

Bayesian Analysis of Censored Spatial Data Based on a Non-Gaussian Model

Vahid Tadayon

Eghlid Higher Education Institute

Received: 3/22/2016 Approved: 1/5/2017

Abstract. In this paper, we suggest using a skew Gaussian-log Gaussian model for the analysis of spatial censored data from a Bayesian point of view. This approach furnishes an extension of the skew log Gaussian model to accommodate to both skewness and heavy tails and also censored data. All of the characteristics mentioned are three pervasive features of spatial data. We utilize data augmentation method and Markov chain Monte Carlo (MCMC) algorithms to do posterior calculations. The methodology is illustrated using simulated data, as well as applying it to a real data set.

Keywords. Censored data; data augmentation; non-Gaussian spatial models; outlier; unified skew Gaussian.

MSC 2010: 62H11, 62F15

1 Introduction

The customary approach to spatial data modeling is to accept that the random field of interest is Gaussian. Nevertheless, this assumption always is threatened by the existence of non-Gaussianity features (such as: heavier tails or skewness) in data sets. Moreover, wrong Gaussian assumptions affect the accuracy of spatial predictions and cause bias in the resulting parameter estimates as well. So, the acceptance of the Gaussianity could be overly restrictive to interpret the quantity of interest. Therefore, there is a demand

for a more flexible class of sampling models to perform efficient inference and spatial prediction in the resulting models.

For modeling only skewed spatial data without heavy tails, the most commonly adopted strategy is to use the previous model after data transformation (Tadayon and Khaledi, 2015). Even so, an appropriate transformation if there exist, may not easy to obtain (see, e.g., De Oliveira et al., 1997). Kim and Mallick (2004) developed the skew Gaussian random field based on the skew normal distribution which has been introduced by Azzalini and Capitanio (1999). Zhang and El-Shaarawi (2010) recently introduced a new class of stationary process with skew normal marginal distributions. Using generalized skew Gaussian spatial field is also one of the other strategies which has been suggested by Prates et al. (2012). More details on skew spatial can be found in Rimstad and Omre (2014) and Karimi and Mohammadzadeh (2011).

On the other hand, the observed data may contain some isolated or grouped outliers which have extreme values compared to their neighboring observation values. Since these observations could provide additional information and knowledge about the model, they play an important role in the spatial statistics. Measurement error is the mainspring of encountering outlying observations and belonging to a region with larger observational variance relative to the rest, is another reason. Since the non-Gaussian characteristics of the data may be derived from these observations, in order to identify and treat with these regions, Palacios and Steel (2006) proposed a flexible class of sampling models called Gaussian log-Gaussian (GLG). In this model which contains a scale mixed random variable, is assumed that the mixed variable is log-normally distributed. Therefore, observations with small mixed random variable belong to a region with larger observational variance. In addition, to identify individual outliers, Fonseca and Steel (2011) considered a similar mixing in the nugget effect component. However, this approach is suitable only for symmetric heavier tail distributions and will fail to handle the skewed data (Zareifard and Khaledi, 2013). An interesting extension, therefore, would be to develop models based on mixtures of skew normal, which would yield a highly flexible and computationally tractable parametric model that could accommodate both multimodality and extreme skewness. This extension which is compatible with Genton and Zhang (2012) remedies, is the unified skew Gaussian-log Gaussian (SUGLG) model introduced by Zareifard and Khaledi (2013). They applied the stochastic approximation expectation-maximization (SAEM) algorithm to maximize the likelihood

function.

In addition, on account of some limitations of the data collection mechanism, the complete spatial data are not always available and so the quantity of interest may not be exactly observed in some sampling locations. In such setting, often the exact values can be recorded only if they fall within a specified range, which they have been called “censored data”. One of the conventional approaches to deal with these data is to impute the censored values with “artificial data”, such as, the detection limits. However, this leads biased estimates (see, e.g., Fridley and Dixon, 2007). Dubrule and Kostov (1986) worked on an interpolation method which motivated by the dual formulation of kriging. They also used geostatistical approaches requiring only the second-order specification of the random field which represent a type of censored data. Utilize likelihood-based approaches for the Gaussian random field is another previous work on the analysis of spatial models in the presence of censored data which well done by Militino and Ugarte (1999). De Oliveira (2005) introduced a Bayesian approach for inference and spatial prediction based on censored data while Fridley and Dixon (2007) incorporate the spatial correlation through an unobserved latent spatial process. Rathbun (2006) applied the Robbins-Monro (1951) stochastic approximation algorithm for estimating the parameters of a spatial regression model with left-censored observations. Eventually, Toscas (2010) proposed a modification on Bayesian approach of De Oliveira (2005) to correct the Bayesian bias in estimation and prediction of spatially correlated left-censored observations.

In all aforementioned works the inference was conducted based on the Gaussianity of the random field. However, in practice, we often observe that the exploratory data analysis shows some version of heavier tails than the normal distribution such as skewness or/and outlier region and consequently, it violates the normality assumption. The application of some transformation is the most commonly adopted strategy. However, the existence of an appropriate transformation is one of the most important problems in this setting and the difficulty in interpretation issues is another. Tadayon and Khaledi (2015) developed and implemented a Markov chain Monte Carlo (MCMC) sampling strategy for inference and fitted the skew Gaussian model based on censored data. Rivaz and Khaledi (2015) utilized a Bayesian framework and performed prediction for the unified skew-Gaussian (SUG) model in the presence of censored observations via a hybrid algorithm. In this article, we intend to fit the SUGLG model defined by Zareifard and Khaledi (2013)

on a set of spatial data which could take both skewness and heavy tails simultaneously and also contains some censored data.

In the following section we describe the form of spatial censored data which considered in this article. The third section introduces the mixture model and derives some of its properties. Section 4 considers Bayesian estimation of the model parameters and describes data augmentation method as well as Bayesian prediction. Using simulated data, the identifiability of the parameters, sensitivity analysis and the application of the suggested model is studied in Section 5. Section 6 illustrates the usage of the proposed methodology on a real spatial data set which contain values of precipitation. Conclusions are presented in Section 7.

2 Censored Spatial Data

Let $\mathbf{y} = (y_1, \dots, y_n)'$ be a realization of $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))'$, represents the data measured at the sampling locations s_1, s_2, \dots, s_n in R , where R is the region under study. In what follows, we assume that the mechanism that produced the censoring is uninformative, meaning that the censoring process is independent of the spatial process $Y(\cdot)$, but may be deterministic or random (De Oliveira, 2005). To argue the credibility of this assumption, consider a case in which the censoring time is independent of the survival time (Militino and Ugarte, 1999). Moreover, notice that this assumption is similar to that of data being missing at random. Thus, the observed data consist of exact observations measured at some sampling locations and interval observations measured in the rest as the result of censoring (Tadayon and Khaledi, 2015; Rivaz and Kaledi, 2015). In this manner, to avoid some difficulties, we consider that the data of size n , consist of m exact observations and $n - m$ interval observations. A reasonable methodology to define m in practice is related to the nature of the data. Therefore, the data would be denoted by

$$\mathcal{D} = \{Y(\mathbf{s}_{k_j}) = y_{k_j}; j \in J\} \cup \{Y(\mathbf{s}_{k_i}) \in A_i; i \in I\},$$

such that $\{y_{k_j}; j \in J\}$ are the observed values and for any $i \in I$, A_i is the interval where $Y(\mathbf{s}_{k_i})$ is known to belong. The censored interval A_i can depend on the sampling situation and it varies from a location to another location.

