

## Extended Generalized Skew Laplace Random Field: Spatial Autoregressive and Moving Average Model for Prediction of Missing Data in Skew and Heavy Tailed Data

M. M. Saber<sup>1\*</sup>, A. Nematollahi<sup>2</sup>, M. Mohammadzadeh<sup>3</sup>

<sup>1</sup> Department of Statistics, Higher Education Center of Eghlid, Eghlid, Islamic Republic of Iran

<sup>2</sup> Department of Statistics, Faculty of Sciences, Shiraz University, Shiraz, Islamic Republic of Iran

<sup>3</sup> Department of Statistics, Faculty of Sciences, Tarbiat Modares University, Tehran, Islamic Republic of Iran

Received: 30 December 2021 / Revised: 6 April 2021 / Accepted: 26 April 2021

### Abstract

In this paper, we define a spatial skew and heavy-tailed random field by an extended version of multivariate generalized skew Laplace distribution. The Bayesian spatial regression model is developed to explain the spatial data. A simulation study is then carried out to validate and evaluate the performance of the proposed model. The application of this model is also demonstrated in an analysis of a geological real data set.

**Keywords:** Multivariate generalized skew Laplace distribution; SARMA model; MCMC algorithm.

### Introduction

Regression analysis of spatial data is an important statistical method that is frequently used in a number of fields such as agriculture, biology, geology, geography and etc. Some methods for introducing the spatial structure into regression models are presented by [1] [2]. For an application of this model to filtering an image, see [3]. The maximum likelihood estimation procedure to estimate the parameters of these kinds of models is applied by [4]. A necessary condition for the consistency of the maximum likelihood estimator of these models has been investigated by [5]. The Bayesian analysis of regression models with spatially correlated errors and missing observations are studied by [6].

The usual assumption in the regression analysis of spatial data is that data come from a Gaussian Random Field (RF). However, this assumption is often based on the simplicity of the Gaussian structures and does not

hold true for the majority of the applications. In real situations, data are often non-Gaussian but a suitable Normalizing transformation for them exists. But Normalizing transformation is usually unknown and interpretation of the transformed data is also more difficult than the original data as indicated by [8]. [9] used Closed Skew-Normal (CSN) RF for spatial regression with correlated errors and missing data, where the distribution of data has an appropriate number of similarities with Normal distribution but is asymmetric. Although multivariate Extended Skew  $t$  (EST) distribution can be used for this circumstance, [10] showed both CSN and EST distributions have two serious problems for defining an RF. An appropriate choice for modeling the skew and heavy-tailed data is the multivariate Generalized Asymmetric Laplace (GAL) distribution introduced by [11]. [10] have used this distribution to define GAL RF for spatial prediction. However, GAL distribution is not closed

\* Corresponding author: Tel: +989171506741; Fax: +7144526655; Email: mmsaber@eghli.ac.ir

under addition with a constant vector. This weakness does not allow us to use the GAL RF for spatial regression. So a version of multivariate GAL is defined named the multivariate Extended Generalized Asymmetric Laplace (EGAL) distribution to determine an applicable RF. The problem of working with spatial and skew data is not restricted to mentioned models. For instance, [12] have been used GAL RF for spatial prediction and [13] studied paned data models for skew-normal data.

The paper is organized as follows. Section 2 introduces the multivariate EGAL distribution and studies its main properties. The spatial EGAL RF based on the multivariate EGAL distribution is defined in Section 3, where a Bayesian spatial regression (including missing observations) with correlated errors follow from a Spatial Autoregressive and Moving Average (SARMA) model is considered. Section 4 is devoted to the prediction of missing values by using a Bayesian estimation approach including the Monte Carlo Markov Chain (MCMC) procedure to generate a sample from the posterior distributions. A simulation study and application to a real data set are presented in Sections 5 and 6, respectively. Discussion and conclusion remarks are given in Section 7.

**Extended Multivariate Generalized Asymmetric Laplace Random Variable**

In this section, we introduce a multivariate skew distribution named multivariate EGAL distribution as an extension of the multivariate GAL distribution introduced by [11].

**Definition 1.** A continuous  $p$ -dimensional random vector  $\mathbf{X}$  has an EGAL distribution, denoted by  $\mathbf{X} \sim EGAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$ , if its characteristic function is given by

$$\phi(\mathbf{t}) = e^{it^T \mathbf{v}} \left( \frac{1}{1 + \frac{1}{2} t^T \boldsymbol{\Sigma} t - i t^T \boldsymbol{\mu}} \right)^q \quad \mathbf{t} \in R^p, \quad (1)$$

where  $q > 0$  is a generalizing parameter,  $\boldsymbol{\mu} \in R^p$  controls both location and skewness,  $\boldsymbol{\Sigma}$  is a non-negative definite dispersion  $p \times p$  matrix and  $\mathbf{v}$  is the pure location parameter. For  $\mathbf{v} = 0$ , we deal with GAL distribution, denoted by  $\mathbf{X} \sim GAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q)$ . Clearly, if  $\mathbf{X} \sim EGAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$  then  $\mathbf{X} = \mathbf{Y} + \mathbf{v}$ , where  $\mathbf{Y} \sim GAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q)$ . For  $\mathbf{v} = 0$  and  $q = 1$ , it reduces to the multivariate Asymmetric Laplace (AL) distribution, denoted by  $\mathbf{X} \sim AL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , introduced by [14]. When  $q = 1$ , we obtain the multivariate Extended Asymmetric Laplace (EAL) distribution denoted by  $\mathbf{X} \sim EAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{v})$ .

Similar to GAL distribution, if the matrix  $\boldsymbol{\Sigma}$  is positive-definite, the EGAL distribution is truly  $p$ -

dimensional and has a probability density function (pdf) of the form

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{2 e^{\boldsymbol{\mu}^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\mathbf{v})}}{(2\pi)^{\frac{p}{2}} \Gamma(q) |\boldsymbol{\Sigma}|^{\frac{1}{2}}} \left( \frac{Q(\mathbf{x}-\mathbf{v})}{\Psi(\boldsymbol{\mu}, \boldsymbol{\Sigma})} \right)^{q-\frac{p}{2}} K_{q-\frac{p}{2}}(Q(\mathbf{x}-\mathbf{v}, \boldsymbol{\Sigma}) \Psi(\boldsymbol{\mu}, \boldsymbol{\Sigma})), \quad (2)$$

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

If  $\mathbf{X} \sim EGAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$ , then the following representation for EGAL distribution holds

$$\mathbf{X} = \mathbf{v} + \boldsymbol{\mu} G + \sqrt{G} \mathbf{N}, \quad (3)$$

where  $G$  has a standard Gamma distribution with shape parameter  $q$ , is independent of  $\mathbf{N} \sim N_p(0, \boldsymbol{\Sigma})$ , which in turn shows that EGAL distributions are location-scale mixtures of the Normal distribution. Stochastic representation (3) leads to many further properties of GAL random vectors, including moments, marginal and linear transformations. In the following propositions, we study some of them.

