Bayesian Analysis (2018) **13**, Number 2, pp. 531–557

Modeling Skewed Spatial Data Using a Convolution of Gaussian and Log-Gaussian Processes

Hamid Zareifard*, Majid Jafari Khaledi†,
Firoozeh Rivaz‡, and Mohammad Q. Vahidi-Asl§

Abstract. In spatial statistics, it is usual to consider a Gaussian process for spatial latent variables. As the data often exhibit non-normality, we introduce a novel skew process, named hereafter Gaussian-log Gaussian convolution (GLGC) to construct latent spatial models which provide great flexibility in capturing skewness. Some properties including closed-form expressions for the moments and the skewness of the GLGC process are derived. Particularly, we show that the mean square continuity and differentiability of the GLGC process are established by those of the Gaussian and log-Gaussian processes considered in its structure. Moreover, the usefulness of the proposed approach is demonstrated through the analysis of spatial data, including mixed ordinal and continuous outcomes that are jointly modeled through a common latent process. A fully Bayesian analysis is adopted to make inference. Our methodology is illustrated with simulation experiments as well as an environmental data set.

Keywords: spatial process, skewness, slice sampling, mixed outcomes, misalignment.

1 Introduction

A popular approach for analyzing spatial data involves the introduction of latent variables. Indeed, this approach creates a statistical tool for describing the underlying structure in the responses. The traditional latent variable models rely on the assumption that the latent random variables are normally distributed (see, e.g., Higgs and Hoeting (2010); Chagneau et al. (2010); Schliep and Hoeting (2013), and references therein). Although the use of the latent Gaussian models facilitates spatial analysis, the normality assumption might be overly restrictive in obtaining an accurate representation of the data structure. In applications, the distribution of the data is frequently skewed and consequently there is a need for flexible and computationally tractable ways to remove the Gaussianity assumption and provide robust estimations and predictions. Accordingly, this work aims to present a skew spatial model using a family of distributions that is analytically tractable, accommodates practical values of asymmetry, and includes the normal distribution as a special case.

*Department of Statistics, Jahrom University, Jahrom 74137-66171, Iran, zareifard@jahrom.ac.ir
†Department of Statistics, Tarbiat Modares University, Tehran, Iran, jafari-m@modares.ac.ir
‡Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran
§Department of Statistics, Faculty of Mathematical Sciences, Shahid Beheshti University, Tehran, Iran

 https://doi.org/10.1214/17-BA1064

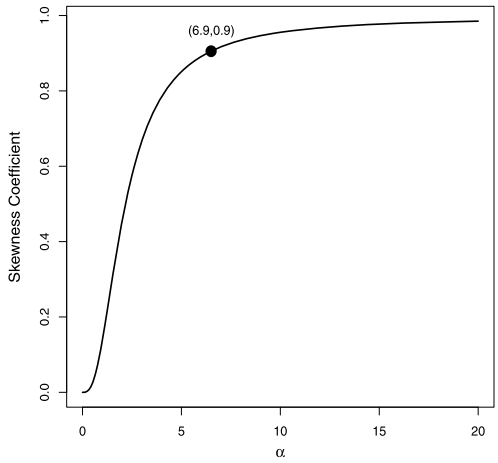

Figure 1: Plot of the skewness coefficient versus α for the univariate skew-normal distribution.

The interest in skew spatial models based on the use of skew-normal distributions (Azzalini, 1985; Azzalini and Capitanio, 1999) has been prevalent in the literature in recent years (e.g., Kim and Mallick, 2004; Allard and Naveau, 2007). Particularly, Zhang and El-Shaarawi (2010) used a stochastic representation of skew-normal distribution to define a stationary skew-Gaussian process with skewed marginal distributions. Zareifard and Jafari Khaledi (2013) recently applied the unified skew-normal distribution (Arellano-Valle and Azzalini, 2006) which unifies a plethora of recent skew-normal models. Generally the stochastic representation of skew-normal variables is based on a convolution of half normal and normal variables as $\alpha|U|+Z$, where α denotes the skewness parameter and the two independent random variables Z and U have normal distributions with zero means. Notice that the random variable $W = |U|$ has a normal distribution truncated below at zero. The skew-Gaussian spatial process may suffer from the fact that the skew-normal distribution is only capable of capturing mild amounts of skewness Azzalini (1985). As seen in Figure 1, the skewness coefficient for the skew-normal distribution rapidly increases from 0 to 0.9 as the skewness parameter goes from zero to 6.5 and then it tends to 1 asymptotically as the skewness parameter tends to infinity.

To overcome the limitation of the existing approach described above, we propose to consider another family of asymmetric distributions. An alternative method of generating skewed distributions is to consider $\alpha W + Z$, where Z has a symmetric distribution, but the non-negative variable W has a specified skewed distribution rather than being truncated. An example of this is the normal-exponential convolution distribution (Aigner et al., 1977; Silver et al., 2009) in which Z is normal and W has an exponential distribution. An extension of this is the normal-gamma convolution distribution introduced by Greene (1990) which includes the normal-exponential convolution distribution as a special case.

Since a convolution of log-normal and normal variables has several desirable properties such as simplicity, flexibility and being easily incorporated into the spatial model,

we develop this approach to construct a random process which is able to efficiently capture skewness. To be more specific, we propose the so-called Gaussian log-Gaussian convolution (GLGC) model based on a convolution of the log-Gaussian and Gaussian processes. As limiting cases, this model includes both well known spatial models (i.e. log-Gaussian as well as Gaussian). In comparison to the existing skew-Gaussian models, the proposed model could capture a greater amount of skewness. Interestingly, the skewness would be a strictly increasing function of α.

Motivated by analyzing a data set with mixed ordinal and continuous outcomes that are used to evaluate an unobservable process, we develop a GLGC factor model to account for the inherent skewness in the data. Specifically, our skew spatial model is built on a hierarchical structure in which latent variables are introduced at two levels within the hierarchy (see for instance Royle and Berliner (1999); Recta et al. (2012); Schliep and Hoeting (2013)). Accordingly, we assume that a skew spatial factor is responsible for the dependencies observed in the data. Since not all variables have been observed at all locations, the model is extended in a way that accounts for the spatial misalignment which has been a long standing issue in the literature (see for instance Gelfand, 2010). Our modeling framework could clearly benefit from a Bayesian estimation scheme since that framework easily handles unobserved variables. In fact, the latent variables can naturally be incorporated as additional parameters in the model using data augmentation (Tanner and Wong, 1987). In order to facilitate Bayesian inference, a Markov chain Monte Carlo (MCMC) sampling approach is implemented. In particular, a slice sampling algorithm (Neal, 2003; Agarwal and Gelfand, 2005) coupled with the Gibbs scheme is developed. Via simulation, we also investigate the impact of latent variable model misspecification on prediction. We found that prediction accuracy of the latent process can be influenced by the violation of the Gaussian assumption.

The rest of the paper is organized as follows. Section 2 introduces the new class of skewed spatial models. In Section 3, the skew spatial factor model which relies on the GLGC process is presented. Section 4 discusses the Bayesian analysis of the model. Sections 5 and 6 illustrate the use of the proposed methodology with simulated and real spatial data sets. Finally, in Section 7 we present some conclusions and final remarks. In this section, we also provide a brief discussion regarding how to extend the univariate skew model to the multivariate case.

2 Skew spatial models

Assume that the data are a subset of the spatial process $Y(\cdot) = \{Y(\mathbf{s}), \mathbf{s} \in D \subseteq \Re^2\}$, which depends on a latent spatial process $H(\cdot) = \{H(\mathbf{s}), \mathbf{s} \in D \subseteq \Re^2\}$ such that

$$Y(\mathbf{s}) = H(\mathbf{s}) + \rho(\mathbf{s}), \tag{1}$$

where $\rho(\mathbf{s})$ is an iid (white noise) measurement error process that is independent of $H(\cdot)$ and has variance τ^2. We define the n-dimensional vector of data as $\mathbf{Y} = (Y(\mathbf{s}_1), \cdots, Y(\mathbf{s}_n))'$ at spatial locations $\mathbf{s}_1, \cdots, \mathbf{s}_n$. Similarly for the process $H(\cdot)$, we define the n-dimensional vector $\mathbf{H} = (H(\mathbf{s}_1), \cdots, H(\mathbf{s}_n))'$. Customary modeling of spatial data proceeds from a Gaussian model for latent process $H(\cdot)$ of the form

$$H(\mathbf{s}) = \mathbf{f}'(\mathbf{s})\boldsymbol{\beta} + Z(\mathbf{s}), \tag{2}$$

where $f'(\mathbf{s})\boldsymbol{\beta}$ describes the spatial trend through a linear model with k covariates $\mathbf{f}'(\mathbf{s}) = (f_1(\mathbf{s}), \cdots, f_k(\mathbf{s}))$. Moreover, $Z(\mathbf{s})$ is a second-order stationary Gaussian process with zero mean, variance σ_1^2, and the isotropic correlation function:

$$Corr(Z(\mathbf{s}_i), Z(\mathbf{s}_j)) = C_{\boldsymbol{\theta}_1}(||\mathbf{s}_i - \mathbf{s}_j||),$$

where $C_{\boldsymbol{\theta}_1}(d)$ is a valid correlation function, parameterized by a vector $\boldsymbol{\theta}_1$. We could extend the model given in (2) to non-Gaussianity by including a process $W(\mathbf{s})$ which is non-negative and independent of $Z(\mathbf{s})$ at each location as

$$H(\mathbf{s}) = \mathbf{f}'(\mathbf{s})\boldsymbol{\beta} + \alpha W(\mathbf{s}) + Z(\mathbf{s}), \tag{3}$$

where $\alpha \in \Re$ is the skewness parameter. This general framework involves some of the skew-Gaussian processes, which have been introduced in the literature. For instance, by considering $W(\mathbf{s}) = |U(\mathbf{s})|$ where $U(\mathbf{s})$ is a second-order stationary Gaussian processes with zero mean, variance 1, and the correlation function

$$Corr(U(\mathbf{s}_i), U(\mathbf{s}_j)) = C_{\boldsymbol{\theta}_2}(||\mathbf{s}_i - \mathbf{s}_j||),$$

we arrive at the model proposed by Zhang and El-Shaarawi (2010). Although this model leads to a wide variety of marginal distributions, which possess rich theoretical properties, it faces two major drawbacks:

1. Since the absolute function is not monotone, the data usually do not provide sufficient information to obtain a precise estimate of the parameters of the unobservable latent variable $U(\mathbf{s})$.