3 Statistical Model

Let $Y(\cdot) = \{Y(s); s \in R\}$ be the random field of interest where $R \subseteq \mathbb{R}^{d^*}$ and $d^* \geq 1$. Our starting point is the model

$$Y(s) = \mu(s) + \frac{W(s)}{\sqrt{\lambda(s)}} + \tau\rho(s), \quad \tau \geq 0, \quad (1)$$

with the mean surface $\mu(s)$ and the scale parameter $\tau \in \mathbb{R}^+$. $\mu(s)$ is assumed to be a linear function of $\mathbf{f}'(s)$, suchlike $\mu(s) = \mathbf{f}'(s)\boldsymbol{\beta}$ where, $\mathbf{f}'(s)$ is a vector of k known functions of the spatial coordinate and $\boldsymbol{\beta} \in \mathbb{R}^k$ is the regression coefficient vector. The term $\rho(s)$ denotes an uncorrelated Gaussian process with mean 0 and unitary variance, modeling the so-called *nugget effect*, which allows for measurement error and small-scale variation. This further, the second-order stationary error process $W(s)$, is a unified skew Gaussian random process corresponding to the model defined by Arellano-Valle and Azzalini (2006). On the other hand, for any vector from this random process like $\mathbf{W} = (W(s_1), \dots, W(s_n))'$, we consider that the distribution of vector \mathbf{W} is $\mathbf{W} \stackrel{d}{=} \alpha\mathbf{U} + \sigma\mathbf{V}$ in which, \mathbf{V} and \mathbf{U} have n -variate normal distribution $N_n(\mathbf{0}, C_{\theta_w})$ and truncated normal distribution $TN_n(\mathbf{0}; \mathbf{0}, C_{\theta_w})$ respectively. The (i, j) th element of the $n \times n$ correlation matrix C_{θ_w} is $C_{\theta_w}(\|s_i - s_j\|)$. Thus, \mathbf{W} has multivariate unified skew-normal (SUN) distribution (See Appendix C)

$$\mathbf{W} \sim \text{SUN}_{n,n}(\mathbf{0}, (\alpha^2 + \sigma^2)C_{\theta_w}, \alpha C_{\theta_w}, \mathbf{0}, C_{\theta_w}).$$

By setting $\boldsymbol{\rho} = (\rho(s_1), \dots, \rho(s_n))'$ and $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$, equation (1) can be written as

$$\mathbf{Y} \stackrel{d}{=} \mathbf{X}\boldsymbol{\beta} + \alpha\Lambda^{-\frac{1}{2}}\mathbf{U} + \sigma\Lambda^{-\frac{1}{2}}\mathbf{V} + \tau\boldsymbol{\rho}.$$

In addition, like Palacios and Steel (2006) we assume that $\ln \lambda(s)$ is a Gaussian random field with finite-dimensional distributions:

$$(\ln(\lambda_1), \dots, \ln(\lambda_n))' \sim N_n\left(-\frac{\nu}{2}\mathbf{J}, \nu C_{\theta_\lambda}\right),$$

for $\mathbf{J}' = (1, \dots, 1)_{n \times 1}$ and $\nu > 0$, which implies a lognormal distribution for each λ_i . Clearly, by tending ν to 0, λ_i tends to 1 and so equation (1) reduces to unified skew Gaussian model. For simplicity, we adopt a same

correlation matrices for \mathbf{W} and $\boldsymbol{\lambda}$ but with different parameters θ_w and θ_λ , respectively, and also we suppose that both \mathbf{W} and $\boldsymbol{\lambda}$ have an isotropic exponential correlation function, separately. However, the first assumption might be somewhat restrictive especially in the case of spatial data without temporal replications (as identifiability problem in these parameters, if exist, can affect the estimates of parameters), but we can hope that this problem is possibly solvable when there are many observations in space. Section 5 confirms this claim. Following Palacios and Steel (2006), we set $\omega^2 = \frac{\tau^2}{\sigma^2}$ for simplicity and identifiability.

The model (1) is able to accommodate each of individual outliers and areas of the space with an inflated variance relative to the rest, because of observations with small values λ_i fall out of the mean surface.

Let $\eta' = (\boldsymbol{\beta}, \alpha, \sigma^2, \omega^2, \nu, \theta_w, \theta_\lambda)$ and $\mathbf{y} = (y_I, y_J) = (y_1, \dots, y_n)$ be a single realization of the considered random field with $y_i = Y(s_i)$, then the conditional joint distribution of \mathbf{y} given $\boldsymbol{\lambda} = (\lambda(s_1), \dots, \lambda(s_n))'$ and η is

$$p(\mathbf{y} | \boldsymbol{\lambda}, \eta) \sim \text{SUN}_{n,n} \left(\mathbf{X}\boldsymbol{\beta}, \Sigma_y, \alpha \Lambda^{-\frac{1}{2}} C_{\theta_w}, 0, C_{\theta_w} \right).$$

Furthermore, the likelihood function for η given \mathcal{D} is

$$\begin{aligned} L(\eta | \mathcal{D}) &= \int_{\mathbb{R}^{+n}} \int_{\mathbb{R}^n} \int_{\prod_{i \in I} A_i} p(Y_I \in \prod_{i \in I} A_i | y_J, \mathbf{w}, \boldsymbol{\lambda}, \eta) \\ &\quad \times p(y_J | \mathbf{w}, \boldsymbol{\lambda}, \eta) p(\mathbf{w} | \boldsymbol{\lambda}, \eta) \\ &\quad \times p(\boldsymbol{\lambda} | \eta) dy_I d\mathbf{w} d\boldsymbol{\lambda}, \end{aligned} \quad (2)$$

where $\mathbf{w} = (w(s_1), \dots, w(s_n))'$, $\Sigma_y = \Sigma_w + \tau^2 \mathbf{I}_n$ and \mathbf{I}_n is the identity matrix of dimension n . By setting $C_{JJ\theta_w} = [C_{\theta_w} (\|s_{J_i} - s_{J_j}\|)]_{m \times m}$, we can easily see that

$$\mathbf{Y}_J | \boldsymbol{\lambda}, \eta \sim \text{SUN}_{m,m} \left(X_J \boldsymbol{\beta}, \Sigma_{y_J}, \alpha \Lambda_J^{-\frac{1}{2}} C_{JJ\theta_w}, 0, C_{JJ\theta_w} \right)$$

and the conditional distributions of interval observations $\mathbf{Y}_I | y_J, \boldsymbol{\lambda}, \eta$ can be

written as $\text{SUN}_{n-m, n-m} \prod_{i \in I} I_{\{y_i \in A_i\}}$ in which its parameters are

$$\begin{aligned}\mu_{y_I|y_J, \boldsymbol{\lambda}, \eta} &= X_I \boldsymbol{\beta} + A (y_J - X_J \boldsymbol{\beta}), \\ \Sigma_{y_I|y_J, \boldsymbol{\lambda}, \eta} &= (\sigma^2 + \alpha^2) \Lambda_I^{-\frac{1}{2}} C_{II\theta_w} \Lambda_I^{-\frac{1}{2}} + \tau^2 I_{n-m} - A \left(\Lambda_J^{-\frac{1}{2}} C_{JI\theta_w} \Lambda_J^{-\frac{1}{2}} \right), \\ \Gamma_{y_I|y_J, \boldsymbol{\lambda}, \eta} &= \Lambda_I^{-\frac{1}{2}} C_{IJ\theta_w} - A \Lambda_J^{-\frac{1}{2}} C_{JJ\theta_w}, \\ \Delta_{y_I|y_J, \boldsymbol{\lambda}, \eta} &= C_{JJ\theta_w} - \alpha^2 C_{JJ\theta_w} \Lambda_J^{-\frac{1}{2}} \Sigma_{y_J}^{-1} \Lambda_J^{-\frac{1}{2}} C_{JJ\theta_w},\end{aligned}$$

where

$$\begin{aligned}A &= (\sigma^2 + \alpha^2) \Lambda_I^{-\frac{1}{2}} C_{IJ\theta_w} \Lambda_J^{-\frac{1}{2}} \Sigma_{y_J}^{-1}, \\ C_{IJ\theta_w} &= [C_{\theta_w} (\|s_{I_i} - s_{J_j}\|)]_{(n-m) \times m}\end{aligned}$$

and the rest is 0. It is necessary to note that $I_{\{A\}}$ shows the indicator function of the set A.

