram estimation by using of data without considering no distribution for them. Therefore, the only unknown parameter is β which is estimated by maximum likelihood method.

The details for prediction is given in the following remark.

Remark 4. The conditional expectations $E(Z(s_0)|Z)$ for $q = 1$ is provided by Equation (4.5). Suppose $\mathbf{Z}^* = (Z(s_0), \mathbf{Z}^T)^T \sim GAL_{n+1}(\mathbf{F}^*\boldsymbol{\beta}, \boldsymbol{\Sigma}^*, 1)$, then from $\mathbf{F}^*\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{f}^T(s_0) \\ \mathbf{F} \end{pmatrix}\boldsymbol{\beta}$ it can be easily concluded that $\boldsymbol{\mu}_2 = \boldsymbol{f}^T(s_0)\boldsymbol{\beta}$ and $\boldsymbol{\mu}_1 = \mathbf{F}\boldsymbol{\beta}$. Now, this is needed to find suitable matrices $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{11}$. So we have to partition the matrix $\boldsymbol{\Sigma}^*$ as follows

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

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

Therefore, $\boldsymbol{\Sigma}_{12} = \boldsymbol{c} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\boldsymbol{f}(s_0)$ and $\boldsymbol{\Sigma}_{11} = \boldsymbol{C} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T$ which leads to

$$E(Z(s_0)|\mathbf{Z}) = \boldsymbol{l}^T\boldsymbol{\Gamma}^{-1}\mathbf{Z} + (\boldsymbol{f}^T(s_0) - \boldsymbol{l}^T\boldsymbol{\Gamma}^{-1}\mathbf{F})\boldsymbol{\beta}\frac{Q(\mathbf{Z},\boldsymbol{\Gamma})}{\Psi(\mathbf{F}\boldsymbol{\beta},\boldsymbol{\Gamma})}R_{1-\frac{n}{2}}(\psi(\mathbf{F}\boldsymbol{\beta},\boldsymbol{\Gamma})\,Q(\mathbf{Z},\boldsymbol{\Gamma}))$$

where $\boldsymbol{\Gamma} = \boldsymbol{C} - \mathbf{F}\boldsymbol{\beta}\boldsymbol{\beta}^T\mathbf{F}^T$ and $\boldsymbol{l} = \boldsymbol{c} - \mathbf{F}\boldsymbol{\beta}\,\boldsymbol{\beta}^T\boldsymbol{f}^T(s_0)$.

Since for $q \neq 1$, $Z(s_0)|Z$ does not have a closed form, the Metropolis-Hasting algorithm is used in order to generate a sample $Z_1, \ldots, Z_M$ from $Z(s_0)|Z$, where the proposal distribution $g_{Z_0}(\boldsymbol{y})\colon N(Z_0, b_1^2)$ in Metropolis-Hasting algorithm is applied. Then $E(Z(s_0)|\mathbf{Z}) \cong \frac{1}{M}\sum_{j=1}^{M} Z_j$.