2. The mean square differentiability of the process $H(\mathbf{s})$ is not guaranteed by that of processes $U(\mathbf{s})$ and $Z(\mathbf{s})$. In other words, even though it turns out that the processes $U(\mathbf{s})$ and $Z(\mathbf{s})$ are q times mean square differentiable ($q \geq 1$), the process $H(\mathbf{s})$ does not necessarily satisfy this property.

On the other hand, if we define $W(\mathbf{s}) \stackrel{d}{=} [U(\mathbf{s})|U(\mathbf{s}) > 0]$, the SUN model proposed by Zareifard and Jafari Khaledi (2013) will be acquired. Since the aforementioned models are constructed based on the stochastic representation of the skew-normal variables, the degree of skewness represented by these models are somewhat limited as described in Section 1 above. In what is to follow, we develop a new class of skew spatial models relying on the log-normal distribution which is of major importance in probability and statistics.

2.1 The GLGC model

Our starting point is to consider a log-Gaussian process for $W(\mathbf{s})$ as

$$W(\mathbf{s}) = \exp(U(\mathbf{s})). \tag{4}$$

Consequently, the spatial process $H(\mathbf{s})$ given in (3) and (4) is a convolution of Gaussian and log-Gaussian processes, so we shall call it the GLGC process. Hereafter, we denote

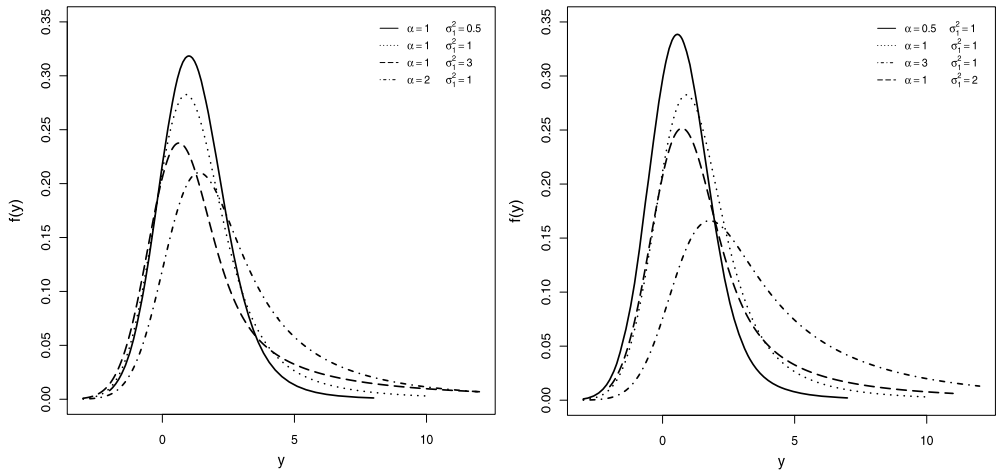

Figure 2: The probability density function of $H(\mathbf{s})$ for $\mu(\mathbf{s}) = 0$, $\sigma_2^2 = 1$ and some different values for α and σ_1^2. The legends indicate the values considered for α and σ_1^2.

the variance and correlation parameters of the Gaussian process $U(\mathbf{s})$ as σ_1^2 and $\boldsymbol{\theta}_1$, respectively, as well as those for the Gaussian process $Z(\mathbf{s})$ as σ_2^2 and $\boldsymbol{\theta}_2$. As we demonstrate further, considering the variance parameter σ_1^2 in the structure of the GLGC process increases its ability in effectively capturing skewness. We now explore the properties and the flexibility of the GLGC process. Interestingly, when $\alpha \longrightarrow 0$ and $\sigma_2^2 \longrightarrow 0$ two important spatial models, i.e. the Gaussian model (2) and the log-Gaussian model

$$H(\mathbf{s}) = \mathbf{f}'(\mathbf{s})\boldsymbol{\beta} + \alpha W(\mathbf{s}),$$

arise as limiting cases of the GLGC model, respectively.

The distribution function of $H(\mathbf{s})$ is a convolution of normal and log-normal distributions as

$$F_{H(\mathbf{s})}(t) = \int_0^{+\infty} \Phi\Big(\frac{t - \mu(\mathbf{s}) - \alpha w}{\sigma_2}\Big) f_{LN}(w|0, \sigma_1^2)dw, \quad t \in \Re,$$

where $\mu(\mathbf{s}) = \mathbf{f}'(\mathbf{s})\boldsymbol{\beta}$, $\Phi(\cdot)$ represents the cumulative distribution function of a standard normal distribution. Also, $f_{LN}(\cdot|\mu, \sigma^2)$ denotes the probability density function of log-normal distribution with location μ and scale σ^2. Figure 2 shows the marginal density function of GLGC model under $\mu(\mathbf{s}) = 0$, $\sigma_2^2 = 1$ and some different pairs for (α, σ_1^2) that are given in the figure legend. It can be seen that the two parameters α and σ_1^2 affect the shape of density function in different ways which leads us to a flexible class of stationary skew processes. In Figure 3, the effect of parameter σ_2^2 on the shape of the density function of GLGC model has been considered. The density functions are plotted under $\mu(\mathbf{s}) = 0$, $\alpha = \sigma_1^2 = 1$, and four different values for σ_2^2 (i.e., $\sigma_2^2 = 0.1, 0.3, 1, 2$). This figure clearly shows that the GLGC model is close to the log-Gaussian model for small values of σ_2^2.

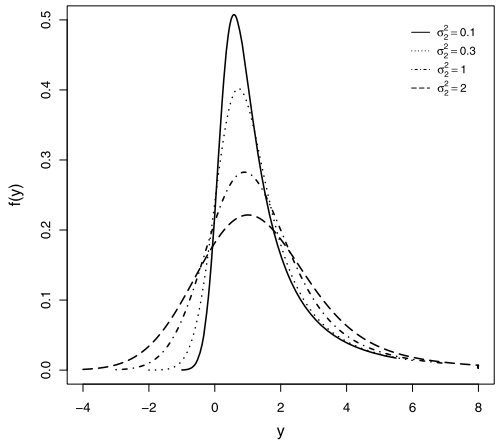

Figure 3: The probability density function of $H(\mathbf{s})$ for $\mu(\mathbf{s}) = 0$, $\alpha = 1$, $\sigma_1^2 = 1$ and some different values for σ_2^2 The legends indicate the values considered for σ_2^2.

One of the main advantages of the GLGC process comes from the fact that the rth moment of the GLGC process can be obtained in closed form as presented in the following proposition.

Proposition. *Let $H(\cdot)$ be the GLGC process given in (3) and (4). Then the rth moment of $H(\mathbf{s})$ around the mean surface $\mu(\mathbf{s})$ is obtained through*

$$E\big(H(\mathbf{s}) - \mu(\mathbf{s})\big)^r = \sum_{i=0}^{[r/2]} \alpha^{r-2i} \binom{r}{2i} e^{\frac{1}{2}(r-2i)^2\sigma_1^2} \sigma_2^{2i} \prod_{j=1}^{\max\{i,1\}} |2i - 2j + 1|, \tag{5}$$

where $r > 0$ takes integer values.

Proof. Using the binomial expansion, we can write

$$E\big(H(\mathbf{s}) - \mu(\mathbf{s})\big)^r = \sum_{i=0}^{r} \binom{r}{i} \alpha^{r-i} E(e^{(r-i)U(\mathbf{s})}) E(Z^i(\mathbf{s})).$$

Now since the odd moments of the standard normal distribution are zero, we have clearly

$$\begin{aligned} E\big(H(\mathbf{s}) - \mu(\mathbf{s})\big)^r &= \sum_{i=0}^{[r/2]} \binom{r}{2i} \alpha^{r-2i} E(e^{(r-2i)U(\mathbf{s})}) E(Z^{2i}(\mathbf{s})) \\ &= \sum_{i=0}^{[r/2]} \binom{r}{2i} \alpha^{r-2i} e^{\frac{1}{2}(r-2i)^2\sigma_1^2} \sigma_2^{2i} \prod_{j=1}^{\max\{i,1\}} |2i - 2j + 1|. \end{aligned}$$

□

As a result, the mean and variance of the proposed model are given by

$$\begin{aligned} E(H(\mathbf{s})) &= \mu(\mathbf{s}) + \alpha e^{\frac{1}{2}\sigma_1^2}, \\ Var(H(\mathbf{s})) &= \alpha^2(e^{2\sigma_1^2} - e^{\sigma_1^2}) + \sigma_2^2, \end{aligned}$$

respectively. Even though $H(\mathbf{s})$ has finite moments, similarly with the log-normal random variable, the moment generating function of $H(\mathbf{s})$ is infinite at any positive number. Also, the characteristic function (CF) of $H(\mathbf{s})$ is given by

$$CF_{\mathbf{s}}(t) = e^{it\mu(\mathbf{s}) - \frac{1}{2}\sigma_2^2 t^2} E(e^{it\alpha e^{U(\mathbf{s})}}) = e^{it\mu(\mathbf{s}) - \frac{1}{2}\sigma_2^2 t^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{it\alpha e^u - \frac{1}{2\pi\sigma_1^2}u^2} du. \quad (6)$$

Since no closed form expressions exist for the right hand integral in (6) (Johnson et al., 1994), the computation of $CF_{\mathbf{s}}(t)$ is a challenging problem. In such a situation it is desirable to approximate the CF function. A wide variety of methods have been developed to approximate the characteristic function of the log-normal distribution (see, e.g., Barouch et al., 1986; Asmussen et al., 2016). These approximations can also be applied for the integral involved in the characteristic function of $H(\mathbf{s})$.