**Proposition 1.** Let  $\mathbf{X} \sim EGAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$ . Then, the expectation and variance-covariance matrix of  $\mathbf{X}$  are given by  $E(\mathbf{X}) = \mathbf{v} + q\boldsymbol{\mu}$  and  $Var(\mathbf{X}) = q(\boldsymbol{\Sigma} + \boldsymbol{\mu} \boldsymbol{\mu}^T)$ .

**Proof:** For this end, (3) can be directly used. However, it is more convenient to use the relation between EGAL and GAL;  $\mathbf{X} = \mathbf{Y} + \mathbf{v}$ . The proof is now complete by using the same results for GAL random vector  $\mathbf{Y}$  given by [11]. ■

**Proposition 2.** Let

$\mathbf{X} = (X_1, \dots, X_p) \sim EGAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$ , and let  $\mathbf{A}$  be a real matrix  $\ell \times p$  and  $\mathbf{B}$  be a real vector with dimension  $\ell$ . Then

$$\mathbf{AX} + \mathbf{B} \sim EGAL_{\ell}(\mathbf{A} \boldsymbol{\mu}, \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T, q, \mathbf{A} \mathbf{v} + \mathbf{B}).$$

**Proof:** The characteristic function of  $\mathbf{AX} + \mathbf{B}$  is

$$\begin{aligned} \phi_{\mathbf{AX}+\mathbf{B}}(\mathbf{t}) &= E(e^{i\mathbf{t}^T(\mathbf{AX}+\mathbf{B})}) \\ &= e^{i\mathbf{t}^T \mathbf{B}} \phi_{\mathbf{X}}(\mathbf{A}^T \mathbf{t}) = e^{i\mathbf{t}^T \mathbf{B}} e^{i(\mathbf{A}^T \mathbf{t})^T \mathbf{v}} \left( \frac{1}{1 + \frac{1}{2} (\mathbf{A}^T \mathbf{t})^T \boldsymbol{\Sigma} (\mathbf{A}^T \mathbf{t}) - i (\mathbf{A}^T \mathbf{t})^T \boldsymbol{\mu}} \right)^q \\ &= e^{i\mathbf{t}^T (\mathbf{A} \mathbf{v} + \mathbf{B})} \left( \frac{1}{1 + \frac{1}{2} \mathbf{t}^T \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^T \mathbf{t} - i \mathbf{t}^T \mathbf{A} \boldsymbol{\mu}} \right)^q. \quad \blacksquare \end{aligned}$$

**Proposition 3.** Let  $\mathbf{X} \sim EGAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$  and consider the partition  $\mathbf{X}^T = (\mathbf{X}_1^T, \mathbf{X}_2^T)$  with  $dim(\mathbf{X}_1) = p_1$ ,  $dim(\mathbf{X}_2) = p_2$ ,  $p_1 + p_2 = p$  and the corresponding

partition of the parameters  $(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{v})$ . Then

$$\mathbf{X}_1 \sim EGAL_{p_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, q, \mathbf{v}_1).$$

**Proof:** Set  $\mathbf{B} = \mathbf{0}$  and  $\mathbf{A}$  somehow  $\mathbf{A}\mathbf{X} = \mathbf{X}_1$ , in Proposition 2. ■

Following [12], the conditional mean and variance of multivariate EAL distribution are given in the following proposition.

**Proposition 4.** Let  $\mathbf{X} \sim EAL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{v})$  and consider the same partition stated in Proposition 2. Then

$$E(\mathbf{X}_2 | \mathbf{X}_1) = \mathbf{v}_2 + \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{X}_1 - \mathbf{v}_1)$$

$$+ (\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\mu}_1) \frac{Q(\mathbf{X}_1 - \mathbf{v}_1, \boldsymbol{\Sigma}_{11})}{\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})} R_{1-\frac{p_1}{2}}(\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) Q(\mathbf{X}_1 - \mathbf{v}_1, \boldsymbol{\Sigma}_{11})), \quad (4)$$

$$\begin{aligned} \text{Var}(\mathbf{X}_2 | \mathbf{X}_1) &= \frac{Q(\mathbf{X}_1 - \mathbf{v}_1, \boldsymbol{\Sigma}_{11})}{\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})} (\boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{12}) \\ &\quad \times R_{1-\frac{p_1}{2}}(\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) Q(\mathbf{X}_1 - \mathbf{v}_1, \boldsymbol{\Sigma}_{11})) \\ &\quad + (\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\mu}_1) (\boldsymbol{\mu}_2 - \boldsymbol{\Sigma}_{21} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\mu}_1)^T \\ &\quad \left( \frac{Q(\mathbf{X}_1 - \mathbf{v}_1, \boldsymbol{\Sigma}_{11})}{\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11})} \right)^2 G(\mathbf{X}_1 - \mathbf{v}_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, p_1), \quad (5) \end{aligned}$$

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

$$G(\mathbf{x}_1, \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}, p_1) = R_{1-\frac{p_1}{2}}(\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) Q(\mathbf{x}_1, \boldsymbol{\Sigma}_{11})) R_{2-\frac{p_1}{2}}(\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) Q(\mathbf{x}_1, \boldsymbol{\Sigma}_{11}))$$

$$- \left( R_{1-\frac{p_1}{2}}(\Psi(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) Q(\mathbf{x}_1, \boldsymbol{\Sigma}_{11})) \right)^2.$$

**Proof:** First, note that  $\mathbf{X} = \mathbf{Y} + \mathbf{v}$ , where  $\mathbf{Y} \sim AL_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . By considering the same partition on  $\mathbf{Y}$ , we have  $\mathbf{X}_1 = \mathbf{Y}_1 + \mathbf{v}_1$  and  $\mathbf{X}_2 = \mathbf{Y}_2 + \mathbf{v}_2$ . Therefore,  $E(\mathbf{X}_2 | \mathbf{X}_1) = E(\mathbf{Y}_2 + \mathbf{v}_2 | \mathbf{Y}_1 + \mathbf{v}_1) = E(\mathbf{Y}_2 | \mathbf{Y}_1) + \mathbf{v}_2$ , and  $\text{Var}(\mathbf{X}_2 | \mathbf{X}_1) = \text{Var}(\mathbf{Y}_2 + \mathbf{v}_2 | \mathbf{Y}_1 + \mathbf{v}_1) = \text{Var}(\mathbf{Y}_2 | \mathbf{Y}_1)$ . The Proof is now complete by substituting the corresponding expressions of  $E(\mathbf{Y}_2 | \mathbf{Y}_1)$  and  $\text{Var}(\mathbf{Y}_2 | \mathbf{Y}_1)$  from [14].

### Spatial Regression Model

In the current work, we consider the multivariate EGAL distribution in a spatial setting. Therefore, the spatial EGAL RF is defined based on the multivariate EGAL distribution.

**Definition 2.** A RF  $\{\mathbf{Z}(\mathbf{s}) : \mathbf{s} \in D \subseteq R^d\}$  is termed a EGAL RF if

$\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)) \sim EGAL_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q, \mathbf{v})$  for all configurations  $(\mathbf{s}_1, \dots, \mathbf{s}_n) \in D \times \dots \times D$  and all  $n \in \mathbb{N}$ , where  $\mathbf{v} + q \boldsymbol{\mu} = E(\mathbf{Z})$  and  $\boldsymbol{\Sigma} = \frac{\text{Var}(\mathbf{Z})}{q} - \boldsymbol{\mu} \boldsymbol{\mu}^T$ .