In order to study the performance of GAL model, a simulation study is performed with 50 realizations. We used a stationary GAL RF with exponential covariance function $C(|\boldsymbol{h}|) = \sigma^2\exp(-|\boldsymbol{h}|/\theta)$ on a regular lattice 500×500 with parameters $\boldsymbol{\beta} = (4,7), \sigma^2 = 1, \theta = 4$. The histogram and P-P plot of simulated data given in Figure 4, shows that data are skewed and cannot follow the Gaussian law.
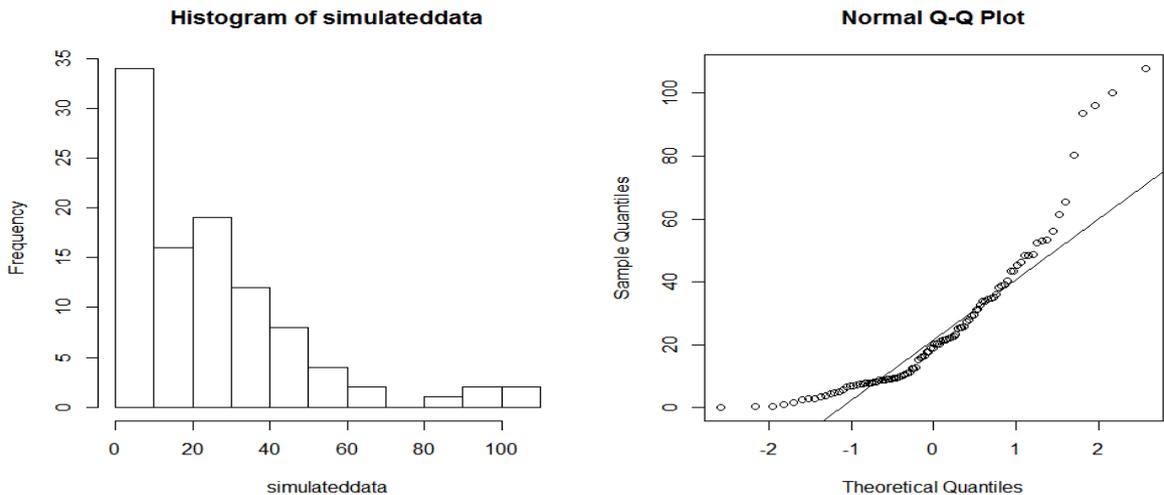

Figure 4. Histogram and Normal Q-Q plot for simulated data show basic similarities with GAL distribution with respect to skewness, heavy tail and non-Gaussian.

Parameter estimates are $\widehat{\boldsymbol{\beta}} = (3.84, 7.32)$, $\hat{\sigma}^2 = 1.76$ and $\hat{\theta} = 4.11$. The parameter estimation has been done for 5000 iteration and values of $\widehat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\hat{\theta}$ are the mean of these estimates. The MSE of estimators are $MSE_{\beta}(\widehat{\boldsymbol{\beta}}) = (0.94, 1.1)$, $MSE_{\sigma^2}(\hat{\sigma}^2) = 2.57$ and $MSE_{\theta}(\hat{\theta}) = 0.28$ which show a suitable estimation has been done. In the next step, the sample was divided at random into an estimation set of size *10* and a validation set of size *40*. 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 Σ^* is more precise than q ≠ 1 and C^*. In the case of q ≠ 1 which Metropolis-Hasting algorithm is applied the parameter $b_1^2 = 2$ has been considered in the proposal distribution.

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)

We finish this section with two examples of real data set which fit to GAL distribution. Data are from Darab region, south of Iran and consist of 45 chemical elements in 811 locations. The histogram and Normal-Q-Q plot of all elements show the skew, heavy tail and non-Gaussian behavior of almost all of them. Histogram and Normal-Q-Q plot of two elements Mg and Na have shown in Figures 5 and 6, respectively. Figure 5, shows that GAL density functions has a good fitness to these data, where the parameters are estimated by using the maximum likelihood estimation method. The details can be found in Saber et al (2014).

The Kolmogorov Smirnov test has done for testing this assumption that data comes from a GAL distribution. The values of Kolmogorov Smirnov's statistic are 0.038 and 0.98 for two element Mg and Na, respectively. These two values will be compared with critical value 0.189 at significance

level 0.05, earned from table of Kolmogorov distribution (Leslie and Miller (1956)). Therefore Mg data follows GAL distribution in level of 0.05 and other one does not follow. The graphs of cumulative distribution function and empirical distribution function for Mg in Figure 7 confirms this fact. We use 100 observations as the training set and 2 observations as the test set. The empirical variogram of data is plotted in Figure 8. The isotropic exponential model $\gamma(\|\boldsymbol{h}\|) = \tau^2 + \sigma^2 \left(1 - \exp\left\{-\frac{\|\boldsymbol{h}\|}{\theta}\right\}\right)$ has been fitted to the empirical variogram. The variogram parameters are estimated by using package geoR. The estimates $\widehat{\sigma^2} = 3.67$, $\hat{\theta} = 11719$, $\widehat{\tau^2} = 4.5$ and $\widehat{\boldsymbol{\beta}} = (7.2, 7.5)$ are earned.