From the marginal moments of $H(\mathbf{s})$ given in (5), we can immediately find an explicit expression for the skewness coefficient of the proposed model as

$$\gamma = \alpha^3 \frac{e^{\frac{3}{2}\sigma_1^2}(e^{3\sigma_1^2} - 3e^{\sigma_1^2} + 2)}{(\alpha^2(e^{2\sigma_1^2} - e^{\sigma_1^2}) + \sigma_2^2)^{\frac{3}{2}}}. \quad (7)$$

Consequently, for $\alpha = 0$, $\alpha < 0$ and $\alpha > 0$, the distribution of $H(\mathbf{s})$ is normal, skewed to the left and skewed to the right, respectively. Moreover, it can be seen that γ is a strictly increasing function in terms of α. Also, the absolute value of γ is a strictly increasing function of σ_1^2. Therefore, the value of skewness is controlled by both parameters α and σ_1^2. This supports the claim that the GLGC model, unlike the skew-Gaussian models, can certainly generate a greater amount of skewness. To emphasize this, the log-normal and skew-normal probability plots of $n = 1000$ simulated data from the proposed model is drawn in Figures 4 and 5. These figures clearly show that the data do not lie exactly along a straight line, indicating lack of fit. To be more specific, there is still a lot of skewness left, and so it can be concluded the log-normal and skew-normal distributions have serious limitations in taking skewness into account. Notice that under the considered values for the parameters, the skewness coefficients are $\gamma = 4.6$ and $\gamma = 11.6$, respectively, which are medium and relatively large values.

Also, in Figures 6 and 7 the GLGC probability plots of $n = 1000$ *i.i.d* data sets from two below models:

- Skew-normal Model: $H = \beta + \alpha|W| + Z; \quad W \sim N(0,1); \quad Z \sim N(0, \sigma_2^2)$
- Log-normal Model: $H = \beta + \alpha W; \qquad W \sim LN(0, \sigma_1^2)$

under $\beta = 0$, $\sigma_1 = \sigma_2 = 1$ and two different values $\alpha = 1$ as well as $\alpha = 1.5$ are plotted. The results shown in these figures are very interesting and informative. They

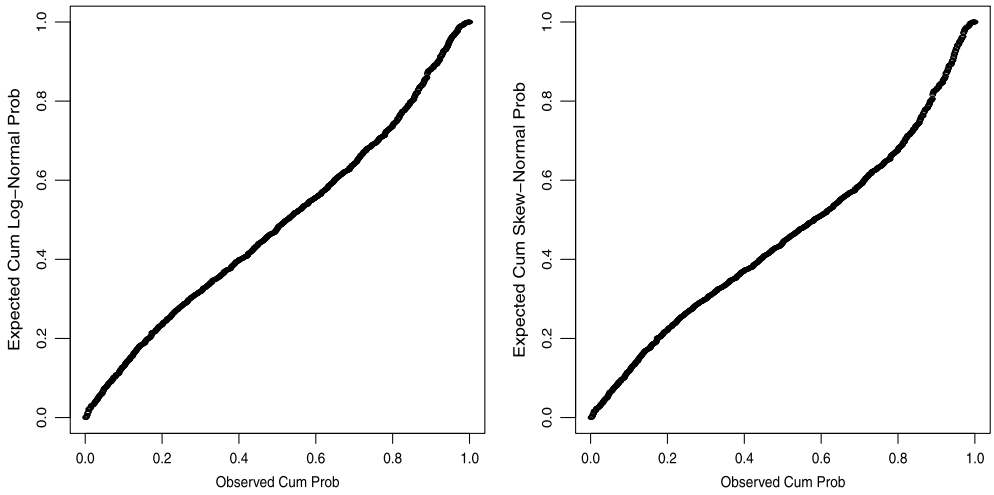

Figure 4: The log-normal and skew-normal probability plots of $n = 1000$ simulated data from the proposed model under $\beta = 0$, $\sigma_2^2 = 1$ and $\alpha = \sigma_1^2 = 1$.

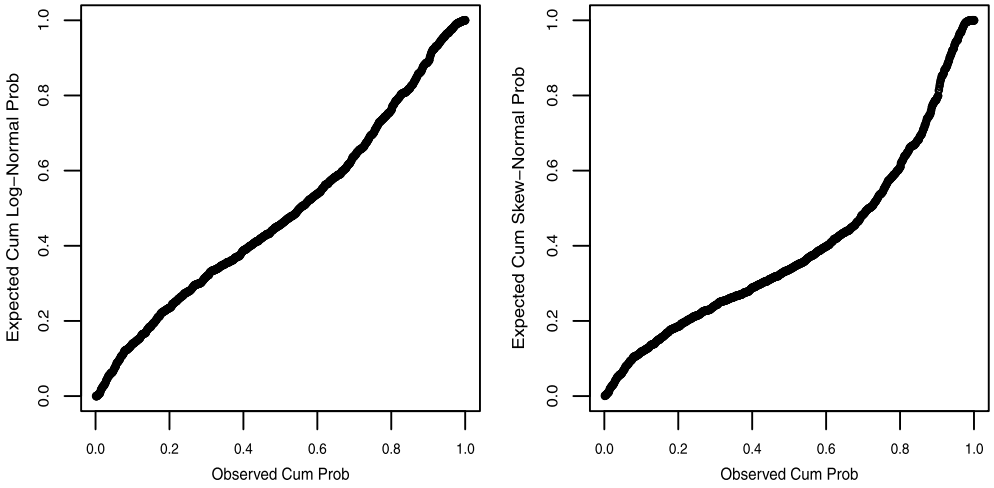

Figure 5: The log-normal and skew-normal probability plots of $n = 1000$ simulated data from the proposed model under $\beta = 0$, $\sigma_2^2 = 1$ and $\alpha = \sigma_1^2 = 1.5$.

obviously indicate that the GLGC model has a good performance even for a data set generated from the skew-normal model. This is not surprising because with regard to (7) the GLGC model is able to capture different distribution shapes which are observed in skewed data. It must be noted that although these results have been reported based on only one simulation, our experiments with more simulations and different values for α consistently confirm the great flexibility of this model in accommodating skewness.

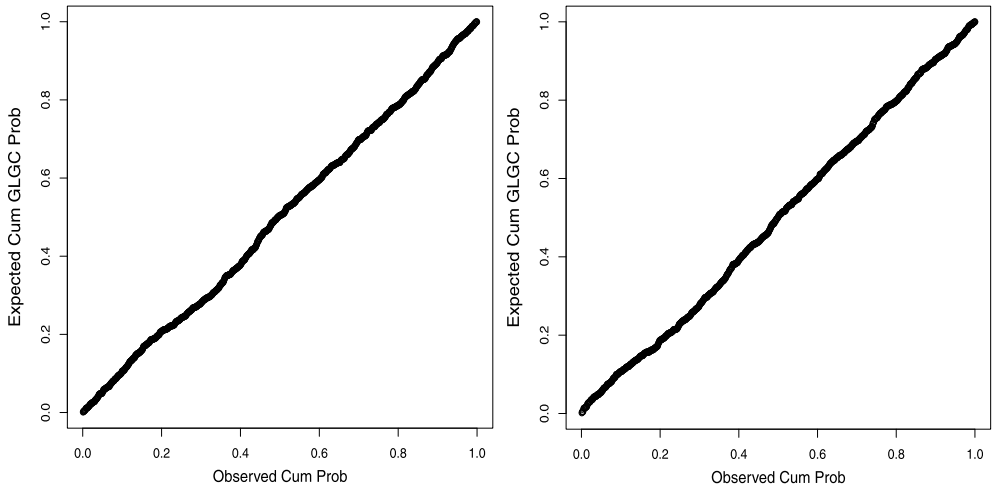

Figure 6: The GLGC probability plots of $n = 1000$ simulated data from the skew-normal model under $\beta = 0$, $\sigma_1 = \sigma_2 = 1$ and two different values for α. The left plot is under $\alpha = 1$ and the right plot is under $\alpha = 1.5$.

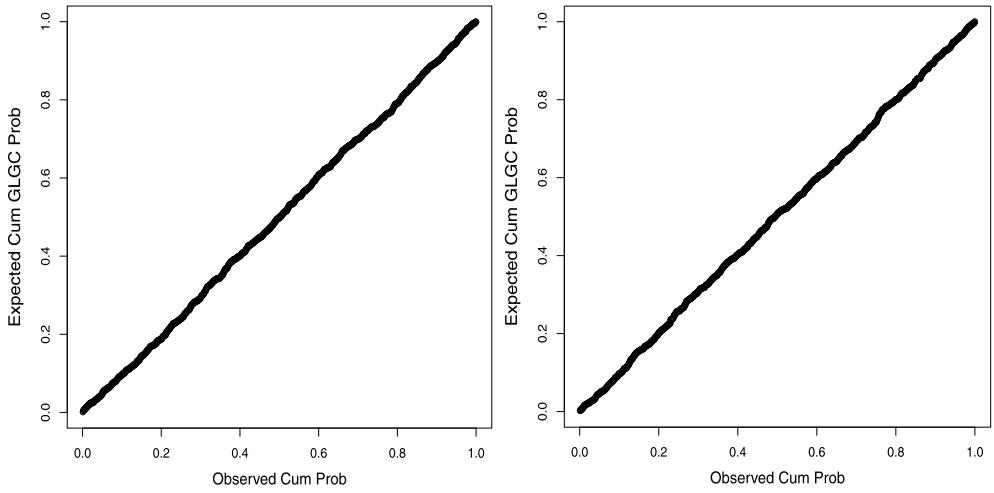

Figure 7: The GLGC probability plots of $n = 1000$ simulated data from the log-normal model under $\beta = 0$, $\sigma_1^2 = 1$ and two different values for α. The left plot is under $\alpha = 1$ and the right plot is under $\alpha = 1.5$.