These choices for  $\mathbf{v}$ ,  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are based on Proposition 1. These choices lead to  $\text{Var}(\mathbf{Z}) = \mathbf{C}$ , in defined RF. All the RF properties, especially its stationarity, are entirely similar for GAL RF, which has been proved in [10].

Based on half plan (unilateral) order, [7] studied the following spatial regression model

$$y_{ij} = x_{ij}^T \boldsymbol{\beta} + z_{ij} \quad i = 1, \dots, m, \quad j = 1, \dots, n, \quad (6)$$

on two dimensional regular lattice  $\{(i, j) : i = 1, \dots, m, \quad j = 1, \dots, n\}$ , where  $y_{ij}$  is response variable,  $x_{ij}$  is an  $r$ -dimensional vector of explanatory variables,  $\boldsymbol{\beta}$  is the vector of regression coefficients and  $z_{ij}$ s are auto-correlated random variables follow a first order Spatially Autoregressive and Moving Average (SARMA(1,1)) model. Based on the quarter plan order, [6] considered the regression model (6), when  $z_{ij}$ s are auto-correlated random variables follow a first order Multiplicative Spatial Autoregressive (MSAR(1)) model. They studied a Bayesian approach to the parameter estimation problem of the model for Gaussian data. The same issue for CSN RF is considered by [9]. Based on the quarter plan order, we consider a 2-dimensional lattice for  $z_{ij}$  via the following SARMA(1,1) model

$$z_{ij} = \theta_1 z_{i-1,j} + \theta_2 z_{i,j-1} + \theta_3 z_{i-1,j-1} + \varphi_1 \varepsilon_{i-1,j} + \varphi_2 \varepsilon_{i,j-1} + \varphi_3 \varepsilon_{i-1,j-1} + \varepsilon_{ij}, \quad (7)$$

where  $|\theta_k| < 1$ ,  $|\varphi_k| < 1$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ ,  $k = 1, 2, 3$ .

Let  $\mathbf{Y} = (y_{11}, y_{12}, \dots, y_{mn})^T$ ,  $\mathbf{X} = (x_{11}, x_{12}, \dots, x_{mn})^T$ ,  $\mathbf{z} = (z_{11}, z_{12}, \dots, z_{mn})^T$ ,  $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \dots, \varepsilon_{mn})^T$ ,  $\mathbf{z}_0 = (z_{00}, z_{01}, \dots, z_{m0})^T$  and  $\boldsymbol{\varepsilon}_0 = (\varepsilon_{00}, \varepsilon_{01}, \dots, \varepsilon_{m0})^T$ . Then (3.1) and (7) can be written as the following matrix form

$$\mathbf{Y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{z} \quad (8)$$

$$\mathbf{z} = \mathbf{B}_1 \mathbf{z} + \mathbf{A}_1 \mathbf{z}_0 + \mathbf{B}_2 \boldsymbol{\varepsilon} + \mathbf{A}_2 \boldsymbol{\varepsilon}_0 + \boldsymbol{\varepsilon}, \quad (9)$$

where  $\mathbf{z}_0$  is unobserved primal values vector of  $z_{ij}$ ,  $\mathbf{B}_1$  is a down triangular  $mn \times mn$  matrix,  $\mathbf{A}_1$  is an upper triangular  $mn \times (m+n+1)$  matrix which their components are zero and functions of  $\theta_1$ ,  $\theta_2$  and  $\theta_3$ . The matrices  $\mathbf{A}_2$  and  $\mathbf{B}_2$  are defined similar to  $\mathbf{A}_1$  and  $\mathbf{B}_1$  with replacing  $\theta_1$ ,  $\theta_2$  and  $\theta_3$  by  $\varphi_1$ ,  $\varphi_2$  and  $\varphi_3$ , respectively. Note that (9) can be written as

$$(\mathbf{I} - \mathbf{B}_1) \mathbf{z} = \mathbf{A}_1 \mathbf{z}_0 + (\mathbf{I} + \mathbf{B}_2) \boldsymbol{\varepsilon} + \mathbf{A}_2 \boldsymbol{\varepsilon}_0. \quad (10)$$

Define  $\boldsymbol{\varepsilon}^* = (\boldsymbol{\varepsilon}^T, \boldsymbol{\varepsilon}_0^T)^T$  and replace  $\mathbf{W} = (\mathbf{I} - \mathbf{B}_1)^{-1}$  to lead (10) to

$$\mathbf{z} = \mathbf{W} \mathbf{A}_1 \mathbf{z}_0 + \mathbf{W} \mathbf{D} \boldsymbol{\varepsilon}^*, \quad (11)$$

where  $\mathbf{D} = [\mathbf{I} + \mathbf{B}_2 | \mathbf{A}_2]$ .

**Theorem 1.** Consider the regression model (6) with autocorrelated errors (7) and let  $\boldsymbol{\varepsilon}^* \sim GAL_{N^*}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, q)$ . Then

$$\mathbf{Y} \sim EGAL_{mn}(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y, q, \mathbf{v}_Y),$$

where  $N^* = mn + m + n + 1$ ,  $\boldsymbol{\mu}_Y = \mathbf{W}\mathbf{D}\boldsymbol{\mu}$ ,  $\boldsymbol{\Sigma}_Y = \mathbf{W}\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T\mathbf{W}^T$  and  $\mathbf{v}_Y = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{A}_1\mathbf{z}_0$ .

**Proof.** From (8) and (11), we have  $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{A}_1\mathbf{z}_0 + \mathbf{W}\mathbf{D}\boldsymbol{\varepsilon}^*$ . Now we conclude that  $\mathbf{Y} \sim EGAL_{mn}(\mathbf{W}\mathbf{D}\boldsymbol{\mu}, \mathbf{W}\mathbf{D}\boldsymbol{\Sigma}\mathbf{D}^T\mathbf{W}^T, q, \mathbf{X}\boldsymbol{\beta} + \mathbf{W}\mathbf{A}_1\mathbf{z}_0)$  by Proposition 2.

**Prediction of the Missing Values**

For predicting of the missing values  $\mathbf{Y}_{mis}$  given the observed values  $\mathbf{Y}_{obs}$ , we consider the square loss function. Therefore, the best predictor of  $\mathbf{Y}_{mis}$  given  $\mathbf{Y}_{obs}$  is  $E(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$ . In order to splitting missing values  $\mathbf{Y}_{mis}$  from observed values  $\mathbf{Y}_{obs}$ , suppose  $\mathbf{Y}^* = (\mathbf{Y}_{obs}^T, \mathbf{Y}_{mis}^T)^T = \mathbf{Q}\mathbf{Y}$ , where  $\mathbf{Q}$  is an appropriate orthogonal matrix. By using Proposition 2, we have

$$\mathbf{Y}^* \sim EGAL_{mn}(\mathbf{Q}\boldsymbol{\mu}_Y, \mathbf{Q}\boldsymbol{\Sigma}_Y\mathbf{Q}^T, q, \mathbf{Q}\mathbf{v}_Y),$$

where  $\mathbf{Q}\boldsymbol{\mu}_Y = (\boldsymbol{\mu}_{obs}^T, \boldsymbol{\mu}_{mis}^T)^T$ ,  $\mathbf{Q}\mathbf{v}_Y = (\mathbf{v}_{obs}^T, \mathbf{v}_{mis}^T)^T$  and  $\mathbf{Q}\boldsymbol{\Sigma}_Y = \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{om} \\ \boldsymbol{\Sigma}_{mo} & \boldsymbol{\Sigma}_{mm} \end{bmatrix}$ . In general

the conditional expectation  $E(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$  has not closed form, so we generate samples from the conditional distribution

$$f_{\mathbf{Y}_{mis}|\mathbf{Y}_{obs}}(\cdot|\mathbf{y}_{obs}) = \frac{f_{\mathbf{Y}^*}(\mathbf{y}^*)}{f_{\mathbf{Y}_{obs}}(\mathbf{y}_{obs})}$$

where by Proposition 3,

$$\mathbf{Y}_{obs} \sim EGAL_{N_{obs}}(\boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo}, q, \mathbf{v}_{obs}).$$