$$\mu = 2.13\,, \sigma^2 = .4, q = 7.14$$

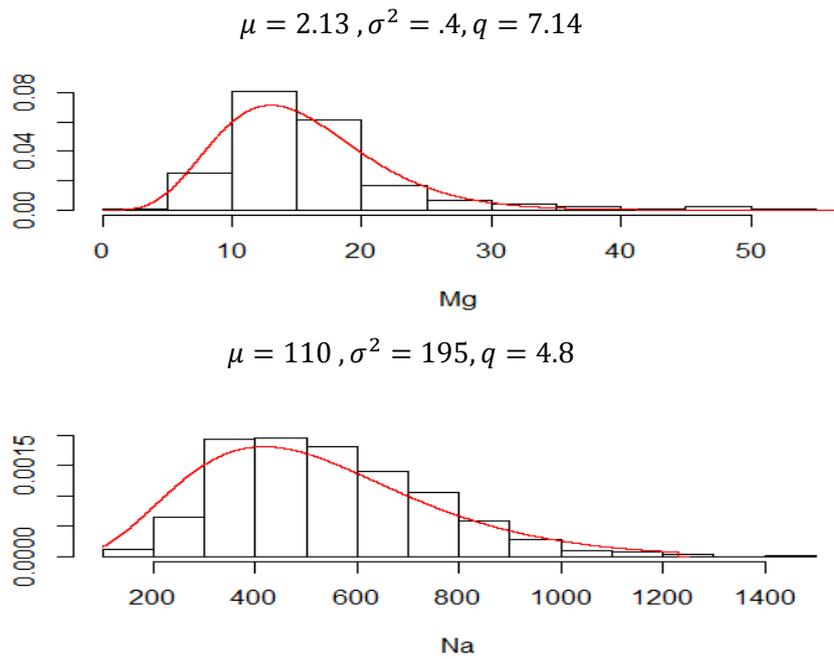$$\mu = 110\,, \sigma^2 = 195, q = 4.8$$

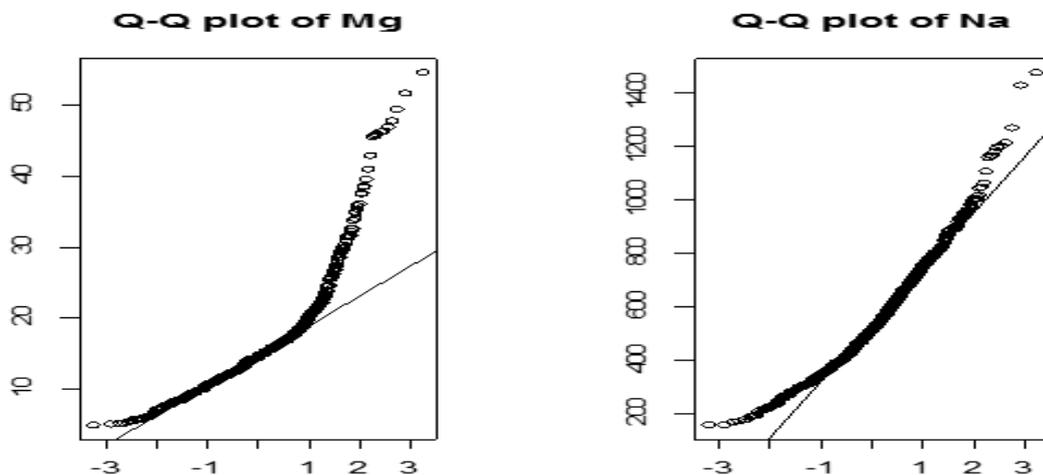Fig 5. Histograms and estimated GAL densities for two elements Mg and Na.