The marginal distribution of the latent vector $\mathbf{H}$ can be written as:

$$p_{\boldsymbol{\psi}}(\mathbf{H}) = \int_{\Re_n^+} f_N^n(\mathbf{H}|\mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{w}, \sigma_2^2\mathbf{C}_{\boldsymbol{\theta}_2}) f_{LN}^n(\mathbf{w}|0, \sigma_1^2\mathbf{C}_{\boldsymbol{\theta}_1}) d\mathbf{w}, \tag{8}$$

where $\boldsymbol{\psi} = (\boldsymbol{\beta}, \alpha, \sigma_1^2, \sigma_2^2, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$, $\mathbf{X} = (\mathbf{f}(\mathbf{s}_1), \cdots, \mathbf{f}(\mathbf{s}_n))'$, $\mathbf{W} = (W(\mathbf{s}_1), \cdots, W(\mathbf{s}_n))'$ and $\mathbf{C}_{\boldsymbol{\theta}_l}$, $l = 1, 2$, is the $n \times n$ correlation matrix with $C_{\boldsymbol{\theta}_l}(||\mathbf{s}_i - \mathbf{s}_j||)$ as its (i, j)th element. Also, $f_N^n(\cdot|\boldsymbol{\mu}_1, \boldsymbol{\Omega}_1)$ and $f_{LN}^n(\cdot|\boldsymbol{\mu}_2, \boldsymbol{\Omega}_2)$ denote the probability density functions of n-variate normal and log-normal distributions with means $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ and covariance matrices $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, respectively. We can easily verify that Kolmogorov consistency conditions hold for this sampling model. We denote the distribution of the n-dimensional random vector $\mathbf{H}$ by $GLGC_n(\mathbf{X}\boldsymbol{\beta}, \sigma_1^2\mathbf{C}_{\boldsymbol{\theta}_1}, \sigma_2^2\mathbf{C}_{\boldsymbol{\theta}_2}, \alpha)$ as well as its density function by $f_{GLGC}^n(\mathbf{h}|\mathbf{X}\boldsymbol{\beta}, \sigma_1^2\mathbf{C}_{\boldsymbol{\theta}_1}, \sigma_2^2\mathbf{C}_{\boldsymbol{\theta}_2}, \alpha)$.

The mean square continuity of the GLGC process is established by the fact that its correlation function is continuous at 0. To show this, we first note that the correlation between $H(\mathbf{s}_i)$ and $H(\mathbf{s}_j)$ is

$$Corr(H(\mathbf{s}_i), H(\mathbf{s}_j)) = \frac{\alpha^2(e^{\sigma_1^2(1+C_{\boldsymbol{\theta}_1}(d))} - e^{\sigma_1^2}) + \sigma_2^2 C_{\boldsymbol{\theta}_2}(d)}{\alpha^2(e^{2\sigma_1^2} - e^{\sigma_1^2}) + \sigma_2^2},$$

where $d = ||\mathbf{s}_i - \mathbf{s}_j||$. Thus, in the case of continuity $C_{\boldsymbol{\theta}_1}(\cdot)$ and $C_{\boldsymbol{\theta}_2}(\cdot)$ at the origin, we can see when d tends to zero, then the correlation between $H(\mathbf{s}_i)$ and $H(\mathbf{s}_j)$ tends to 1 which shows the mean square continuity of the process $H(\mathbf{s})$ in the absence of a nugget effect. To determine the degree of smoothness of the process, the degree of mean square differentiability can be considered. Stein (1999) showed that a process is q times mean square differentiable if and only if the $2q$th derivative of the covariance function at 0 exists and is finite. Since the covariance function of the GLGC process is of the form

$$C(d) = \alpha^2(e^{\sigma_1^2(1+C_{\boldsymbol{\theta}_1}(d))} - e^{\sigma_1^2}) + \sigma_2^2 C_{\boldsymbol{\theta}_2}(d), \tag{9}$$

it can easily be seen that the degree of mean square differentiability of the process is equal to the degree of mean square differentiability of the underlying processes $U(\mathbf{s})$ and $Z(\mathbf{s})$. To be more specific, if the processes $U(\mathbf{s})$ and $Z(\mathbf{s})$ are q times mean square differentiable, then the process $H(\mathbf{s})$ also satisfies this property. Therefore, the smoothness properties of $U(\mathbf{s})$ and $Z(\mathbf{s})$ carry over to the process $H(\mathbf{s})$.

As our working family of isotropic correlation functions for the underlying processes $U(\mathbf{s})$ and $Z(\mathbf{s})$, we use the Whittle correlation function (Whittle, 1954)

$$C_\theta(d) = \frac{d}{\theta} K_1\left(\frac{d}{\theta}\right), \quad \theta > 0, \tag{10}$$

where θ is the range parameter that controls how fast the correlation decays with distance and $K_1(\cdot)$ is the modified Bessel function of the second kind of the first order. This model is, in fact, Whittle's generalization of the exponential correlation function. We may further concentrate on some other families of correlation functions and then easily choose an appropriate model via a model selection process. However, in order to reduce the complexity of the current discussion, we have not addressed this task in our simulations or applied example.

Parameter	True values	$\alpha = 1,\ \sigma_1^2 = 1$	$\alpha = 2,\ \sigma_1^2 = 2$
α		$1.35_{(0.81,1.76)}$	$1.89_{(0.67,2.96)}$
σ_1^2		$0.78_{(0.23,1.73)}$	$2.28_{(1.79,3.38)}$
θ_1	1	$1.07_{(0.52,1.69)}$	$1.14_{(0.74,1.72)}$
σ_2^2	2	$1.62_{(0.37,2.48)}$	$2.80_{(0.92,4.85)}$
θ_2	1	$0.83_{(0.38,1.41)}$	$0.66_{(0.47,0.90)}$
τ^2	1	$0.63_{(0.46,1.14)}$	$0.61_{(0.31,1.56)}$
β_0	5	$4.34_{(4.25,4.64)}$	$5.45_{(4.86,5.93)}$
β_1	-1	$-0.71_{(-1.28,-0.15)}$	$-0.83_{(-1.32,-0.25)}$
β_2	1	$0.78_{(0.18,1.39)}$	$0.60_{(0.16,1.15)}$

Table 1: Posterior mean (95% credible interval) for the model parameters.

2.2 Model assessment

To examine the performance of the proposed model, we conducted a simple simulation study. Throughout, we use a sample size of $n = 100$ points which were selected from a $[0, 5] \times [0, 5]$ regular grid. We generated data from the GLGC model with $\sigma_2^2 = 2$, $\theta_1 = \theta_2 = 1$ and $\tau^2 = 1$. In addition, assuming that $\mathbf{s} = (s_1, s_2)$, we consider the mean function $\mu(\mathbf{s}) = \beta_0 + \beta_1 s_1 + \beta_2 s_2$ with $\boldsymbol{\beta} = (5, -1, 1)$. To assess the identifiability of skewness parameters (i.e. α and σ_1^2), two data sets were generated by assuming $\alpha = \sigma_1^2 = 1$ and $\alpha = \sigma_1^2 = 2$. We fit the GLGC model to the simulated data using Bayesian inference. Details have been omitted, though based on the techniques provided subsequently in Section 4, one can easily arrive at them. Table 1 displays some posterior results for the parameters. It is clear that these parameters are all identifiable from the data and hence they do not suffer from the non-identifiability problem. In addition, the results show that the GLGC model is able to recover the true parameters used in simulating the data. Now, in order to investigate how much better our model fits relative to the two models below:

- Log-Gaussian Model: $\mathbf{H} = \mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{W}; \qquad \mathbf{W} \sim LN_n(0, \sigma_1^2\mathbf{C}_{\theta_1})$
- SUN Model: $\mathbf{H} = \mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{W} + \mathbf{Z}; \quad \mathbf{W} \sim TN_n(0; 0, \mathbf{C}_{\theta_1}); \quad \mathbf{Z} \sim N_n(0, \sigma_2^2\mathbf{C}_{\theta_2})$

where $TN_n(\mathbf{c}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the $N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ distribution truncated below at a point $\mathbf{c}$, the marginal likelihoods of data for GLGC, log-Gaussian and SUN models are first computed using the modified harmonic mean estimator $\hat{p}_4$ of Newton and Raftery (1994). Then we compute Bayes factors between models (see Kass and Raftery, 1995). As seen from Table 2, the GLGC model noticeably outperforms the two other models.