This conditional density has not closed form and so the Metropolis-Hastings (MH) algorithm is now applied to generate data from this conditional distribution, where the proposal distribution  $g_{\mathbf{y}_{mis}}(\mathbf{y}): N_{N_{mis}}(\mathbf{y}_{mis}, \text{diag}(b^2))$  is used. For every missing value, we generate a sample with size  $k$ . Mean and variance of these samples are considered as the predicted value and variance of prediction, respectively.

For the case of  $q = 1$ , we have

$$\mathbf{Y}^* \sim EAL_{mn}(\mathbf{Q}\boldsymbol{\mu}_Y, \mathbf{Q}\boldsymbol{\Sigma}_Y, \mathbf{Q}\mathbf{v}_Y). \tag{12}$$

Therefore  $E(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$  and  $Var(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$  are computed by Proposition 4. In the following theorem, the best predictor of  $\mathbf{Y}_{mis}$  given  $\mathbf{Y}_{obs}$  with its variance is given for the special case of  $q = 1$ , when we deal with EAL distributed errors.

**Theorem 2.** Consider the regression model (6) with autocorrelated errors (7) and let  $\boldsymbol{\varepsilon}^* \sim AL_{N^*}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Let the partition  $\mathbf{Y}^T = (\mathbf{Y}_{obs}^T, \mathbf{Y}_{mis}^T)$  with  $\text{dim}(\mathbf{Y}_{obs}) = N_{obs}$ , and the corresponding partition of the parameters

$$(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{v}) \text{ where } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{oo} & \boldsymbol{\Sigma}_{om} \\ \boldsymbol{\Sigma}_{mo} & \boldsymbol{\Sigma}_{mm} \end{bmatrix}.$$

Then

$$E(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}) = \mathbf{v}_{mis} + \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}(\mathbf{Y}_{obs} - \mathbf{v}_{obs}) + (\boldsymbol{\mu}_{mis} - \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\mu}_{obs})$$

$$\frac{Q(\mathbf{Y}_{obs} - \mathbf{v}_{obs}, \boldsymbol{\Sigma}_{oo})}{\Psi(\boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo})} R_{1 - \frac{N_{obs}}{2}}(\Psi(\boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo}) Q(\mathbf{Y}_{obs} - \mathbf{v}_{obs}, \boldsymbol{\Sigma}_{oo})),$$

and

$$Var(\mathbf{Y}_{mis}|\mathbf{Y}_{obs}) = \frac{Q(\mathbf{Y}_{obs} - \mathbf{v}_{obs}, \boldsymbol{\Sigma}_{oo})}{\Psi(\boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo})} (\boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\Sigma}_{om}) \times R_{1 - \frac{N_{obs}}{2}}(\Psi(\boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo}) Q(\mathbf{Y}_{obs} - \mathbf{v}_{obs}, \boldsymbol{\Sigma}_{oo})) + (\boldsymbol{\mu}_{mis} - \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\mu}_{obs}) \times (\boldsymbol{\mu}_{mis} - \boldsymbol{\Sigma}_{mo}\boldsymbol{\Sigma}_{oo}^{-1}\boldsymbol{\mu}_{obs})^T \left( \frac{Q(\mathbf{Y}_{obs} - \mathbf{v}_{obs}, \boldsymbol{\Sigma}_{oo})}{\Psi(\boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo})} \right)^2 G(\mathbf{Y}_{obs} - \mathbf{v}_{obs}, \boldsymbol{\mu}_{obs}, \boldsymbol{\Sigma}_{oo}, N_{obs}).$$

**Proof.** The theorem is simply proved by Proposition 4 and Equation (12).

In applications, the regression coefficients and spatial correlation parameters are unknown and have to be estimated. In this work, a Bayesian approach is used to estimate model parameters. We assume  $\boldsymbol{\mu} = (\alpha_1 J_{N_1}^T, \dots, \alpha_k J_{N_k}^T)^T$  for having  $k$  different skewness where  $\sum_{i=1}^k N_i = N_{obs}$  and  $J_N$  is an  $N$ -dimensional unite vector. Also, let  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}$  and  $q$  is known. By this assumption, we have an appropriate reduction in size of computations. Let  $\boldsymbol{\eta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\alpha}^T, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T)^T$  be vector of unknown parameters, where  $\boldsymbol{\alpha}^T = (\alpha_1, \dots, \alpha_k)$ ,  $\boldsymbol{\theta}^T = (\theta_1, \theta_2, \theta_3)$  and  $\boldsymbol{\varphi}^T = (\varphi_1, \varphi_2, \varphi_3)$ . By considering independence of parameters, the joint prior distribution of  $\boldsymbol{\eta}$  can be written as

$$\pi(\boldsymbol{\eta}) = \pi(\boldsymbol{\beta}) \pi(\sigma) \pi(\boldsymbol{\alpha}) \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\varphi}).$$

Choosing the priors  $\boldsymbol{\beta} \sim N_r(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0)$ ,  $\sigma^2 \sim IG(\lambda_0, \sigma_0)$ ,  $\boldsymbol{\alpha} \sim N_k(\boldsymbol{\alpha}_0, \text{diag}(\tau_0^2))$ ,  $\theta_i \sim TN(-1, 1, \theta_{0i}, \psi_i^2)$  and  $\varphi_i \sim N(\varphi_{0i}, \zeta_i^2)$ , where  $TN(-1, 1, \theta_{0i}, \psi_i^2)$  is a truncated Normal distribution to  $(-1, 1)$  with mean  $\theta_i$  and variance  $\psi_i^2$ , the posterior density of  $\boldsymbol{\eta}$  is then given by

$$\pi(\boldsymbol{\eta}|\mathbf{y}_{obs}) \propto f(\mathbf{y}_{obs}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}).$$