4 Bayesian Inference

To complete the Bayesian model specification, we need to select some prior distributions for all unknown parameters. One can then easily see that a convenient strategy of avoiding improper posterior distribution, is to utilize proper (but diffuse) priors. However, in this case, the insignificance of the prediction sensitivity to the hyperparameters should be established. For convenience but not necessary optimal, we assume elements of the parameter vector η to be independent a priori, which means that

$$\pi(\eta) = \pi(\boldsymbol{\beta}) \pi(\alpha) \pi(\sigma^2) \pi(\omega^2) \pi(\nu) \pi(\theta_w) \pi(\theta_\lambda). \quad (3)$$

The prior distributions adopted are as follows: $\boldsymbol{\beta} \sim N_k(\mathbf{0}, c_0 \mathbf{I}_k)$, $\alpha \sim N(0, c_1)$, $\sigma^{-2} \sim \text{Gamma}(c_2, c_3)$, $\omega^{-2} \sim \text{GIG}(0, c_4, c_5)$, $\nu \sim \text{GIG}(0, c_6, c_7)$, $\theta_w \sim \text{Exp}\left(\frac{c_8}{\text{med}(d)}\right)$ and $\theta_\lambda \sim \text{Exp}\left(\frac{c_9}{\text{med}(d)}\right)$, where the hyperparameters c_1, \dots, c_9 are chosen to reflect vague prior information and $\text{med}(d)$ is the median of all distance between the data locations. Moreover, the closed form of generalized inverse-Gaussian (GIG) distribution and some justification for utilizing the above priors can be found in Palacios and Steel (2006).

Now, we try to derive the posterior distribution of the parameters by combining the likelihood function (2) and the prior distribution (3). Since a direct Bayesian analysis of the model is computationally impractical because of multiple integrals in posterior distribution, we adopt the data augmentation method and use the latent variables \mathbf{U} and $\boldsymbol{\lambda}$ to produce some samples $\{(\hat{I}^{(i)}, u^{(i)}, \lambda^{(i)}, \eta^{(i)})\}_{i=1}^l$ from the extended model $p(y_I, \mathbf{U}, \boldsymbol{\lambda}, \eta, |\mathcal{D})$ via MCMC methods. Details are given in Appendix A.

It must be noted that in many applications, the prediction of values at unsampled locations is a usual task of interest. The spatial prediction of the response in an unobserved location y_0 would be based on the Bayesian predictive distribution:

$$p(y_0 | \mathbf{y}) = \int_{\prod_{i \in I} A_i} \int_{R_+^n} \int_{R_+^p} p(y_0 | \mathbf{y}, \boldsymbol{\psi}, \boldsymbol{\psi}_0) p(\boldsymbol{\psi}_0 | \mathbf{y}, \boldsymbol{\psi}) p(\boldsymbol{\psi} | \mathbf{y}) d\boldsymbol{\psi}_0 d\boldsymbol{\psi} dy_I,$$

where $\boldsymbol{\psi}(s) = \ln \boldsymbol{\lambda}(s)$. Evidently, the full predictive distribution can not be evaluated in closed form but can be approximated using Monte Carlo samples.

5 Simulation Studies

The main aim of this section is to assess the performance of the proposed methodology on detecting two levels of censoring and also represent prediction values at some new unobserved locations. But since the SUGLG model may have a problem to indentify the correlation parameters (see Section 3), so, we first use simulation to evaluate the identifiability of these parameters. Then we present a banal method to revalue the ability of the this model to correctly identify outlying observations as well. Throughout, results are based on every 100th draw from an MCMC chain of length 150,000 with a burn-in of 100,000 which provides enough output for convergence. The spatial sampling points were considered on coordinates of the data file *97data.dat* available from GSLIB software (Deutsch and Journel, 1998) in which, 97 locations are taken on a pseudo-regular grid over a bidimensional region 50 by 50 miles.

First of all, we appoint a lattice locations

$$\{12, 21, 28, 40\} \times \{10, 20, 30, 40\}, \quad (4)$$

as a hold-out data set to evaluate the performance of the model in spatial prediction and then data are generated on these 113 locations based on the sampling model

$$\mathbf{Y} = \beta_0 + \alpha\mathbf{U} + \sigma\mathbf{V} + \tau\rho,$$

with a constant mean surface $\beta_0 = 0$, $\alpha = 3$, $\sigma = 1$, $\tau = 0.1$ and $\theta_\lambda = \theta_w = 0.5$. After that, a cluster of outliers at locations 29, 37, 59, 78 and 84 has been created by imposing two units to the simulated values. Then, by leaving the simulated value in the lattice locations (4), the model will be fitted on the rest (of size 97). For model validation, we now implement two levels of left-censored design, one 17.5% and other 67% as a *balanced* and *extreme* cases of censoring respectively, such that in the first of them, the 17th ordered observation will be substituted for 1st up to 17th smallest values and in the second, 1st up to 65th smallest observations will be replaced with 65th ordered observation. Figure 1 shows a schematic description of the region that displays the sampling locations as well as the region with inflated variance and censored locations. As an exploratory analysis, Figure 2 which shows the histogram with the nonparametric density estimator of generated data, confirm the existence of an outlier region in these data as well as skewness.

Since the inference may be challenging in identifying the correlation parameters θ_w and θ_λ , here, we focus on this problem to see what extent information about these parameters can be recovered from data. This study will be done just based on the balanced case of censoring because we do not expect a high ability to identify the parameters in the extreme case. To check the aforementioned problem, three data sets were generated with $\theta_w = 0.25, 1, 1.5$ and $\theta_\lambda = 0.1, 0.7, 1.1$ (in which, the other values have been fixed) and the estimate of the model parameters are obtained. However there are quite moderate sample size, the results which are reported in Table 1, indicates that there is no significant problem on the identifying of the correlation parameters and so the data almost allow for acceptable inference on these parameters.

At the moment, we concentrate on the performance of the Bayesian estimation, in which we must do one of these actions: choose some benchmark values for the various hyperparameters used in the priors, using empirical Bayes or employing hierarchical Bayes. To facilitate the computations, we consider the first method. But, we must investigate the robustness of the pos-

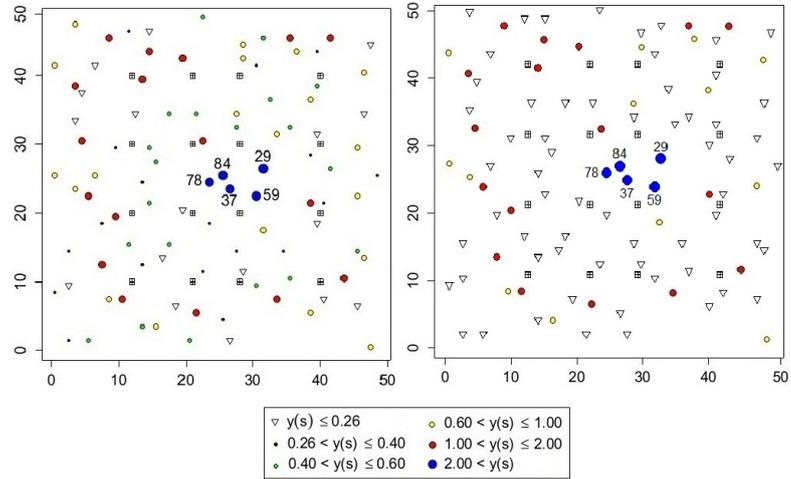


Figure 1. Study region based on balanced case (on the left) and extreme case (on the right) of censoring.