3 A GLGC factor model

In some applied problems, several types of measurements to evaluate an unobservable process are collected from sampling locations (see e.g., Schliep and Hoeting, 2013; Hogan and Tchernis, 2004). In such a situation, one way to jointly model the different types of

$\alpha = \sigma_1^2 = 1$	The Bayes factor in favor of the GLGC versus the log-Gaussian	1.3×10^9
	The Bayes factor in favor of the GLGC versus the SUN	7.9×10^6
$\alpha = \sigma_1^2 = 2$	The Bayes factor in favor of the GLGC versus the log-Gaussian	2.6×10^4
	The Bayes factor in favor of the GLGC versus the SUN	7.9×10^{13}

Table 2: The Bayes factor in favor of the GLGC model.

data is to consider a low-dimensional spatial model using latent factors. Classical spatial factor models rely on the assumption that factors are normally distributed which may be a restrictive assumption. Generally, in some applications, considering a skewed distribution for underlying latent variables can significantly improve the model fit (see for instance Chen et al. (1999) and Chen (2004)). In this section, we use the GLGC process as an underlying process to develop a skew spatial factor model. Hence, the resulting model inherits the advantages of the GLGC model. More specifically, let $J_o(\geq 1)$ and $J_c(\geq 1)$ denote the number of ordinal and continuous response variables, respectively. Therefore $J = J_o + J_c$ is the number of all considered responses on the spatial domain of interest D. We assume that the ordinal random process $Y_j(\mathbf{s})$, $j \in \{1, \cdots, J_o\}$, is created by clipping a latent process, say $T_j(\mathbf{s})$, meaning

$$Y_j(\mathbf{s}) = \sum_{l=1}^{L} l I_{\{\kappa_{j,l-1} < T_j(\mathbf{s}) \leq \kappa_{j,l}\}}, \tag{11}$$

where I denotes the indicator function, and $\kappa_{j,0}, \cdots, \kappa_{j,L}$ are the ordered cut-points defined by

$$-\infty = \kappa_{j,0} < \kappa_{j,1} < \cdots < \kappa_{j,L} = +\infty. \tag{12}$$

Note that the model can easily be generalized to include ordinal variables with a varying number of categories, e.g. $Y_j(\mathbf{s}) \in \{1, \cdots, L_j\}$. Further details of this discussion can be found in Schliep and Hoeting (2013). Also for all continuous variables j in $J_o+1, \cdots, J$, we assume that $Y_j(\mathbf{s}) \stackrel{d}{=} T_j(\mathbf{s})$. We now consider a linear relationship between each of the latent continuous processes (introduced in the first level of latency) and a spatial factor process $H(\mathbf{s})$ in which covariates are involved in its mean structure. More specifically, we assume that

$$T_j(\mathbf{s}) = \nu_j + \omega_j H(\mathbf{s}) + \rho_j(\mathbf{s}), \quad j = 1, \cdots, J, \tag{13}$$

where the mean of $T_j(\mathbf{s})$ is a metric-specific linear combination of the latent process $H(\mathbf{s})$, and $\rho_j(\mathbf{s})$ is an uncorrelated Gaussian process that is independent of $H(\cdot)$ and has zero mean and variance τ_j^2, representing the nugget effect. The fixed effect ν_j is the intercept and the fixed effect ω_j is the factor loading of the spatial process $H(\mathbf{s})$ which allows us to quantify the relationship between each of the response variables and $H(\mathbf{s})$.

Here since not all the variables have been observed over all locations, we suppose that a single realization of the j'th considered random process at n_j locations $\mathbf{s}_{j_1}, \cdots, \mathbf{s}_{j_{n_j}}$ is observed as $\mathbf{Y}_j = (Y_j(\mathbf{s}_{j_1}), \cdots, Y_j(\mathbf{s}_{j_{n_j}}))'$ for j in $1, \cdots, J$. Now we denote $\mathbf{s} = (\mathbf{s}_1, \cdots, \mathbf{s}_n)$ as all sampling locations that the variables are observed. These locations can then be shared across the variables observed at the same location. If we define $\mathbf{H} = (H(\mathbf{s}_1), \cdots, H(\mathbf{s}_n))'$ that is being shared across the variables and $\mathbf{T}_j = (T_j(\mathbf{s}_{j_1}), \cdots, T_j(\mathbf{s}_{j_{n_j}}))'$, then

$$\mathbf{T}_j|\mathbf{H} \sim N_{n_j}(\nu_j \mathbf{1}_{n_j} + \omega_j \mathbf{K}_j \mathbf{H}, \tau_j^2 \mathbf{I}_{n_j}), \quad j = 1, \cdots, J, \tag{14}$$

where $\mathbf{K}_j$ is an $n_j \times n$ indicator matrix that maps the n locations to the n_j locations where the variable j was observed. Also $\mathbf{I}_{n_j}$ is the $n_j \times n_j$ identity matrix and $\mathbf{1}_{n_j}$ is a $n_j \times 1$ vector of ones. Consequently, $\mathbf{T}_j$'s are conditionally independent given the latent vector $\mathbf{H}$.

A classical spatial factor model is constructed based on considering a Gaussian process for $H(\mathbf{s})$. However, in order to enhance the model fit for skewed spatial data, we propose to use a GLGC process for the spatial latent process. In this setting, we assume that

$$\mathbf{H} \sim GLGC_n(\mathbf{X}\boldsymbol{\beta}, \sigma_1^2 \mathbf{C}_{\theta_1}, \sigma_2^2 \mathbf{C}_{\theta_2}, \alpha), \tag{15}$$

where $\mathbf{X}$ contains k location-specific observable covariates, $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients and $\mathbf{C}_{\theta_l}$, $l = 1, 2$, is an $n \times n$ correlation matrix with $C_{\theta_l}(||\mathbf{s}_i - \mathbf{s}_j||)$ as its (i, j)th element. The latent vector $\mathbf{H}$ can be displayed under the following hierarchical representation

$$\begin{aligned} \mathbf{H}|\mathbf{U} &\sim N_n(\mathbf{X}\boldsymbol{\beta} + \alpha\mathbf{W}, \sigma_2^2 \mathbf{C}_{\theta_2}), \\ \mathbf{U} &\sim N_n(\mathbf{0}, \sigma_1^2 \mathbf{C}_{\theta_1}), \end{aligned} \tag{16}$$

where here $\mathbf{W} = \exp(\mathbf{U})$. As the resulting spatial factor model has been built on the GLGC process, we call it the GLGC factor (GLGCF) model.

3.1 Identification constraints

Before providing details of the Bayesian analysis, in what follows, we have placed some constraints that will be taken into account to ensure that the model parameters are identifiable.

1. To avoid the lack of identifiability with the intercept parameters $\nu_1, \cdots, \nu_J$, it is necessary to place a restriction on the cut-points and the matrix $\mathbf{X}$. Without loss of generality, we assume that the matrix $\mathbf{X}$ does not include the column of ones and $\kappa_{j,1} = 0$ for $j = 1, \cdots, J_o$. Therefore, we are left with the estimation of the cut-points vectors $\boldsymbol{\kappa}_1, \cdots, \boldsymbol{\kappa}_{J_o}$ where $\boldsymbol{\kappa}_j = (\kappa_{j,2}, \cdots, \kappa_{j,L-1})$.

2. The information contained in the ordinal responses pertains to probability of being in a particular category and thus we can not estimate the nugget effect parameters $\tau_1^2, \cdots, \tau_{J_o}^2$ together with the cut-points. Accordingly, we set $\tau_1^2 = \cdots \tau_{J_o}^2 = 1$.

3. One value of J factor loadings must be fixed. Here, we set $\omega_J = 1$.

Under these assumptions, the parameters of the GLGCF model are $\boldsymbol{\kappa} = (\boldsymbol{\kappa}_1, \cdots, \boldsymbol{\kappa}_{J_o})$, $\boldsymbol{\nu} = (\nu_1, \cdots, \nu_J)$, $\boldsymbol{\omega} = (\omega_1, \cdots, \omega_{J-1})$, $\boldsymbol{\tau} = (\tau_{J_o+1}, \cdots, \tau_J)$, $\boldsymbol{\beta}$, α, σ_1, σ_2, θ_1 and θ_2 all of which are now identifiable. The unaugmented likelihood function of the observed data is given by

$$p(\mathbf{y}_1, \cdots, \mathbf{y}_J|\boldsymbol{\zeta}) = \int_{\Re^n} \prod_{j=1}^{J} p(\mathbf{y}_j|\mathbf{h}, \boldsymbol{\zeta}) f_{GLGC}^n(\mathbf{h}|\mathbf{X}\boldsymbol{\beta}, \sigma_1^2\mathbf{C}_{\theta_1}, \sigma_2^2\mathbf{C}_{\theta_2}, \alpha) d\mathbf{h}, \tag{17}$$

where $\boldsymbol{\zeta} = (\boldsymbol{\beta},\ \boldsymbol{\nu},\ \boldsymbol{\omega},\ \boldsymbol{\tau}, \sigma_1,\ \sigma_2,\ \alpha,\ \theta_1,\ \theta_2, \boldsymbol{\kappa})$ is the vector of parameters. Moreover, if we define $A(y_{j_i}) = (\kappa_{j,l-1}, \kappa_{j,l}]$ when $y_{j_i} = l$ for $i = 1, \cdots, n_j$, $j = 1, \cdots, J_o$ and $l = 1, \cdots, L$ then

$$\begin{aligned} p(\mathbf{y}_j|\mathbf{h}, \boldsymbol{\zeta}) &= \int_{A(y_{j_1})} \cdots \int_{A(y_{j_{n_j}})} f_N^{n_j}(\mathbf{t}_j|\nu_j\mathbf{1}_{n_j} + \omega_j\mathbf{K}_j\mathbf{h}, \mathbf{I}_{n_j}) d\mathbf{t}_j, \quad j = 1, \cdots, J_o, \\ p(\mathbf{y}_j|\mathbf{h}, \boldsymbol{\zeta}) &= f_N^{n_j}(\mathbf{y}_j|\nu_j\mathbf{1}_{n_j} + \omega_j\mathbf{K}_j\mathbf{h}, \tau_j^2\mathbf{I}_{n_j}), \qquad j = J_o + 1, \cdots, J. \end{aligned}$$