This posterior density has also a complicated form, and so the MCMC method is used to generate a sample from the posterior distribution. To use Gibbs sampler, the derived full conditional distributions of  $\boldsymbol{\xi}$  as an arbitrary component of  $\boldsymbol{\eta}$  are given by

$$\pi(\boldsymbol{\xi}|\mathbf{y}_{obs}, \boldsymbol{\eta}_{-\boldsymbol{\xi}}) \propto f(\mathbf{y}_{obs}|\boldsymbol{\eta})\pi(\boldsymbol{\xi}),$$

where  $\boldsymbol{\eta}_{-\boldsymbol{\xi}}$  is a  $\boldsymbol{\eta}$  which  $\boldsymbol{\xi}$  is deleted from it. These distributions do not have closed forms. For generating data from these densities, MH algorithm in Gibbs sampler is applied. The considered proposal distributions for  $\boldsymbol{\xi} \in \{\boldsymbol{\beta}, \boldsymbol{\alpha}\}$ ,  $\sigma^2$  and  $\delta \in \{\theta_1, \theta_2, \theta_3, \varphi_1, \varphi_2, \varphi_3\}$  are  $g_{\boldsymbol{\xi}}(\cdot): N_r(\boldsymbol{\xi}, \text{diag}(b_{\boldsymbol{\xi}}^2))$ ,

$g_{\sigma^2}(\cdot): \text{Gamma}(\alpha_0, \frac{1}{\sigma^2})$  and  $g_{\delta}(\cdot): \text{TN}(-1, 1, \delta, b_{\delta}^2)$ , respectively. Here  $b_{\xi}^2$ ,  $\alpha_0$  and  $b_{\delta}^2$  are suitable numbers that control efficiency in MH sampling.

**Remark 1.** There is no concern about non-identifiable parameters. The EGAL density function (2) is determined uniquely by its parameters. On the other hand, in Theorem 2 we need to  $\boldsymbol{\mu}_Y$ ,  $\boldsymbol{\Sigma}_Y$  and  $\boldsymbol{v}_Y$  which are identifiable parameters of an EGAL distribution. However, there does not seem to be a guarantee that the parameters  $\boldsymbol{\beta}$ ,  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  are identifiable, which fortunately we do not need to estimate.

### Simulation Study

In order to study the performance of the EGAL model in Bayesian estimation of model parameters and Bayesian prediction, a simulation study is performed with calculations done in R (R Development Core Team, 2003). Also, the accuracy of the model is compared with the Gaussian model. Let  $m = n = 8$ ,  $q = 1$ ,  $k = 2$ ,  $N_1 = 40$ ,  $N_2 = 41$ ,  $\alpha_1 = 2$ ,  $\alpha_2 = 1$ ,  $\boldsymbol{\beta}^T = (4, 7)$ ,  $\sigma^2 = 1$ ,  $\theta_1 = 0.4$ ,  $\theta_2 = 0.8$ ,  $\theta_3 = 0.9$ ,  $\varphi_1 = 0.5$ ,  $\varphi_2 = 0.6$ ,  $\varphi_3 = 0.8$ ,  $\boldsymbol{x}_{ij}^T = (1, i)$  and  $\boldsymbol{z}_0$  are  $m + n + 1$  random samples with standard Gaussian distribution. After generating  $\boldsymbol{\varepsilon}^*$  from  $GAL_{81}(\boldsymbol{\mu}, \boldsymbol{I}, 1)$ , we generate  $\boldsymbol{z}$  by (3.6) and then  $\boldsymbol{Y}$  by (8). The nine random locations are chosen for missing values  $\boldsymbol{y}_{mis} = (y_3, y_{14}, y_{27}, y_{38}, y_{44}, y_{47}, y_{51}, y_{59}, y_{61})$ . The histogram and Q-Q plot of simulated data given in Figure 1, show that data are skewed and not Gaussian. The small  $p$ -values of the Kolmogorov-Smirnov test and the Shapiro-Wilk test for simulated data confirm that data are not Gaussian even at the level of 0.01. The values of skewness and kurtosis coefficients are 1.14 and -0.89,

respectively.

The parameters  $\boldsymbol{\eta} = (\boldsymbol{\beta}^T, \sigma^2, \boldsymbol{\alpha}^T, \boldsymbol{\theta}^T, \boldsymbol{\varphi}^T)^T$  are firstly estimated by using Gibbs sampler with iteration 20000. Parameter estimates are  $\hat{\boldsymbol{\beta}} = (3.94, 7.02)$ ,  $\hat{\sigma}^2 = 0.86$ ,  $\hat{\alpha}_1 = 1.91$ ,  $\hat{\alpha}_2 = 0.98$ ,  $\hat{\theta}_1 = 0.39$ ,  $\hat{\theta}_2 = 0.63$ ,  $\hat{\theta}_3 = 0.84$ ,  $\hat{\varphi}_1 = 0.61$ ,  $\hat{\varphi}_2 = 0.64$  and  $\hat{\varphi}_3 = 0.76$ .

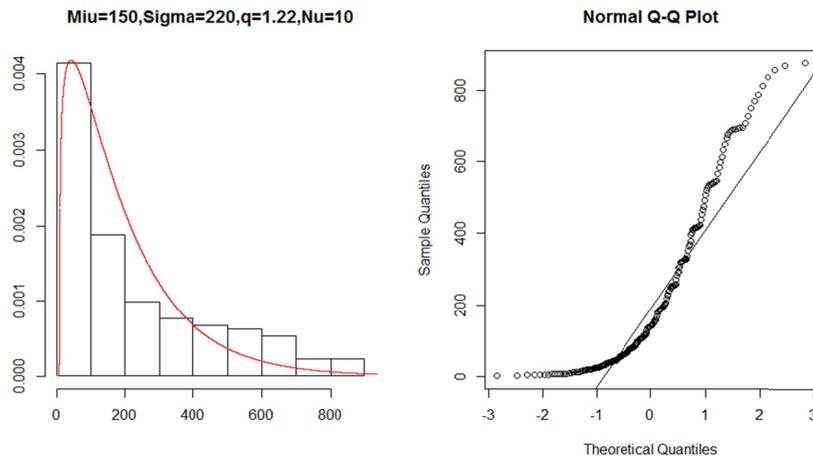
The plots of convergence for the mean of the simulated parameters given in Figure 2 show that the Bayes estimators are converging to the real parameters. Besides, the Gellman-Rubin test, as a convergence criterion, was accomplished for all parameters. Table 1 shows that all test statistic values were approximately close to 1, which confirm the convergence of the MCMC algorithms.