○ = exact observation.

▽ = left-censored observation.

⊠ = location which is added to spatial prediction.

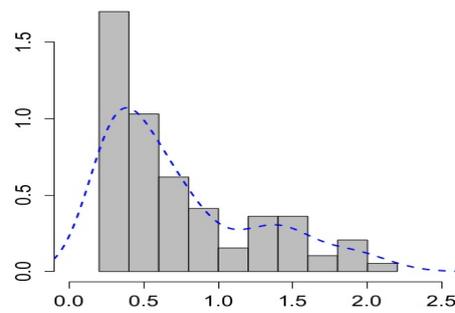


Figure 2. Histogram with the nonparametric density estimator of the generated data.

Table 1. Identifiability of the correlation parameters based on the balanced case of censoring.

θ_w	True value	0.2	1	1.5
	Estimated value	0.98	1.15	1.41
θ_λ	True value	0.1	0.7	1.1
	Estimated value	0.14	0.81	0.98

Table 2. Prior Sensitivity Analysis: Setup of the Experiment and Maximum Relative Change (MRC).

Hyperparameter	Benchmark	Alternative values	MRC		
			17.5%	67%	
β_0	c_0	10^4	$10^2, 10^6$	0.0611	0.0926
α	c_1	10^5	$10^3, 10^7$	0.132	0.1915
σ^2	(c_2, c_3)	$(10^{-6}, 10^{-6})$	$(10^{-4}, 10^{-8}), (10^{-8}, 10^{-4})$	0.1709	0.1135
ω^2	(c_4, c_5)	$(0.1, 9)$	$(0.7, 1.5), (0.5, 0.7)$	0.3147	0.652
θ_w	c_8	0.7	0.4, 1.4	0.393	0.4117
θ_λ	c_9	0.7	0.35, 1.60	0.421	0.4705
ν	(c_6, c_7)	$(0.5, 1.5)$	$(0.30, 1.0), (0.75, 0.8)$	0.8003	0.96

terior results under changes in the prior hyperparameters. It must be noted that we do not expect stability of the posteriors per se, but rather our interest is often in prediction. For this purpose, we employ the *relative change* criterion, which is the absolute value of the induced change in the marginal posterior mean of each parameter divided by the standard deviation computed under the benchmark prior. Table 2 which lists the alternative prior hyperparameters as well as the maximum relative change recorded for each parameter, shows no significant changes.

Therefore, the sensitivity of posterior can be partitioned into three groups: First, the prior on $(\beta_0, \alpha, \sigma^2)$ which does not play a vital role, $(\omega^2, \theta_w, \theta_\lambda)$ where the data is reasonably informative on these parameters as the second and the last is ν by the largest influence of prior changes. Since there is relatively little direct information in the data on ν , choice of prior is quite important. However, the adoption of hierarchical or empirical Bayes can be helpful.

Table 3 summarizes the results of fitting four models:

- Gaussian (GAUS): $Y(s) = \beta_0 + \tau\rho(s)$,

Table 3. The estimation of the parameters in four models: GAUS, SUG, GLG and SUGLG.

			β_0	α	σ^2	ω^2	θ_w	θ_λ	ν
		True value	0	3	1	0.1	0.5	0.5	1
Estimated values	GAUS	17.5%	1.67	-	2.17	0.9	0.38	0.15	-
		67%	2.03	-	1.86	0.91	0.46	0.4	-
	SUG	17.5%	1.21	2.03	1.68	0.83	0.45	0.12	-
		67%	1.42	1.35	1.4	0.71	0.58	0.39	-
	GLG	17.5%	0.98	-	1.74	0.66	0.22	0.34	0.51
		67%	1.5	-	1.38	0.59	0.55	0.59	1.23
	SUGLG	17.5%	0.05	2.77	1.33	0.2	0.56	0.45	1.12
		67%	0.8	2.13	1.15	0.23	0.46	0.53	0.92

Table 4. The root of mean square error (RMSE) of the predicted response values based on GAUS, SUG, GLG and SUGLG models.

		GAUS	SUG	GLG	SUGLG
RMSE	17.5%	0.191	0.083	0.095	0.031
	67%	0.211	0.187	0.166	0.102

- Unified Skew Gaussian (SUG): $Y(s) = \beta_0 + W(s) + \tau\rho(s)$,
- Gaussian-Log Gaussian (GLG): $Y(s) = \beta_0 + \lambda(s)^{-\frac{1}{2}} + \tau\rho(s)$,
- Unified Skew Gaussian-Log Gaussian (SUGLG):

$$Y(s) = \beta_0 + \frac{W(s)}{\sqrt{\lambda(s)}} + \tau\rho(s).$$

According to this table, it is found that the SUGLG model has a better performance. Moreover, to compare four models in prediction, we compute the root of mean-square error of the predicted response values on the determined lattice locations. Table 4 suggests the desirability of using the SUGLG model.

Finally, by plotting the Bayesian estimate of λ_i 's for $i = 1, \dots, n$, it can be seen that how posterior values of λ_i 's might be used to identify outliers (because of the effect of small values of λ_i 's in the denominator, see equation

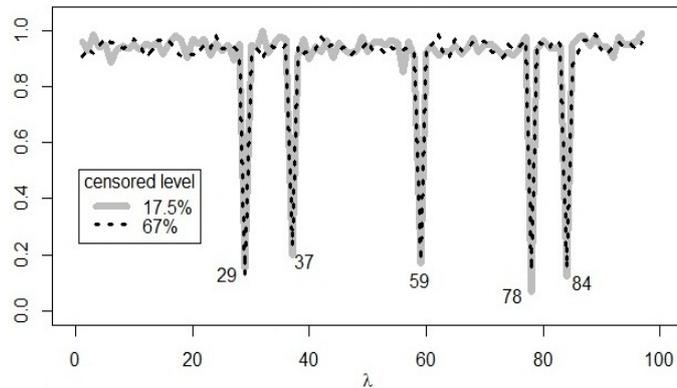


Figure 3. Mean of the posterior values λ_i 's for $i = 1, \dots, 97$. Small values of λ_i 's lead to large prediction in response.

(1). Figure 3 which shows mean of the posterior values λ_i 's for $i = 1, \dots, 97$, confirms the existence of an outlier region in the clustered locations 29, 37, 59, 78 and 84.

6 Rainfall Data Analysis

The analyzed data set is comprised of rainfall amounts, measured in inch and accumulated on the first and coldest Wednesday in December 2012 at 30 stations located in Fars province of Iran. A schematic description of the region, the stations and rainfall amounts is shown in Fig. 4. In this figure it can be seen that there is a region with larger observation relative to the rest, along the north-eastern of the study region. In this data set (which are presented in Appendix B), the zero values are reported for five stations. Nevertheless, since the nonzero values are reported for the countrysides of these stations by residents and the stations operators, we consider their values as censored with censored intervals $[0, 0.01)$, where the value 0.01 is the least value of precipitation in their countrysides. These locations are indicated by "▽" as a left-censored observations in Figure 4.

First of all, we performed the exploratory data analysis (EDA) such as the histogram, Q-Q plot, the Shapiro-Wilk's test, the empirical semivariogram and the Haining (1991) method to detect outlier observations. Some of those results are presented in Fig. 5.

The histogram and Q-Q plot suggest that the data set has moderately

Table 5. Shapiro-Wilk's normality test.