4 Bayesian analysis

In this section, we will discuss the Bayesian inference for the GLGCF model. To avoid the calculation of high dimensional integrals in the likelihood function (17), we employ data augmentation to incorporate the latent variables $\mathbf{t}_o = (\mathbf{t}_1, \cdots, \mathbf{t}_{J_o})$, $\mathbf{h}$ and $\mathbf{u}$. The distribution of the unknown parameters and latent variables, given the observed data, can be written as

$$\begin{aligned} p(\mathbf{t}_o, \mathbf{h}, \mathbf{u}, \boldsymbol{\zeta}|\mathbf{y}_1, \cdots, \mathbf{y}_J) &\propto \prod_{j=1}^{J_o} p(\mathbf{y}_j|\mathbf{t}_j, \boldsymbol{\zeta}) p(\mathbf{t}_j|\mathbf{h}, \boldsymbol{\zeta}) \prod_{j=J_o+1}^{J} p(\mathbf{y}_j|\mathbf{h}, \boldsymbol{\zeta}) p(\mathbf{h}|\mathbf{u}, \boldsymbol{\zeta}) p(\mathbf{u}|\boldsymbol{\zeta}) \pi(\boldsymbol{\zeta}) \\ &= \prod_{j=1}^{J_o} \prod_{i=1}^{n_j} I_{\{t_{j_i} \in A(y_{j_i})\}} f_N(t_{j_i}|\nu_j + \omega_j h_{j_i}, 1) \\ &\times \prod_{j=J_o+1}^{J} f_N^{n_j}(\mathbf{y}_j|\nu_j\mathbf{1}_{n_j} + \omega_j\mathbf{K}_j\mathbf{h}, \tau_j^2\mathbf{I}_{n_j}) \\ &\times f_N^n(\mathbf{h}|\mathbf{X}\boldsymbol{\beta} + \alpha\exp(\mathbf{u}), \sigma_2^2\mathbf{C}_{\theta_2}) f_N^n(\mathbf{u}|\mathbf{0}, \sigma_1^2\mathbf{C}_{\theta_1}) \pi(\boldsymbol{\zeta}). \end{aligned}$$

We now define the prior distribution of the vector of the model parameters, i.e. $\pi(\boldsymbol{\zeta})$ to complete the Bayesian model specification. We first assume the components of the parameter vector $\boldsymbol{\zeta}$ to be independent a priori. Then, we assign proper (but diffuse) prior distributions for the unknown parameters. Additionally, a family of conjugate priors is assigned for most of the parameters. To be more specific, the conjugate multivariate normal and inverse gamma prior distributions will be taken for components of $\boldsymbol{\zeta}_1 = (\boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\omega}, \alpha)$ and $\boldsymbol{\zeta}_2 = (\sigma_1^2, \sigma_2^2, \boldsymbol{\tau}^2)$, respectively. More specifically,

$$\begin{aligned} \pi(\boldsymbol{\zeta}_1) &= f_N^k(\boldsymbol{\beta}|\mathbf{0}, c_\beta\mathbf{I}_k) f_N^J(\boldsymbol{\nu}|\mathbf{0}, c_\nu\mathbf{I}_J) f_N^{J-1}(\boldsymbol{\omega}|\mathbf{0}, c_\omega\mathbf{I}_{J-1}) f_N(\alpha|0, c_\alpha), \\ \pi(\boldsymbol{\zeta}_2) &= f_{IG}(\sigma_1^2|s_{\sigma_1}, r_{\sigma_1}) f_{IG}(\sigma_2^2|s_{\sigma_2}, r_{\sigma_2}) \prod_{j=J_o+1}^{J} f_{IG}(\tau_j^2|s_{\tau_j}, r_{\tau_j}), \end{aligned} \tag{18}$$

where $f_{IG}(\cdot|s,r)$ denotes the probability density function of the inverse gamma distribution with the shape parameter s and the rate parameter r. The hyperparameters c_β, c_ν, c_ω and c_α are chosen to be large in order to ensure that the corresponding prior distributions are vague. Additionally, by choosing very small values for the hyperparameters s_{σ_1}, r_{σ_1}, s_{σ_2}, r_{σ_2}, s_τ and r_τ, we use vague prior distributions for the scale parameters involved in the vector $\boldsymbol{\zeta}_2$.

For θ_1 and θ_2, we assign independent gamma prior distributions

$$\pi(\theta_1,\theta_2) = f_G(\theta_1|s_{\theta_1},r_{\theta_1})f_G(\theta_2|s_{\theta_2},r_{\theta_2}), \tag{19}$$

where $f_G(\cdot|s,r)$ denotes the probability density function of gamma distribution with the shape parameter s and the rate parameter r. We consider $s_{\theta_1} = s_{\theta_2} = s_\theta$ and $r_{\theta_1} = r_{\theta_2} = r_\theta$. In the Supplementary Appendix A (Zareifard et al., 2017), we discuss how to choose appropriate values for the hyperparameters of these priors.

For $j = 1,\cdots,J_o$, the $L-2$ unknown cut-points $\kappa_{j,l}; l = 2,\cdots,L-1$ are ordered values, so the uniform prior distribution can be considered for the vector $\boldsymbol{\kappa}_j$ such that

$$\pi(\kappa_{j,l}|\kappa_{j,l-1},\kappa_{j,l+1}) \propto I_{(\kappa_{j,l-1}<\kappa_{j,l}<\kappa_{j,l+1})}, \quad l = 2,\cdots,L-1, \quad j = 1,\cdots,J_o.$$

However, this prior can lead to a poor mixing in the Markov chain. To overcome this obstacle, following Albert and Chib (1997), an unconstrained and real-valued vector $\boldsymbol{\delta}_j = (\delta_{j,2},\cdots,\delta_{j,L-1})$ is used in which $\delta_{j,2} = \log(\kappa_{j,2})$ and $\delta_{j,l} = \log(\kappa_{j,l} - \kappa_{j,l-1})$ for $l = 3,\cdots,L-1$. The inverse transformation is then given by $\kappa_{j,l} = \sum_{i=2}^{l}\exp(\delta_{j,i})$. We then impose an unrestricted multivariate normal prior distribution with mean $\mathbf{0}$ and covariance matrix $c_\delta I_{L-2}$ for the $(L-2)$-dimensional vector $\boldsymbol{\delta}_j$, $j = 1,\cdots,J_o$.

4.1 Markov chain Monte Carlo-based inference

We shall now turn to a discussion of how to generate samples from the joint distribution of model parameters and latent variables via the Gibbs sampling algorithm. The full conditionals of the latent variables are as follows:

$$\begin{aligned}
p(\mathbf{t}_o|\mathbf{rest}) &\propto \prod_{j=1}^{J_o}\prod_{i=1}^{n_j} I_{\{t_{j_i}\in A(y_{j_i})\}} f_N(t_{j_i}|\nu_j+\omega_j h_{j_i},1),\\
p(\mathbf{h}|\mathbf{rest}) &= f_N^n(\mathbf{h}|\boldsymbol{\Sigma}_h(\frac{1}{\sigma_2^2}\mathbf{Q}_{\theta_2}(\mathbf{X}\boldsymbol{\beta}+\alpha\exp(\mathbf{u}))+\sum_{j=1}^{J}\frac{\omega_j}{\tau_j^2}\mathbf{K}_j'(\mathbf{T}_j-\nu_j\mathbf{1}_{n_j})),\boldsymbol{\Sigma}_h),\\
p(\mathbf{u}|\mathbf{rest}) &\propto \exp\{-\frac{\alpha^2}{2\sigma_2^2}(\exp(\mathbf{u})-\boldsymbol{\mu}_u)'\mathbf{Q}_{\theta_2}(\exp(\mathbf{u})-\boldsymbol{\mu}_u)\}f_N^n(\mathbf{u}|\mathbf{0},\sigma_1^2\mathbf{C}_{\theta_1}),
\end{aligned}$$

where $\mathbf{t} = (\mathbf{t}_1',\cdots,\mathbf{t}_J')'$, $\mathbf{Q}_{\theta_j} = \mathbf{C}_{\theta_j}^{-1}$, $j = 1,2$, $\boldsymbol{\mu}_u = \frac{1}{\alpha}(\mathbf{h}-\mathbf{X}\boldsymbol{\beta})$ and $\boldsymbol{\Sigma}_h = (\frac{1}{\sigma_2^2}\mathbf{Q}_{\theta_2} + \sum_{j=1}^{J}\frac{\omega_j^2}{\tau_j^2}\mathbf{K}_j'\mathbf{K}_j)^{-1}$. Now the algorithm will proceed in the following steps:

1. For $j = 1,\cdots,J_o$ and $i = 1,\cdots,n_j$, update the latent variable t_{j_i} from its full conditional distribution which is in fact a truncated normal distribution.
2. Update the latent vector $\mathbf{H}$ from its full conditional distribution.

3. Update the parameters $(\boldsymbol{\nu}, \boldsymbol{\omega})$ in block form from their joint full conditional distribution.

4. For $j = J_o + 1, \cdots, J$, update the nugget effect parameter τ_j^2 from its full conditional distribution.

5. Update the cut-points $\boldsymbol{\kappa}$ by first drawing $\boldsymbol{\delta}$ from its full conditional distribution via the random-walk Metropolis-Hastings algorithm and then inverse mapping to get $\boldsymbol{\kappa}$. See Higgs and Hoeting (2010) for explicit details on the reparameterization and updating scheme for $\boldsymbol{\kappa}$.

6. Update the parameters $(\boldsymbol{\beta}, \alpha, \sigma_2^2, \theta_2)$ in block form from their joint full conditional distribution.

7. Update the latent vector $\mathbf{U}$ from its full conditional distribution. The necessary details in this step are explained in the remainder of this section.