In the second stage, we are going to predict the missing values by two EGAL and Gaussian models. For the Gaussian model, (6)-(11) remain valid. The only difference is in the assumption of distribution errors which is  $\boldsymbol{\varepsilon}^* \sim N_N(\mathbf{0}, \boldsymbol{\Sigma})$ . Therefore,

$$\begin{aligned} \boldsymbol{Y}^* &\sim N_{mn}(\boldsymbol{Q}\boldsymbol{v}_Y, \boldsymbol{Q}\boldsymbol{\Sigma}_Y), \\ E(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}) &= \boldsymbol{v}_{mis} + \boldsymbol{\Sigma}_{mo} \boldsymbol{\Sigma}_{oo}^{-1} (\boldsymbol{Y}_{obs} - \boldsymbol{v}_{obs}) \\ &\text{and} \\ \text{Var}(\boldsymbol{Y}_{mis}|\boldsymbol{Y}_{obs}) &= \boldsymbol{\Sigma}_{mm} - \boldsymbol{\Sigma}_{mo} \boldsymbol{\Sigma}_{oo}^{-1} \boldsymbol{\Sigma}_{om}. \end{aligned} \quad (13)$$

The Bayesian prediction of missing values and standard deviations of predictions by two EGAL, when  $q = 1$ , and Gaussian models are shown in Table 2. The results show that EGAL model is more efficient than the Gaussian model, however the predicted values for two models are very near together. The Prediction Mean Square Errors (PREMS) computed by

$$\text{PREMS}(\hat{\boldsymbol{Y}}_{mis}) = \frac{\sum_{i=1}^{N_{mis}} (\hat{\boldsymbol{Y}}_{mis,i} - \boldsymbol{Y}_{mis,i})^2}{n}, \quad \text{are illustrated in Table 2.}$$



**Figure 1.** Histogram and Normal Q-Q plot for simulated data.

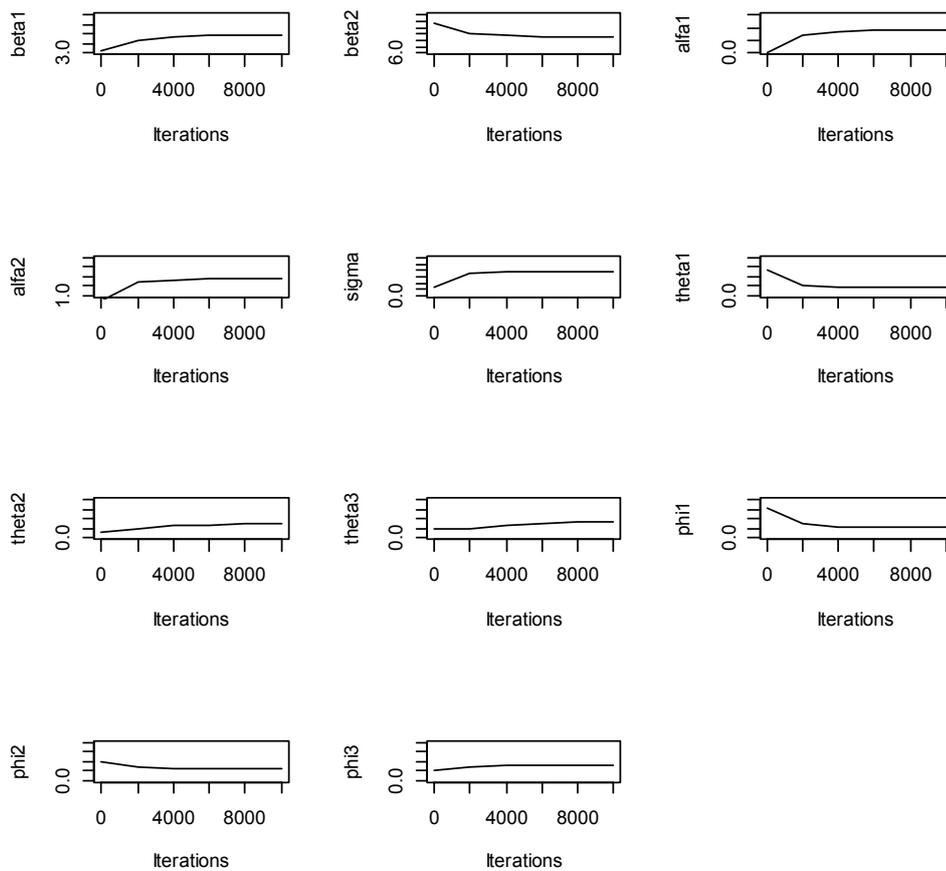


Figure 2. The plot of convergence for parameters

Table 1. The Gellman-Rubin statistics for convergence test.

Estimator	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\sigma}^2$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$
Test's statistics	1.18	1.29	1.07	1.02	1.05	1.12	1.12	1.09	1.25	1.14	1.004

The results show that not only the variances of predictions by the EGAL model are less than the Gaussian model but also the prediction values by the EGAL model are nearer to real values than these values for the Gaussian model. The prediction by the EGAL model for the case  $q \neq 1$  cannot be done by Theorem 1, as we mentioned in the previous section. Therefore, we do simulate with  $q = 3$  and let all other parameters be similar to the case of  $q = 1$ . The same results hold as the case  $q = 1$  for the histogram, Q-Q plot, and Bayesian estimation of parameters. Bayesian prediction of missing values and standard deviations of predictions for two models EGAL when  $q = 2$  and Gaussian have shown in Table 3. The results are approximately similar to the case of  $q = 1$ .

As we expected, the estimation in the EGAL model for  $q = 1$  has less error than EGAL model for

$q \neq 1$ . Indeed, for  $q = 1$ , the conditional expectation  $E(\mathbf{Y}_{mis}|\mathbf{Y}_{obs})$  has a closed-form but for  $q \neq 1$  this conditional expectation has to be computed by MCMC algorithms. We finish this section by noting the following immediate consequence of simulation.

**Note 1.** The computational time in this case, is considerably more than the case  $q = 1$ , since the prediction by EGAL model for the case  $q \neq 1$  cannot be done by Theorem 1 and have to be done by the MCMC method. The results show that for  $q \in (0,1)$ , the prediction is even better than of prediction obtained by the MCMC and is very exact. For  $q \in (1,5)$ , the results are approximately good. Table 4 shows the maximum PREMS for prediction by Theorem 1 for some different values of  $q$ . Simulation has been done 10 times in 9 locations that had been randomly selected.

**Note 2.** In order to have knowledge about sensitivity

**Table 2.** Real values, predictions and variances for two models Gaussian and EGAL for  $q = 1$ .

Real value	EGAL		Gaussian	
	Prediction	Variance	Prediction	Variance
<b>10.462</b>	10.450	0.009	11.139	1.602
<b>16.187</b>	16.112	0.008	16.007	3.165
<b>27.778</b>	27.737	0.004	27.731	3.971
<b>33.995</b>	33.984	0.009	33.967	4.511
<b>39.744</b>	39.739	0.008	39.737	4.312
<b>39.824</b>	39.911	0.009	39.921	4.561
<b>45.405</b>	45.398	0.001	45.386	3.840
<b>51.196</b>	51.195	0.011	51.304	2.123
<b>51.833</b>	51.799	0.011	51.788	4.028
<b>PREMS</b>	0.00278		0.05747	

**Table 3.** Real values, predictions and prediction error for two models Gaussian and EGAL for  $q \neq 1$ .

Real value	EGAL		Gaussian	
	Prediction	Variance	Prediction	Variance
<b>23.512</b>	21.189	2.530	20.517	1.602
<b>34.066</b>	34.518	2.037	34.191	3.465
<b>54.683</b>	56.515	1.965	56.349	3.971
<b>64.802</b>	64.219	2.842	63.319	4.511
<b>62.923</b>	64.721	2.089	64.413	4.312
<b>68.105</b>	68.061	2.150	67.774	4.561
<b>61.452</b>	60.728	2.170	60.672	3.840
<b>67.394</b>	67.256	3.732	65.830	4.503
<b>74.030</b>	73.128	2.157	71.867	4.128
<b>PREMS</b>	1.543		2.670	

**Table 4.** PREMS for some values  $q \neq 1$  by (12)

$q$	<b>0.1</b>	<b>0.5</b>	<b>0.9</b>	<b>4</b>	<b>10</b>
<b>PREMS</b>	0.00014	0.02887	1.75417	4.82543	11.35119

of prediction with respect to the estimated parameters, we predicted  $Y_{mis}$  by assuming parameters which have remarkable difference with real parameters. The results, not given here for the reason of space, show that this prediction has a very weak sensitivity to the values of estimated parameters.