Shapiro-Wilk	P-value
0.8765	0.002343

right-skewed distribution which was confirmed by using the Shapiro-Wilk test where reported in Table 5. In addition, comparing the Haining bounds and the response values, indicate a region in the northings of studying space with larger observational variance relative to the rest, where we observed large amounts of precipitation. Finally, it is clear from the empirical semi-variograms that there exists a strong spatial correlation as well as a nugget effect in the data set.

Before starting to compute estimations and spatial predictions, we mention that, since explanatory analysis of the data did not show any significant relation between response value and the spatial locations, the mean function is assumed to be constant, so we have just β_0 and thus $k = 1$. This exploratory data analysis suggests that SUGLG model is a suitable option for doing the analysis. To perform Bayesian inference, benchmark values of the hyperparameter have chosen same as the given values in Table 2. The results is represented in Table 6. Although the DIC and the LMPL ($= \sum_{i=1}^n \log(CPO_i)$) (The conditional predictive ordinate (CPO): The CPO model comparison is a Bayesian cross-validation approach (e.g., Geisser, 1993)) criterion present a versus results in comparison between two SUG and GLG models in some situations, but both of them confirm the better performance of the SUGLG model.

Moreover, the prediction of rainfall amounts in the five mentioned stations, is presented in Table 7. One of the attractive feature of the Bayesian approach of course, is that it allows to make inference about the censored values, and in particular to quantify uncertainty about them. Figure 6 displays the histograms of the posterior samples of the censored values, where vertical lines are placed at their respective censoring limits. It is clear that for some of these locations there is a lot of uncertainty about the true value, while for others the respective censoring limit is a poor estimate of the true value.

Finally, the contour map corresponding to the predictive mean surface under the SUGLG case, is shown in Figure 7 which are computed from the predictive distribution over a regular grid of 20×20 . According to this figure, the predictions are highest in the area of inflated variance that contains

Table 6. Estimation (Est) and Standard deviation (SD) of the parameters in four models: GAUS, SUG, GLG and SUGLG with the DIC and the LPML criterion to comparison these models.

		β_0	α	σ^2	ω^2	θ_w	θ_λ	ν	DIC	LPML
GAUS	Est	0.006	-	2.13	0.291	0.036	0.521	-		
	SD	1.71	-	1.03	0.84	2.17	1.97	-	57.3	19.4
SUG	Est	0.043	2.76	1.98	0.284	0.251	0.56	-		
	SD	1.65	1.19	1.11	0.761	2.061	2.01	-	40.7	31.1
GLG	Est	0.044	-	1.9	0.276	0.232	0.473	0.32		
	SD	1.13	-	1.86	0.874	1.81	1.68	0.24	39.6	30
SUGLG	Est	0.011	2.22	1.4	0.197	0.461	0.465	2.61		
	SD	0.33	0.95	0.41	0.29	0.925	0.71	0.13	11.4	43.7

Table 7. Bayesian spatial prediction of rainfall values in five stations.

	Observed value	Predicted value
Darab	0	0.0070121
Firuzabad	0	0.009761
Lamerd	0	0.0099013
Larestan	0	0.0095678
Qir-o-Karzin	0	0.0095411

the aforementioned observations. Figure 8 presents the standard deviation surfaces for the SUGLG case as well.

7 Conclusions

In this article we have considered a unified skew Gaussian-log Gaussian model to analyze of non-Gaussian random fields based on censored data. Overall, the Bayesian framework, data augmentation method and MCMC algorithms have been extended for parameter estimation. However, using this algorithm entail more complication in the computations, recently, some studies have been performed to find replacing framework and it has led to some new

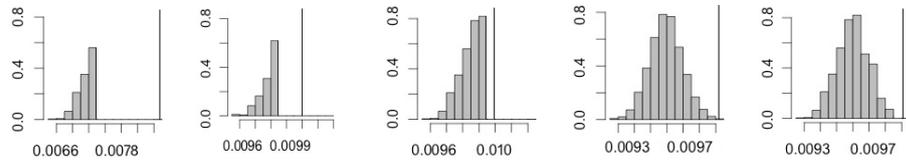


Figure 6. Estimates of the posterior distribution of $Y(s_i)$ for five censored observations. Vertical lines are placed at the respective censoring limits. (a) to (e) correspond to Darab, Firuzabad, Lamerd, Larestan and Qir-o-Karzin, respectively.

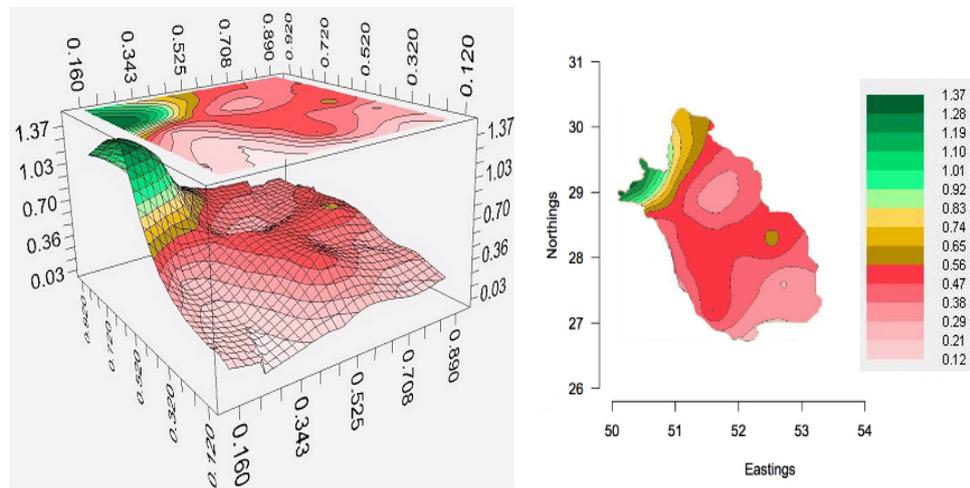


Figure 7. The contour map corresponding to the predictive mean for rainfall dataset.

methods such as Variational Bayes' method (Ren et al. 2011). Although this method requires more complex theoretic calculations, it could increase the speed of calculations. So, assessing the performance of this method in Bayesian inference of Non-Gaussian random fields in the presence of censored values is an interesting area to investigate in further research.

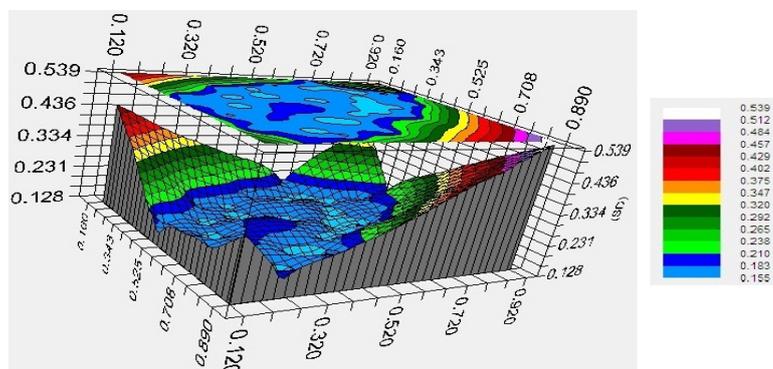


Figure 8. Topographic data: standard deviation surface of predictive mean.