8. Update the parameters (σ_1^2, θ_1) in block form from their joint full conditional distribution.

As observed, because the elements of $\mathbf{U}$ are not conditionally independent given other variables, a direct simulation from the full conditional of $\mathbf{U}$, that is

$$p(\mathbf{u}|\mathbf{rest}) \propto p(\mathbf{h}|\mathbf{u}, \mathbf{rest}) f_N^n(\mathbf{u}|\mathbf{0}, \sigma_1^2 \mathbf{C}_{\theta_1}),$$

is not a trivial task. To deal with this issue, the slice sampling method has recently been found to provide an attractive strategy and therefore has received the utmost attention by those who use MCMC procedures to simulate from complex non-normalized multivariate densities (Neal, 2003; Agarwal and Gelfand, 2005). In this framework, we sample from $p(\mathbf{u}|\mathbf{rest})$ by sampling uniformly from the $(n+1)$-dimensional region that lies under the plot of $p(\mathbf{h}|\mathbf{u}, \mathbf{rest}) f_N^n(\mathbf{u}|\mathbf{0}, \sigma_1^2 \mathbf{C}_{\theta_1})$. This idea can be formalized by sampling from an augmented distribution $p(\mathbf{u}, \mathrm{v}|\mathbf{rest})$, where V is an auxiliary variable. Then, it is possible to obtain the marginal samples $\mathbf{u}^{(t)}$ by drawing $(\mathbf{u}^{(t)}, \mathrm{v}^{(t)})$ from $p(\mathbf{u}, \mathrm{v}|\mathbf{rest})$ and, subsequently, ignoring the samples $\mathrm{v}^{(t)}$. More precisely, we introduce an auxiliary variable V such that $[V|\mathbf{u}, \sigma_2, \boldsymbol{\mu}_u, \theta_2]$ has the uniform distribution on the interval $[0, \exp\{-\frac{\alpha^2}{2\sigma_2^2}(\exp(\mathbf{u}) - \boldsymbol{\mu}_u)' \mathbf{Q}_{\theta_2} (\exp(\mathbf{u}) - \boldsymbol{\mu}_u)\}]$. Now, if e represents the exponential distribution with mean 1, say exp(1), then $\log V = -\frac{\alpha^2}{2\sigma_2^2}(\exp(\mathbf{u}) - \boldsymbol{\mu}_u)' \mathbf{Q}_{\theta_2} (\exp(\mathbf{u}) - \boldsymbol{\mu}_u) - e$ and so it will be routine to sample. Accordingly,

$$p(\mathbf{u}|\mathrm{v}, \mathbf{rest}) \quad \propto \quad I_{\{\mathrm{v} < \exp\{-\frac{\alpha^2}{2\sigma_2^2}(\exp(\mathbf{u}) - \boldsymbol{\mu}_u)' \mathbf{Q}_{\theta_2} (\exp(\mathbf{u}) - \boldsymbol{\mu}_u)\}\}} f_N^n(\mathbf{u}|\mathbf{0}, \sigma_1^2 \mathbf{C}_{\theta_1}).$$

In order to sample from the above distribution, we would draw $\mathbf{u}$ from $f_N^n(\mathbf{u}|\mathbf{0}, \sigma_1^2 \mathbf{C}_{\theta_1})$ and retain it only if $\mathbf{u}$ is such that $\mathrm{v} < \exp\{-\frac{\alpha^2}{2\sigma_2^2}(\exp(\mathbf{u}) - \boldsymbol{\mu}_u)' \mathbf{Q}_{\theta_2} (\exp(\mathbf{u}) - \boldsymbol{\mu}_u)\}$.

Finally, given $\sigma_2^{(t)}, \theta_2^{(t)}, \alpha^{(t)}$ and $\boldsymbol{\mu}_u^{(t)}$, we summarize the main steps in the iteration $(t+1)$ of the slice sampling algorithm as follows:

1. Draw $e^{(t+1)}$ of exp(1), and let $a_t = \boldsymbol{\epsilon}_u^{(t)'} \mathbf{Q}_{\theta_2^{(t)}} \boldsymbol{\epsilon}_u^{(t)} + 2\frac{\sigma_2^{2(t)}}{\alpha^{2(t)}} e^{(t+1)}$ where $\boldsymbol{\epsilon}_u = \exp(\mathbf{u}) - \boldsymbol{\mu}_u^{(t)}$,
2. Draw $\mathbf{u}^{(t+1)}$ from a truncated normal distribution on $\{\boldsymbol{\epsilon}_u' \mathbf{Q}_{\theta_2^{(t)}} \boldsymbol{\epsilon}_u < a_t\}$.

For details see Supplementary Appendix B (Zareifard et al., 2017).

4.2 Bayesian spatial prediction

Prediction of the responses at unobserved locations is often a major objective of the study. In the spatial factor models, obtaining a prediction map of the underlying spatial process $H(\cdot)$ which explains all the systematic variability of responses, is also usually of great importance. This map can be used to relate the observed variables to the unobservable actual process and hence to better understand the multivariate phenomenon under study. In what follows, we only concentrate on the prediction of the latent factor. Accordingly, prediction of the responses is, of course, straightforward.

We should predict the latent variable at each observed location or at some new locations. Before addressing this, we first denote $\mathbf{H}_0 = (H_{0_1}, \cdots, H_{0_p})$ as the latent vector at arbitrary locations $\mathbf{s}_{0_1}, \cdots, \mathbf{s}_{0_p}$. Also similarly with $\mathbf{H}_0$, we define $\mathbf{U}_0 = (U_{0_1}, \cdots, U_{0_p})$. Now to make Bayesian spatial prediction, we need to obtain the posterior predictive distribution

$$p(\mathbf{h_0}|\mathbf{y}_1, \cdots, \mathbf{y}_J) = \int p(\mathbf{h}_0|\mathbf{h}, \mathbf{u}, \mathbf{u}_0, \boldsymbol{\eta}) p(\mathbf{u}_0|\mathbf{u}, \boldsymbol{\eta}) p(\boldsymbol{\eta}, \mathbf{u}, \mathbf{h}|\mathbf{y}_1, \cdots, \mathbf{y}_J), \tag{20}$$

where

$$\begin{aligned} p(\mathbf{u}_0|\mathbf{u}, \boldsymbol{\eta}) &= N_p(\mathbf{C}_{\theta_1}^{po}\mathbf{C}_{\theta_1}^{-1}\mathbf{u}, \sigma_1^2(\mathbf{C}_{\theta_1}^{pp} - \mathbf{C}_{\theta_1}^{po}\mathbf{C}_{\theta_1}^{-1}\mathbf{C}_{\theta_1}^{op})), \\ p(\mathbf{h}_0|\mathbf{h}, \mathbf{u}, \mathbf{u}_0, \boldsymbol{\eta}) &= N_p(\mathbf{X}_p\boldsymbol{\beta} + \alpha\exp(\mathbf{u}_0) + \mathbf{C}_{\theta_2}^{po}\mathbf{C}_{\theta_2}^{-1}(\mathbf{h} - \mathbf{X}\boldsymbol{\beta} - \alpha\exp(\mathbf{u})), \\ &\quad \sigma_2^2(\mathbf{C}_{\theta_2}^{pp} - \mathbf{C}_{\theta_2}^{po}\mathbf{C}_{\theta_2}^{-1}\mathbf{C}_{\theta_2}^{op})), \end{aligned} \tag{21}$$

where X_p is $p \times k$ matrix of k location-specific observable covariates and

$$\mathbf{C}_{\theta_i}^{pp} = [C_{\theta_i}(||\mathbf{s}_{0_i} - \mathbf{s}_{0_j}||)]_{p\times p}, \quad \mathbf{C}_{\theta_i}^{po} = [C_{\theta_i}(||\mathbf{s}_{0_i} - \mathbf{s}_j||)]_{p\times n}, \quad i = 1, 2.$$

The posterior predictive distribution (20) can not be evaluated in a closed form but sampling from it is straightforward. Then the Bayesian spatial predictor is determined by

$$\hat{E}(\mathbf{H}_0|\mathbf{y}_1, \cdots, \mathbf{y}_J) = \frac{1}{G}\sum_{g=1}^{G} \mathbf{h}_0^{(g)},$$

where $\{\mathbf{h}_0^{(g)}; g = 1, \cdots, G\}$ denotes a sample from (20), and G is its sample size.

5 Simulation study

In this section, we perform a simulation study to evaluate the performance of the GLGCF model. We also compare the results with those obtained from a Gaussian

Parameter	True Value	$\alpha=1,\ \sigma_1^2=1$	$\alpha=2,\ \sigma_1^2=1$	$\alpha=1,\ \sigma_1^2=2$	$\alpha=2,\ \sigma_1^2=2$
α		$0.51_{(-0.53,1.65)}$	$2.04_{(0.95,2.94)}$	$0.59_{(-0.12,2.37)}$	$1.64_{(0.65,3.37)}$
σ_1^2		$1.24_{(0.39,2.34)}$	$0.85_{(0.21,2.52)}$	$1.82_{(0.67,4.51)}$	$2.28_{(0.92,4.02)}$
θ_1	3	$2.61_{(1.53,3.89)}$	$2.73_{(2.18,3.61)}$	$3.19_{(2.31,4.36)}$	$3.58_{(2.05,4.79)}$
σ_2^2	2	$1.88_{(0.21,3.62)}$	$2.50_{(0.63,4.52)}$	$2.65_{(0.73,5.21)}$	$1.87_{(0.73,6.31)}$
θ_2	3	$2.21_{(1.65,3.38)}$	$3.38_{(2.57,4.51)}$	$3.60_{(2.07,5.12)}$	$2.25_{(1.03,5.18)}$
ν_1	0	$0.15_{(-0.87,1.08)}$	$-0.23_{(-1.27,0.97)}$	$-0.56_{(-1.38,0.41)}$	$1.03_{(-0.14,1.95)}$
ν_2	0	$0.16_{(-0.61,0.77)}$	$-0.32_{(-1.53,0.67)}$	$0.81_{(0.07,1.19)}$	$0.70_{(-0.31,1.43)}$
ω	1	$1.18_{(0.83,1.55)}$	$0.85_{(0.40,1.18)}$	$1.38_{(0.36,2.53)}$	$0.83_{(0.36,1.60)}$
κ_2	2	$2.68_{(1.31,4.17)}$	$2.26_{(0.82,4.03)}$	$2.59_{(0.95,4.71)}$	$3.07_{(1.63,4.81)}$
τ^2	1	$0.82_{(0.21,1.76)}$	$1.40_{(0.17,2.91)}$	$1.16_{(0.15,5.13)}$	$1.74_{(0.27,7.51)}$

Table 3: Posterior mean (95% credible interval) for the parameters of the GLGCF model.

factor model (denoted by GF) as well as a factor model in which we use a log-Gaussian process for $H(\mathbf{s})$ (denoted by LGF). The GF and LGF models can be seen as limiting cases of the GLGCF model. For each model, we ran two parallel chains starting from distant initial points with a total of 100,000 iterations for each chain. Convergence of the MCMC was verified through the Gelman and Rubin convergence diagnostics (Gelman and Rubin, 1992) implemented in the CODA package of R language. We have used the last 25,000 iterations of each chain for the purposes of inference.