**Application**

In this section, we briefly describe the analysis of a geological real data set. The data consist of 45 chemical elements in 110 locations in a region near Darab city of Iran which has been shown in Figure 3. The histogram and Normal Q-Q plot of all elements show the skewness, heavy tail, and non-Gaussian behavior of the data. However, due to the limited space, we do not include all diagrams. An element that has remarkable similarities with EGAL distribution is Barium (Ba) which has been measured in ppm. In order to predict this element over the whole of the region, we consider a  $15 \times 15$  regular lattice on the region which has been shown in Figure 4. The Histogram and Q-Q plot of data

given in Figures 5, show that the data are really non-Gaussian. The skewness and kurtosis coefficients are 1.7 and 0.51, respectively. Also, the small  $p$ -values of the Kolmogorov-Smirnov test and the Shapiro-Wilk test confirm that the data are not Gaussian. Figure 5 shows that the EGAL density function has good fitness to data, where the parameters are estimated by using the maximum likelihood estimation method. The scatter plot of data given in Figure 5 also shows the possibility of a harmonic trend in data.

By using regression based on its coordinates, we see a trend in data in the form of  $Ba_{(s_1, s_2)} \sim L(s_1, s_2)$ , where  $L(s_1, s_2) = 5.6e + 14(s_1^{-2}) + 5.5e + 15(s_2^{-1.6})$ . The data are detrended.

by  $Ba_{ij}^* = Ba_{ij} - L(s_i, s_j)$ . Consider  $q = 1.3, k = 2, N_1 = 110$  and  $N_2 = 115$ . Now, (8) is written in the modified form  $Y = J_{225} \beta_0 + z$ , and other equations will be changed according to (13). We consider an observation on lattice if its distance with lattice is less than 0.1 lattice width. Here it is about 440. By this assumption, we have 13 observed values on a lattice.

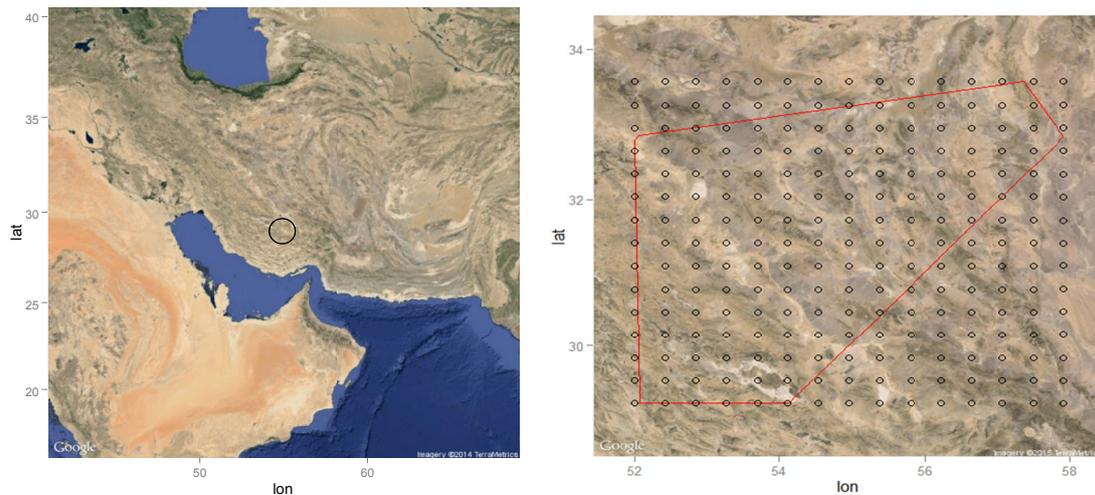


Figure 3. The geographical position of the studied region and considered lattice.

The Gibbs sampler with 10000 iterations in order to estimate parameters is also applied. Estimated parameters by MCMC method are  $\hat{\beta}_0 = 15200$ ,  $\hat{\sigma}^2 = 19.4$ ,  $\hat{\alpha}_1 = 11$ ,  $\hat{\alpha}_2 = 6.3$ ,  $\hat{\theta}_1 = 0.73$ ,  $\hat{\theta}_2 = 0.61$ ,  $\hat{\theta}_3 = 0.36$ ,  $\hat{\varphi}_1 = 0.15$ ,  $\hat{\varphi}_2 = 0.54$  and  $\hat{\varphi}_3 = 0.93$ . Because  $q$  is 1.3, by Note 1, we did prediction by Theorem 1. The graph of surface prediction and its contour plot is shown in Figures 6 and 7. From Figure 7, one can see the spatial structure in the predicted value of

Ba in the whole of the region. This point comes from the fact that contour lines with near numbers are in a neighborhood. Figure 8 shows less prediction error for inner locations. This result has a logical justification from the statistical point of view. There is more information for prediction in an inner location than a marginal site by considering the number of observations with a basic rule in prediction at the mentioned location.

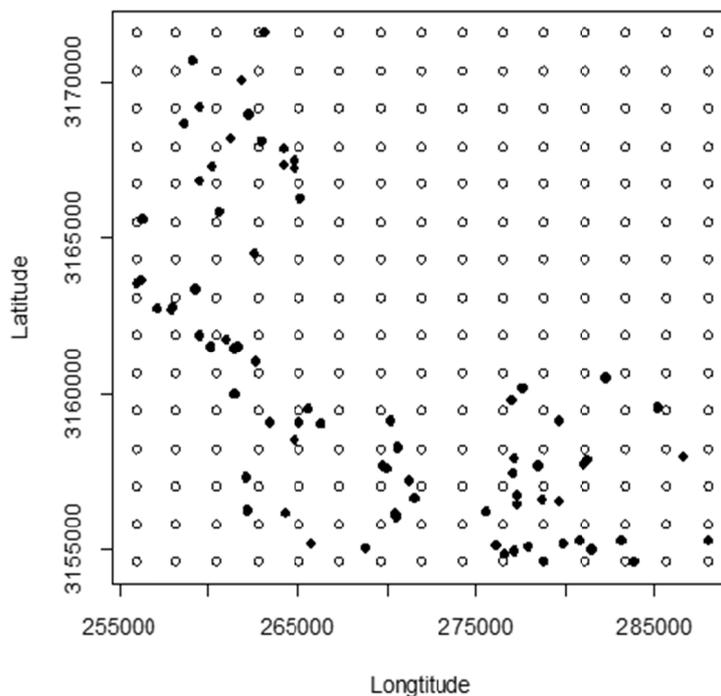


Figure 4. The observed values

### Results and Discussion

In order to define an RF in terms of the multivariate skew distributions, we proposed the multivariate EGAL distribution. We used the EGAL RF in spatial regression for skew and heavy-tailed data. The proposed RF can also be used for many spatial goals such as spatial prediction and spatial generalized mixed linear models which have been previously studied by SN RF and CSN RF. Although the work has been done only for the first order SARMA model, the results can be extended for the general model SARMA(p, q) with higher-order, in a similar manner. In spite of the basic difference in (7) for the general case, but other equations remain valid without any changes. The structure of matrices  $A_i$  and  $B_i$ ,  $i = 1, 2$  in SARMA(1,1) differ from SARMA(p, q) models. In the second case, these matrices are more complicated. Throughout the paper, a 2-dimensional lattice was considered. This assumption can be also generalized by considering a d-dimensional lattice. However, more complicated relations are expected to obtain for the general cases. For instance, a model SARMA(1,1) on a 3-dimensional lattice has the following form analogue with (6)