References

- Arellano-Valle, R.B. and Azzalini, A. (2006). On the Unification of Families of Skew-normal Distributions. *Scandinavian Journal of Statistics*, **33**, 561-574.
- Azzalini, A. and Capitanio, A. (1999). Statistical Applications of the Multivariate Skew Normal Distribution. *Journal of the Royal Statistical Society, Soc. B*, **61**, 579-602.
- De Oliveira, V. (2005). Bayesian Inference and Prediction of Gaussian Random Fields based on Censored Data. *Journal of Computational and Graphical Statistics*, **14**, 95-115.
- De Oliveira, V., Kedem, B. and Short, D.A. (1997). Bayesian Prediction of Transformed Gaussian Random Fields. *Journal of the American Statistical Association*, **92**, 1422-1433.
- Deutsch, C.V. and Journel, A.G. (1998). *GSLIB: Geostatistical Software Library and Users Guide*. New York: Oxford University Press.
- Dubrule, O. and Kostov, C. (1986). An Interpolation Method taking into Account Inequality Constraints. *Mathematical Geosciences*, **18**, 33-51.
- Fonseca, T. and Steel, M. (2011). Non-Gaussian Spatiotemporal Modelling Through Scale Mixing. *Biometrika*, **98**, 761-774.
- Fridley, B.L. and Dixon, P. (2007). Data Augmentation for a Bayesian Spatial Model Involving Censored Observations. *Environmetrics*, **18**, 107-123.
- Geisser, S. (1993). *Predictive Inference: An Introduction*, London: Chapman & Hall Ltd.
- Genton, M.G. and Zhang, H. (2012). Identifiability Problems in Some non-Gaussian Spatial Random Fields. *Chilean Journal of Statistics*, **3**, 171-179.

-
- Haining, R. (1991). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge: Cambridge University Press.
- Karimi, O. and Mohammadzadeh, M. (2011). Bayesian Spatial Prediction for Discrete Closed Skew Gaussian Random Field. *Mathematical Geosciences*, **43.5**, 565-582.
- Kim, H.M. and Mallik, B.K. (2004). A Bayesian Prediction using the Skew-Gaussian Distribution. *Journal of Statistical Planning and Inference*, **120**, 85-101.
- Militino, A.F. and Ugarte, M.D. (1999). Analyzing Censored Spatial Data. *Mathematical Geosciences*, **31**, 551-562.
- Palacios, M. and Steel, M. (2006). Non-Gaussian Bayesian Geostatistical Modeling. *Journal of the American Statistical Association*, **101**, 604-618.
- Prates, Marcos, O., Dey, K.D. and Lachos, V.H. (2012). A Dengue fever Study in the State of Rio de Janeiro with the Use of Generalized Skew-normal/Independent Spatial Fields. *Chilean Journal of Statistics*, **3**, 33-45.
- Rathbun, S. L. (2006). Spatial Predictions with Left-censored Observations. *Journal of Agricultural, Biological, and Environmental Statistics*, **11**, 317-336.
- Ren, Q., Banerjee, S., Finley, A.O. and Hodges, J.S. (2011). Variational Bayesian Methods for Spatial Data Analysis. *Computational Statistics & Data Analysis*, **55**, 3197-3217.
- Rimstad, K. and Omre, H. (2014). Skew-Gaussian Random Fields. *Spatial Statistics*, **10**: 43-62.
- Rivaz, F. and Khaledi, M.J. (2015). Bayesian Spatial Prediction of Skew and Censored Data via a Hybrid Algorithm. *Journal of Applied Statistics*, doi:10.1080/02664763.2015.1014895.
- Robbins, H. and Monro, S. (1951). A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, **22**, 400-407.
- Tadayon, V. and Jafari Khaledi, M. (2015). Bayesian Analysis of Skew Gaussian Spatial Models based on Censored Data. *Communications in Statistics - Simulation and Computation*, **44**, 2431-2441.
- Toscas, P.J. (2010). Spatial Modeling of Left-censored Water Quality Data. *Environmetrics*, **21**, 632-644.
- Zareifard, H. and Jafari Khaledi, M. (2013). Non-Gaussian Modeling of Spatial Data using Scale Mixing of a Unified Skew-Gaussian Process. *Journal of Multivariate Analysis*, **114**, 16-28.
- Zhang, H. and El-Shaarawi, A. (2010). On Spatial Skew-Gaussian Processes and Applications. *Environmetrics*, **21**, 33-47.

Appendix A: The Full Conditional Distributions

In sequel, we try to present the full conditional distributions of all unknown quantities $p(y_I, \mathbf{U}, \boldsymbol{\lambda}, \eta, |\mathcal{D})$ in the Gibbs sampler framework. Moreover, to facilitate the computations we put the notation $\eta_{-\phi}$ to show the vector η without the element ϕ and set $Z = \mathbf{y} - X\boldsymbol{\beta}$.

- **Full conditional distribution of censored values**

By considering the partition of $\mathbf{u} = (u_I, u_J)$ we have

$$\mathbf{Y}_I | \mathbf{u}, v, \boldsymbol{\lambda}, \eta, \mathcal{D} \sim N_{n-m}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*) \prod_{i \in I} I_{\{y_i \in A_i\}},$$

where,

$$\begin{aligned} \boldsymbol{\mu}^* &= X_I \boldsymbol{\beta} + \alpha \Lambda_I^{-\frac{1}{2}} u_I + \Sigma_{y_I y_J} \Sigma_{y_J y_J}^{-1} \left(y_J - X_J \boldsymbol{\beta} - \alpha \Lambda_J^{-\frac{1}{2}} u_J \right), \\ \boldsymbol{\Sigma}^* &= \Sigma_{y_I y_I} - \Sigma_{y_I y_J} \Sigma_{y_J y_J}^{-1} \Sigma_{y_J y_I}. \end{aligned}$$

which is a $(n - m)$ -variate truncated normal distribution and so it can easily generated by standard algorithm.

- **Full conditional distribution of latent variable \mathbf{U}**

$$\begin{aligned} p(\mathbf{U} | \mathbf{y}, \boldsymbol{\lambda}, \eta) &\propto p(\mathbf{y} | \mathbf{u}, \boldsymbol{\lambda}, \eta) p(\mathbf{U} | \boldsymbol{\lambda}, \eta) \pi(\boldsymbol{\lambda} | \eta) \\ &\propto \exp \left\{ -\frac{1}{2} \left(Z - \alpha \Lambda^{-\frac{1}{2}} \mathbf{U} \right)' \boldsymbol{\Sigma}_y^{-1} \left(Z - \alpha \Lambda^{-\frac{1}{2}} \mathbf{U} \right) \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \mathbf{U}' C_{\theta_w}^{-1} \mathbf{U} \right\} \\ &\propto \exp \left[-\frac{1}{2} \left\{ \mathbf{U}' \left(C_{\theta_w}^{-1} + \alpha^2 \Lambda^{-\frac{1}{2}} \boldsymbol{\Sigma}_y^{-1} \Lambda^{-\frac{1}{2}} \right) \right. \right. \\ &\quad \left. \left. \times \mathbf{U} + 2\alpha \mathbf{U}' \Lambda^{-\frac{1}{2}} \boldsymbol{\Sigma}_y^{-1} Z \right\} \right], \end{aligned}$$

which means that $\mathbf{U} | \mathbf{y}, \boldsymbol{\lambda}, \eta \sim \text{TN}_n(\mathbf{0}; \boldsymbol{\mu}_u^*, H_u^{-1})$ by

$$\begin{aligned} \boldsymbol{\mu}_u^* &= H_u^{-1} \left(\Lambda^{-\frac{1}{2}} \boldsymbol{\Sigma}_y^{-1} \right) Z, \\ H_u &= \frac{1}{\alpha^2} C_{\theta_w}^{-1} + \Lambda^{-\frac{1}{2}} \boldsymbol{\Sigma}_y^{-1} \Lambda^{-\frac{1}{2}}. \end{aligned}$$