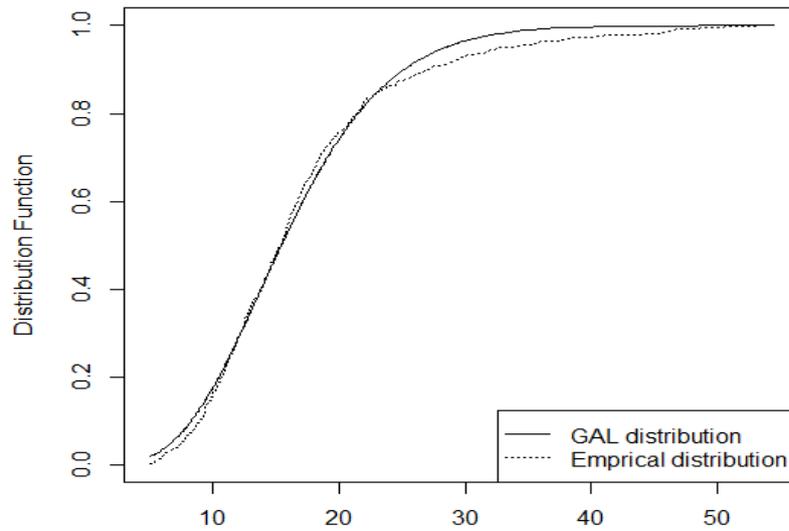Fig 6. Normal Q-Q plot for the data set of Darab region

Fig 7. Comparison between GAL distribution and empirical distribution of Mg.

Figure 8. Empirical variogram and fitted isotropic exponential model.

In order to compare performance of GAL model with a full Gaussian model, we predicted in a test location by two methods. Table 2 shows surprising results of this comparison. This comparison shows the efficiency of GAL model with respect to Kriging. It is remarkable that using Σ^* instead of 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.

Table 2. Comparison between Kriging and GAL (NaN refers to not a number).

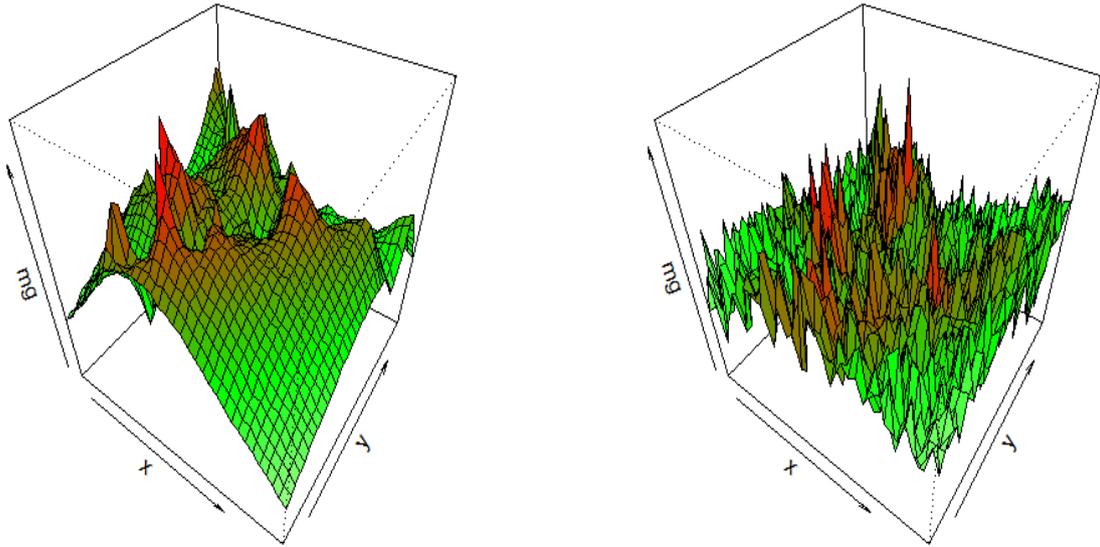| Real value | Predicted value (standard deviation) | | | |
| | Kriging | | GAL | |
	Σ^*	C^*	Σ^*	C^*
15.29	15.52, NaN	16.63 (1.86)	15.29 (1.4)	16.21 (3.4)
18.05	19.45, NaN	17.69 (1.87)	17.93 (1.3)	14.74 (2.8)

Figure 9. 3D plot of surface prediction based on $\mathbf{C}^*$(right) and $\mathbf{\Sigma}^*$ (left).