$$\begin{aligned}
 z_{ijk} = & \theta_1 z_{i-1,j,k} + \theta_2 z_{i,j-1,k} + \theta_3 z_{i,j,k-1} \\
 & + \theta_4 z_{i-1,j-1,k} + \theta_5 z_{i-1,j,k-1} \\
 & + \theta_6 z_{i,j-1,k-1} + \theta_7 z_{i-1,j-1,k-1} + \\
 & \varphi_1 \varepsilon_{i-1,j,k} + \varphi_2 \varepsilon_{i,j-1,k} + \varphi_3 \varepsilon_{i,j,k-1} \\
 & + \varphi_4 \varepsilon_{i-1,j-1,k} + \varphi_5 \varepsilon_{i-1,j,k-1} + \varphi_6 \varepsilon_{i,j-1,k-1} + \\
 & \varphi_7 \varepsilon_{i-1,j-1,k-1} + \varepsilon_{ijk},
 \end{aligned}$$

where  $i = 1, \dots, n_1$ ,  $j = 1, \dots, n_2$ ,  $k = 1, \dots, n_3$  and  $|\theta_\ell| < 1$ ,  $|\varphi_\ell| < 1$  for  $\ell = k = 1, \dots, 7$ . A combination of these two generalizations may be needed in some circumstances; however, it is out of the scope of this study.

### Acknowledgment

The authors are thankful to the referees for their many helpful comments that greatly improved this paper. We also wish to acknowledge for the support from the Center of Excellence in Analysis of Spatial and Spatio-Temporal Correlated Data at Tarbiat Modares University.

### References

1. Anselin L. Spatial Econometrics, methods and models (Kluwer, Dordrecht) 1988.

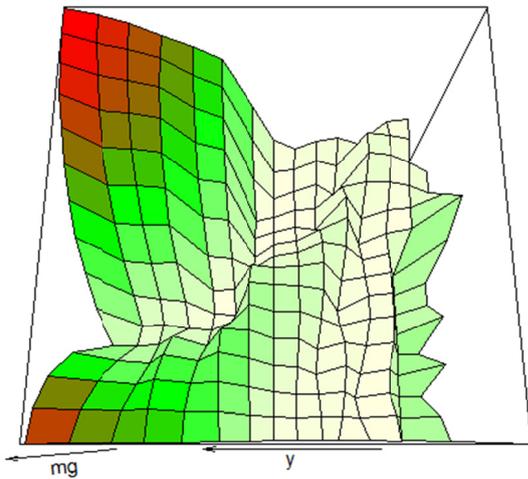


Figure 6. 3D plot of surface predictions based on EGAL model

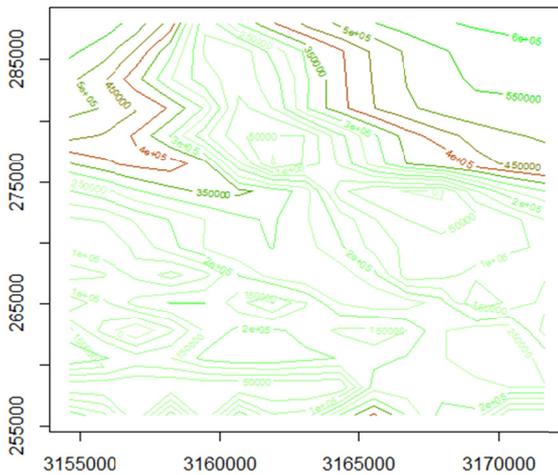


Figure 7. Contour plot of predictions.

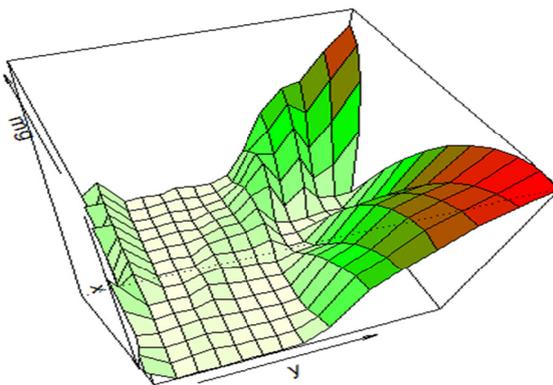


Figure 8. 3D plot of surface for prediction errors

2. Anselin L. Spatial dependence and spatial structural instability in applied regression analysis. *J Region Sci.* 1990; 30: 185-207.
3. Bustos O, Ojeda S and Vallejos R. Spatial ARMA models and its applications to image filtering. *Brazil J Probab Stat.* 2009; (23) 2: 141-165.
4. Shin DW and Sarkar S. Parameter estimation in regression models with auto-correlated errors using irregular data. *Commun Stat Theory Methods.* 1994; 23: 3567-3580.
5. Dogan O. Heteroskedasticity of unknown form in spatial autoregressive models with a moving average disturbance term. *Econometrics.* 2015; 3: 101-127.
6. Oh M, Shin DW and Kim HJ. Bayesian analysis of regression models with spatially correlated errors and missing observations. *Comput Stat Data Analys.* 2002; 39: 387-400.
7. Basu S and Reinsel GC. Regression models with spatially correlated errors. *J Am Stat Assoc.* 1994; 89: 88-99.
8. Kim H-M and Mallick BA. Bayesian prediction using the skew Gaussian distribution. *J Stat Plan Infer.* 2004; 120: 85-101.
9. Karimi O and Mohammadzadeh M. Bayesian spatial regression models with CSN correlated errors and missing observations. *Stat Papers.* 2012; 53: 205-218.
10. Saber MM, Nematollahi AR and Mohammadzadeh M. Generalized asymmetric Laplace random fields: existence and application. *J Data Sci.* 2018; 18: 51-68.
11. Kozubowski TJ, Podgórski K and Rychlik I. Multivariate generalized Laplace distribution and related random fields. *J Multivar Anal.* 2013; 113: 59-72.
12. Saber MM, Nematollahi AR and Mohammadzadeh M. Generalized Skew Laplace Random Fields: Bayesian Spatial Prediction for Skew and Heavy Tailed Data. *J Stat Theory Appl*, In Press.
13. Farzammehr MA, Zadkarami MR, McLachlan GJ and Lee SX. Skew-normal Bayesian spatial heterogeneity panel data models. *J Appl Stat.* 2020; 47(5): 804-826.
14. Kotz S, Kozubowski TJ and Podgórski K. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering and Finance*, Birkhäuser, Boston 2001.