• **Full conditional distribution of regression coefficients**

$$\begin{aligned}
 \pi(\boldsymbol{\beta} | y_I, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\boldsymbol{\beta}}, \mathcal{D}) &\propto \pi(\boldsymbol{\beta} | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\boldsymbol{\beta}}) \\
 &\propto p(\mathbf{y} | \mathbf{u}, \boldsymbol{\lambda}, \eta) \pi(\boldsymbol{\beta}) \\
 &\propto \exp \left\{ -\frac{1}{2} \left(\mathbf{y}' \Sigma_y^{-1} X \boldsymbol{\beta} - \boldsymbol{\beta}' X' \Sigma_y^{-1} \mathbf{y} \right. \right. \\
 &\quad \left. \left. + \boldsymbol{\beta}' X' \Sigma_y^{-1} X \boldsymbol{\beta} + \boldsymbol{\beta}' X' \Sigma_y^{-1} \alpha \Lambda^{-\frac{1}{2}} \mathbf{u} \right. \right. \\
 &\quad \left. \left. + \alpha \mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} X \boldsymbol{\beta} + \boldsymbol{\beta}' (c_0 I_k)^{-1} \boldsymbol{\beta} \right) \right\} \\
 &\propto \exp \left[-\frac{1}{2} \left\{ \boldsymbol{\beta}' \left(X' \Sigma_y^{-1} X + (c_0 I_k)^{-1} \right) \boldsymbol{\beta} \right. \right. \\
 &\quad \left. \left. - 2 \boldsymbol{\beta}' X' \Sigma_y^{-1} \left(\mathbf{y} - \alpha \Lambda^{-\frac{1}{2}} \mathbf{u} \right) \right\} \right].
 \end{aligned}$$

So, $\boldsymbol{\beta} | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\boldsymbol{\beta}} \sim N_k(\boldsymbol{\mu}_{\boldsymbol{\beta}}^*, \mathbf{H}_{\boldsymbol{\beta}}^{-1})$ with

$$\begin{aligned}
 \boldsymbol{\mu}_{\boldsymbol{\beta}}^* &= \mathbf{H}_{\boldsymbol{\beta}}^{-1} X' \Sigma_y^{-1} \left(\mathbf{y} - \alpha \Lambda^{-\frac{1}{2}} \mathbf{u} \right), \\
 \mathbf{H}_{\boldsymbol{\beta}} &= X' \Sigma_y^{-1} + \frac{1}{c_0} I_k.
 \end{aligned}$$

• **Full conditional distribution of parameter α**

$$\begin{aligned}
 \pi(\alpha | y_I, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\alpha}, \mathcal{D}) &\propto \pi(\alpha | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\alpha}) \\
 &\propto p(\mathbf{y} | \mathbf{u}, \boldsymbol{\lambda}, \eta) \pi(\alpha) \\
 &\propto \exp \left\{ -\frac{1}{2} \left(-\mathbf{y}' \Sigma_y^{-1} \alpha \Lambda^{-\frac{1}{2}} \mathbf{u} + \boldsymbol{\beta}' X' \Sigma_y^{-1} \alpha \Lambda^{-\frac{1}{2}} \mathbf{u} \right. \right. \\
 &\quad \left. \left. - \mathbf{u}' \Lambda^{-\frac{1}{2}} \alpha \Sigma_y^{-1} \mathbf{y} + \alpha \mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} X \boldsymbol{\beta} \right. \right. \\
 &\quad \left. \left. + \alpha^2 \mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} \Lambda^{-\frac{1}{2}} \mathbf{u} + \frac{1}{c_1} \alpha^2 \right) \right\} \\
 &\propto \exp \left[-\frac{1}{2} \left\{ \alpha^2 \left(\mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} \Lambda^{-\frac{1}{2}} \mathbf{u} + \frac{1}{c_1} \right) \right. \right. \\
 &\quad \left. \left. - 2 \alpha \mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} Z \right\} \right].
 \end{aligned}$$

Thus, $\alpha | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\alpha} \sim N(\mu_{\alpha}^*, h_{\alpha}^{-1})$ where,

$$\begin{aligned}\mu_{\alpha}^* &= h_{\alpha}^{-1} \mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} Z, \\ h_{\alpha} &= \mathbf{u}' \Lambda^{-\frac{1}{2}} \Sigma_y^{-1} \Lambda^{-\frac{1}{2}} \mathbf{u} + \frac{1}{c_1}.\end{aligned}$$

- **Full conditional distribution of parameter σ^2** By assume that $\mathbf{b} = \mathbf{y} - X\boldsymbol{\beta} - \alpha\Lambda^{-\frac{1}{2}}\mathbf{u} = Z - \alpha\Lambda^{-\frac{1}{2}}\mathbf{u}$ and $\mathbf{B} = \Lambda^{-\frac{1}{2}}C_{\theta_w}\Lambda^{-\frac{1}{2}} + \omega^2\mathbf{I}_n$ we have

$$\begin{aligned}\pi(\sigma^2 | y_I, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\sigma^2}, \mathcal{D}) &\propto \pi(\sigma^2 | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\sigma^2}) \\ &\propto p(\mathbf{y} | \mathbf{u}, \boldsymbol{\lambda}, \eta) \pi(\sigma^2) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2} \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}\right\} |\sigma^2\mathbf{B}|^{-\frac{1}{2}} \\ &\quad \times \left(\frac{1}{\sigma^2}\right)^{c_2-1} \exp\left\{-\frac{c_3}{\sigma^2}\right\}.\end{aligned}$$

Since this full conditional distribution does not have any standard form, we use the Metropolis-Hastings method.

- **Full conditional distribution of parameter ω^2** By recalling \mathbf{b} and \mathbf{B} from the above item, we also have

$$\begin{aligned}\pi(\omega^2 | y_I, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\omega^2}, \mathcal{D}) &\propto \pi(\omega^2 | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\omega^2}) \\ &\propto p(\mathbf{y} | \mathbf{u}, \boldsymbol{\lambda}, \eta) \pi(\omega^2) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2} \mathbf{b}'\mathbf{B}^{-1}\mathbf{b}\right\} |\sigma^2\mathbf{B}|^{-\frac{1}{2}} \\ &\quad \times \omega^{-2} \exp\left\{-\frac{1}{2}(c_4^2\omega^{-2} + c_5^2\omega^2)\right\}\end{aligned}$$

and hence $\omega^2 | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\omega^2} \sim \text{GIG}\left(-\frac{n}{2}, \sqrt{c_4^2 + \frac{\mathbf{b}'\mathbf{B}^{-1}\mathbf{b}}{\sigma^2}}, c_5\right)$.

- **Full conditional distribution of parameter ν**

$$\begin{aligned}\pi(\nu | y_I, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\nu}, \mathcal{D}) &\propto \pi(\nu | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\nu}) \\ &\propto p(\mathbf{y} | \mathbf{u}, \boldsymbol{\lambda}, \eta) \pi(\boldsymbol{\lambda} | \mathbf{u}, \eta) \pi(\nu) \\ &\propto \pi(\boldsymbol{\lambda} | \eta) \pi(\nu) \\ &\propto \exp \left\{ -\frac{1}{2} \left(\boldsymbol{\psi} + \frac{\nu}{2} \mathbf{J} \right)' (\nu C_{\theta_\lambda})^{-1} \left(\boldsymbol{\psi} + \frac{\nu}{2} \mathbf{J} \right) \right\} \\ &\quad \times \frac{1}{\nu^{1+\frac{n}{2}}} \exp \left\{ -\frac{1}{2} (c_6^2 \nu^{-1} + c_7^2 \nu) \right\}.\end{aligned}$$