Prediction surfaces in Figure 9 show no smoothness for predictions based on C^*, while prediction based on $\mathbf{\Sigma}^*$, is more smooth. Contour graph of surface prediction is shown in Figure 10. Because of having an erratic shape of surface prediction based on $\boldsymbol{C}^*$, the contour graph only has plotted for the left panel of Figure 9. From Figure 10, 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.
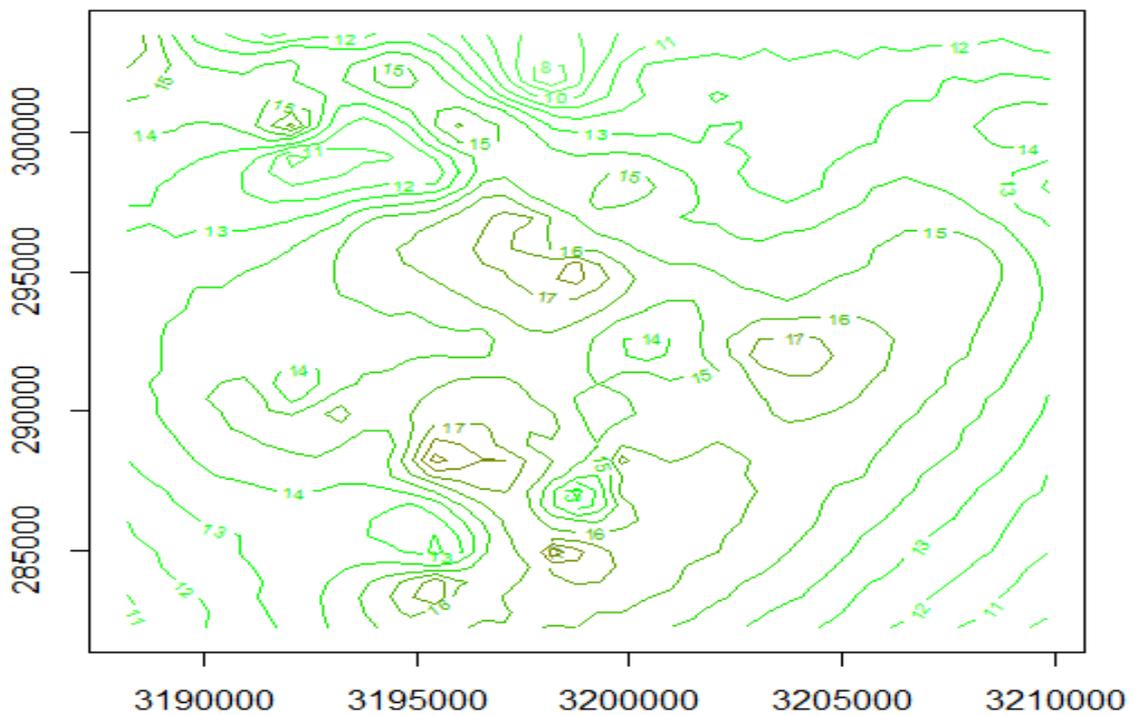

Figure 10. Contour plot of surface for predictions.

6. Discussion and Conclusion

In this paper, we have shown that the multivariate CSN and multivariate EST distributions are not suitable for defining a RF because of their skewness property, (Theorems 3.1and 3.2). In the other words, these kinds of multivariate distributions are applicable to define a RF only in the symmetric case. However, there are some other multivariate distributions that cannot be used for defining a RF even in their symmetric case, e.g. the multivariate Dirichlet distribution and multivariate inverted Dirichlet distribution (Kotz et al, 2000) and the generalized multivariate Dirichlet distribution (Wong, 1998).