In this case, the Metropolis-Hastings method can be used for sampling from this full conditional as well. Recall that this full conditional also can be represent in terms of $\mathbf{a} = \ln \boldsymbol{\lambda} + \frac{\nu}{2} \mathbf{J} = \boldsymbol{\psi} + \frac{\nu}{2} \mathbf{J}$ in the penultimate line which is used in the next item. Therefore,

$$\pi(\nu | y_I, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\nu}, \mathcal{D}) \propto \frac{1}{\nu^{1+\frac{n}{2}}} \exp \left[-\frac{1}{2} \left\{ c_6^2 \nu^{-1} + c_7^2 \nu + \mathbf{a}' (\nu C_{\theta_\lambda})^{-1} \mathbf{a} \right\} \right].$$

- **Full conditional distributions of the correlation parameters θ_λ and θ_w** In terms of \mathbf{a} , \mathbf{b} and \mathbf{B} , we have

$$\begin{aligned}p(\theta_w | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\theta_w}) &\propto \exp \left\{ -\frac{c_8}{\text{med}(d)} \theta_w \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{b}' \mathbf{B}^{-1} \mathbf{b} \right\} |\sigma^2 \mathbf{B}|^{-\frac{1}{2}} \\ &\quad \times \exp \left\{ -\frac{1}{2} \mathbf{u}' C_{\theta_w}^{-1} \mathbf{u} \right\}\end{aligned}$$

and

$$\begin{aligned}p(\theta_\lambda | \mathbf{y}, \mathbf{u}, \boldsymbol{\lambda}, \eta_{-\theta_\lambda}) &\propto \exp \left\{ -\frac{c_9}{\text{med}(d)} \theta_\lambda \right\} \\ &\quad \times \exp \left\{ -\frac{1}{2} \mathbf{a}' (\nu C_{\theta_\lambda})^{-1} \mathbf{a} \right\}.\end{aligned}$$

These full conditional distributions are of nonstandard form, so a Metropolis-Hastings step would be used.

Appendix B: Rainfall Data

Table 8. Rainfall amounts of Fars province measured in inch, geographical locations, and elevation in meter.

station	precipitation	longitude	latitude	elevation
Abadeh	1.09345684	52.40	31.11	2030
Arsanjan	0.48489401	53.16	29.56	1703
Bavanat	0.21459461	53.40	30.28	2231
Darab	0	54.17	28.47	1098
Eqlid	1.29816493	52.38	30.54	2300
Estahban	0.41391564	54.02	29.05	1609
Farashband	0.37737203	52.06	28.48	782
Fasa	0.47246634	53.41	28.56	1288
Firuzabad	0	52.33	28.53	1362
Gerash	0.29725655	54.15	27.69	403
Jahrom	0.21737661	53.32	28.29	1082
Kavar	0.27805494	52.65	29.16	651
Kazerun	0.31700190	51.39	29.36	860
Kherameh	0.26663226	53.29	29.63	875
Khonj	0.55668826	53.40	27.98	511
Khorrambid	0.77678033	53.09	30.35	2251
Lamerd	0	53.12	27.22	405
Larestan	0	54.17	27.42	792
Mamasani	0.83764997	51.32	30.04	972
Marvdasht	0.51173630	52.54	29.56	1605
Mohr	0.23188202	52.88	27.55	659
Neyriz	0.58328311	54.20	29.12	1632
Pasargad	0.74427868	53.21	30.19	1614
Qir-o-Karzin	0	53.03	28.28	746
Rostam	1.49020491	51.51	30.25	864
Sarvestan	0.42793923	53.21	29.27	719
Sepidan	1.61606073	52.00	30.14	2201
Shiraz	0.68990878	52.36	29.32	1484
Zarghan	0.22756008	52.43	29.47	1596
Zarrindasht	0.34310578	54.25	28.21	1029

Appendix C: Details on the Used Distributions

An n -dimensional random vector Y is said to have a multivariate closed skew-normal distribution, denoted by $CSN_{n,m}(\boldsymbol{\mu}, \Sigma, D, \mathbf{v}, \Theta)$, if its density is $\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma) \Phi_m(D(\mathbf{y} - \boldsymbol{\mu}); \mathbf{v}, \Theta) / \Phi_m(\mathbf{0}; \mathbf{v}, \Theta + D\Sigma'D)$, where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\mathbf{v} \in \mathbb{R}^m$ and $\Sigma \in \mathbb{R}^{n \times n}$ and $\Theta \in \mathbb{R}^{m \times m}$ are both covariance matrices, $D \in \mathbb{R}^{m \times n}$, $\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma)$ and $\Phi_m(\mathbf{y}; \boldsymbol{\mu}, \Sigma)$ are the probability density function (pdf) and cumulative distribution function (cdf), respectively, of the n -dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ .

Arellano-Valle and Azzalini (2006) presented a family of skew-normal distributions which unifies a plethora of SN distributions including CSN with no over-parametrization problem. According to their approach, an n -dimensional random vector Y has a multivariate unified skew-normal distribution, denoted by $SUN_{n,m}(\boldsymbol{\mu}, \Sigma, \Gamma, \mathbf{v}, \Delta)$, if its density is

$$\phi_n(\mathbf{y}; \boldsymbol{\mu}, \Sigma) \Phi_m(\Gamma'\Sigma^{-1}(\mathbf{y} - \boldsymbol{\mu}); \mathbf{v}, \Delta - \Gamma'\Sigma^{-1}\Gamma) / \Phi_m(\mathbf{0}; \mathbf{v}, \Delta),$$

where $\Delta \in \mathbb{R}^{m \times m}$ is a correlation matrix, $\Gamma \in \mathbb{R}^{n \times m}$. The SUN and CSN classes are equivalent when $\Delta = \Theta + D\Sigma'D$ and $\Gamma = \Sigma D$.

Let $\{G(\mathbf{s}), \mathbf{s} \in R \subseteq \mathbb{R}^d\}$, $d \geq 1$, be a spatial, ergodic, stationary, zero-mean Gaussian process with stationary covariance function $c(\mathbf{h})$ and denoted the covariance matrix of the random vector $\mathbf{G} = (G(s_1), \dots, G(s_n))'$ by Σ . Also consider $V \sim N_n(\mathbf{0}, C_\theta)$ and $U \sim TN_n(\mathbf{0}; \mathbf{0}, C_\theta)$ are independent in which, $TN_n(\mathbf{c}; \boldsymbol{\mu}, \Sigma)$ denotes the $N_n(\boldsymbol{\mu}, \Sigma)$ distribution truncated below at a point \mathbf{c} . Then if $\mathbf{W}_\lambda \stackrel{d}{=} \alpha\Lambda^{-\frac{1}{2}}\mathbf{U} + \sigma\Lambda^{-\frac{1}{2}}\mathbf{V}$ and \mathbf{T} be a normal vector of dimension n , we have

$$\begin{pmatrix} \mathbf{T} \\ \mathbf{G} \end{pmatrix} \Big| \boldsymbol{\lambda} \sim N_{2n} \left(\mathbf{0}, \begin{bmatrix} C_\theta & \alpha C_\theta \Lambda^{-\frac{1}{2}} \\ \alpha \Lambda^{-\frac{1}{2}} C_\theta & \Sigma_w \end{bmatrix} \right).$$

Now, we define a $SUGLG$ process $\{W_\lambda(s)\}$ based on the following hierarchical representation:

1. $[W_\lambda(\mathbf{s}) | \boldsymbol{\lambda}(\mathbf{s})] \stackrel{d}{=} [G(\mathbf{s}) | \mathbf{T} > \mathbf{0}, \boldsymbol{\lambda}(\mathbf{s})]$,
2. $\boldsymbol{\lambda}(\mathbf{s})$ is a log-Gaussian stochastic process.

Vahid Tadayon

Department of Statistics,
Eghlid Higher Education Institute,
Eghid, Iran.
email: *vahidtadayon24@gmail.com*