It is also easy to show that some well-known multivariate distributions, studied by Kotz et al (2000) are not compatible, e.g. the truncated multivariate normal distribution and truncated multivariate t distributon. On the other hand, however some other multivariate distributions such as Linnik's distribution, Gamma distribution and Logistic distribution are compatible. So far we only studied compatibility of continuous distributions although this concept is about all multivariate distribution. Just like continuous distribution, some discrete multivariate distribution are compatible while some others are not.

Though not proved here, it seems that the similar results about compatibility still hold for more skew elliptical distributions with closed formed marginal distributions. Therefore, a wide class of skew distributions is not applicable to define a RF. This subject is left for future research work.

References

[1]   Adcock, C. J. (2010). Asset pricing and portfolio selection based on the multivariate extended skew-Student-t distribution. *Ann. Oper. Res*, 176, 221–234.

[2]   Allard, D. and Naveau, P. (2007). A new spatial skew-normal random field model. Communications in Statistics: *Theory and Methods*, **36** (9), 1821–1834.

[3]   Arellano-Valle, R. B., and Genton, M. G. (2010). Multivariate extended skew-t distributions and related families. *Metron*, **68**, 201-234.

[4]   Arnold, B. C. and Beaver, R. J. (2000). The skew-Cauchy distribution. *Statistics and Probability Letters*, **49**, 285–290.

[5]   Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171-178.

[6]   Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, **83**, 715-726.

[7]   Azzalini, A. and Capitanio, A. (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t -distribution, *Journal of the Royal Statistical Society Series B*, **65**, 367–389.

[8]   Bagnato, L. and Minozzo, M. (2014). A latent variable approach to modelling multivariate geostatistical skew-normal data. in: M. Carpita et al., eds., *Advances in Latent Variables: Methods, Models and Applications*, Studies in Theoretical and Applied Statistics, Springer, pp. 113–126.

[9]   Gaetan, C. and Guyon, X. (2010). *Spatial Statistics and Modeling*, Springer.

[10] Gonzales-Farias, G., Dominguez-Molina and Gupta, A. K. (2004). The closed skew normal distribution. In M. G. Genton, (ed), skew-elliptical distribution and their applications: A journey beyond normality. Boca Rayton, FL: Chapman and Hall, 25-42.

[11] Karimi, O. and Mohammadzadeh, M. (2011). Bayesian spatial for discrete closed skew Gaussian random fields. *Mathematical Geosciences*, **43**, 565-582.

[12] Karimi, O., Omre, H. and Mohammadzadeh, M. (2010). Bayesian closed-skew Gaussian inversion of seismic AVO data for elastic material properties. *Geophysics*, **75**, 1-11.

[13] Kim, H-M. and Mallick, B. (2004). A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, **120**, 85-101.

[14] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). Continuous multivariate distributions. Volume 1: Models and Applications. New York: Wiley.

[15] Kotz, S., Kozubowski, T. J. and Podgórski, K. (2001). The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, *Economics, Engineering and Finance,* Birkhäuser, Boston.

[16] Kozubowski, T. J., Podgórski, K and Rychlik, I. (2013). Multivariate generalized Laplace distribution and related random fields. *Journal of Multivariate Analysis*, **113**, 59–72.

[17] Leslie, H. and Miller. (1956). Table of percentage points of Kolmogorov statistics. *Journal of the American Statistical Association*, **51**, 111-121.

[18] Minozzo, M. and Ferracuti, L. (2012). On the existence of some skew normal stationary processes. *Chilean Journal of Statistics*, **20**, 2036-2919.

[19] Revelle, W. (2015). psych: Procedures for Personality and Psychological Research. Northwestern University, *Evanston. R package version 1.5.8.*

[20] Saber, M. M., Nematollahi, A. R., and Mohammadzadeh, M. (2014). Bayesian spatial prediction for generalized skew Laplace random field, Technical report, Shiraz, Iran: Shiraz University.

[21] Wong, T. T. (1998). Generalized Dirichlet distribution in Bayesian analysis. *Applied Mathematics and Computation*, **97**, 165-181.