Example 1. To address the identifiability of the parameters, we generate the data from the GLGCF model without any covariates. Spatial sampling points are randomly taken on a $[0,10]\times[0,10]$ square. We focus on two parameters which have effects on the skewness of the model (i.e. α and σ_1^2), because we would expect inference to be most challenging for these parameters. We then consider a data set consisting of an ordinal variable Y_1 with three categories and a continuous variable Y_2. Next, we shall assume that the locations of the two variables are misaligned. Throughout, we use a sample size of $n_1=n_2=100$ with $\nu_1=\nu_2=0$, $\omega=1$, $\tau^2=1$, $\kappa_2=2$, $\sigma_2^2=2$, $\theta_1=\theta_2=3$. To assess the identifiability of skewness parameters, four data sets were generated assuming $\alpha,\sigma_1^2\in\{1,2\}$. The vectors $\mathbf{T}_1$ and $\mathbf{T}_2$ were simulated from the GLGCF model and finally, we consider Y_{1_i} as

$$Y_{1_i}=\begin{cases}1, & \text{if } T_{1_i}\le 0;\\ 2, & \text{if } 0<T_{1_i}\le 2;\\ 3, & \text{otherwise.}\end{cases}\quad i=1,\cdots,100$$

The Bayesian analysis is implemented under the proper priors (18) with $c_\beta=c_\alpha=c_{\nu_1}=c_{\nu_2}=c_\omega=10^4$ and $s_{\sigma_1}=r_{\sigma_1}=s_{\sigma_2}=r_{\sigma_2}=s_\tau=r_\tau=10^{-6}$. We also choose a diffuse prior for δ with considering $c_\delta=100$. Table 3 displays some posterior results for the parameters under the GLGCF model. This table clearly indicates that the data allow for meaningful inference on the skewness parameters, even with this quite moderate sample size. As a result, the GLGCF model provides satisfactory and reliable results.

Example 2. We want to investigate whether the choosing of an appropriate factor model is important in predicting the spatial factor $H(\mathbf{s})$. To address this issue, we conduct a simulation study to compare the predictive performance of the GLGCF model

True model⇒	GF	LGF	GLGCF($\alpha = \sigma_1^2 = 1$)	GLGCF($\alpha = \sigma_1^2 = 2$)
⇓ Fitted model				
GLGCF	0.114	0.164	0.098	0.121
GF	0.076	3.416	2.153	7.937
LGF	1.057	0.086	1.160	3.084

Table 4: The average of MSPE over 20 simulated data sets under the considered models.

against the GF and LGF models. To be more specific, we consider the situation where the factor process $H(\mathbf{s})$ in (13) is either Gaussian, log-Gaussian or GLGC and investigate the effect of model misspecification on the prediction. Data sets consist of an ordinal variable with three categories as well as a continuous variable. Altogether, 80 data sets (i.e., 20 data sets from the GF model, 20 data sets from the LGF model under $\alpha = \sigma_1^2 = 1$, 20 data sets from the GLGCF model under $\alpha = \sigma_1^2 = 1$ as well as 20 data sets from the GLGCF model under $\alpha = \sigma_1^2 = 2$) were simulated by choosing the values of other parameters as given in Example 1. We generated the data on sampling locations assigned in the applied example (notice that in the real example, $n_1 = 45$ and $n_2 = 163$ introducing misalignment, see Figure 8). In addition, a regular 15×15 grid located in the region is left out as a hold-out sample to assess the predictive performance of the fitted models.

After fitting three models to simulated data sets, we predicted the values of latent process $H(\mathbf{s})$ in locations of the considered grid. Using these predictions and the true values, the mean squared prediction error (i.e., $MSPE = \frac{\sum_{i=1}^{225}(h_i - \hat{h}_i)^2}{225}$ where h_i is the i^{th} latent factor located in the grid and $\hat{h}_i$ is the posterior mean) were computed. Table 4 provides an average of MSPE over 20 simulated data sets. It is evident from this table that when the true models are GF or LGF, there is rather a small difference between the GLGCF model and them. However, when the distribution of the spatial factor is inadequate in fitting the data from the GLGCF model, both models (GF and LGF) yield rather big MSPE values particularly for the large value of skewness parameters. Still, as expected, the LGF model fits slightly better than GF. We can therefore conclude that a misspecification of the latent variable model can lead to poor prediction.

Finally, for some simulated data sets from the GLGCF model, we also assessed the accuracy of the considered models in predicting the bivariate response at unobservable locations. Consistent with the prediction results of the latent process, our experiments (not presented here) demonstrated the success of the GLGCF model for prediction of the continuous variable. Of course, the predictive performance of GLGCF and LGF models for the ordinal variable was approximately similar.

6 Illustrative example

We now turn from the simulation studies to a real example to demonstrate the applicability of the GLGCF model. The data used in this study consist of esophageal cancer

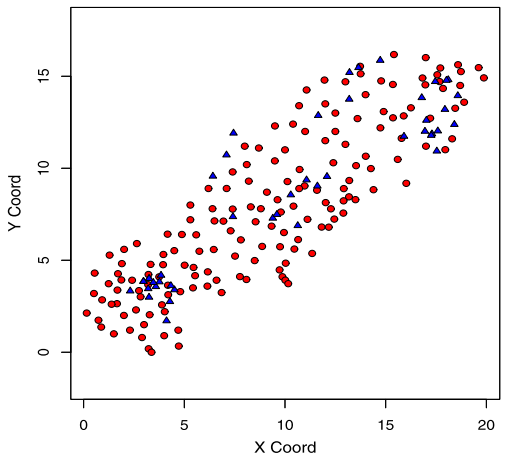

Figure 8: Sampling locations of $n_1 = 45$ sampled villages (blue triangles) and $n_2 = 163$ Cesium data (red circles).

mortality rates (per 100,000) for 45 sampled villages recorded between March 2006 and March 2007 provided by the cancer registration department in Golestan province. According to recent cancer registry data in Iran, esophageal cancer has great prevalence in this province which spans 20437 km^2 and receives very high annual rainfall (Roshandel et al., 2012). The rate has reported as an ordinal variable scaled with low, medium and high. Figure 8 shows longitude and latitude coordinates of the centroid of the sampled villages. Another part of the data includes the amount of Cs^{137} which is a continuous response and it was measured by a university research group at 163 points during 2006 (see Figure 8). Due to the fact that northern Iran has been affected by the Chernobyl disaster, through precipitation, Cesium-137 (Cs^{137}), which is a heavy metal, has found its way into this area and is now regarded as one of the most rampantly dangerous elements that contribute to soil pollution in this region. So, we consider the esophageal cancer rates and the Cs^{137} values as two responses that might explain soil pollution. In fact, soil pollution is considered as the latent process of interest and two data sources together will help infer on the spatial process of soil pollution. We aim to understand how the soil pollution level affects the esophageal cancer mortality level and to evaluate the spatial distribution of pollution in the soil.

Since the latent factor of the soil pollution which is of our interest, is inherently a point-referenced outcome, we treat areas (villages) as point-referenced data. There are several possible ways to jointly model point-location and areal bivariate data (see for instance Berrocal et al., 2010). In a simple way, one may assign the same health value to all points in a given areal unit and treat as 163 bivariate response vectors. However, our strategy seems to be more reasonable here since the area of each village relative to the total area of the province is negligible. Particularly in all cases, the sampling locations of Cs^{137} data are far away from the villages. Indeed, two variables are observed at different points. Moreover, in the application considered here, we actually have two different types of outcome instead of bivariate responses that are closely tied to one another.

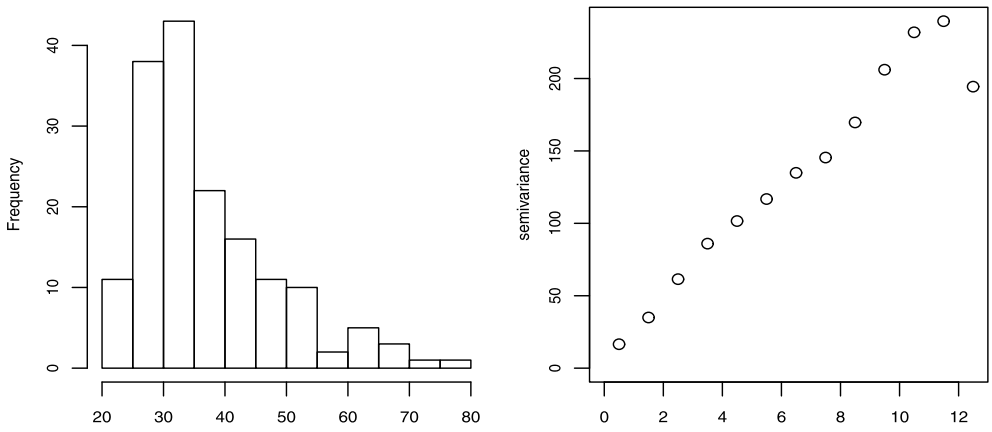

Figure 9: Histogram (left column) and robust empirical semivariograms (right column) for the continuous response.

Since there are no explanatory variables that might be used for the modeling purposes, we assume the mean to be constant. For an exploratory data analysis, the histogram and the robust empirical semivariograms of continuous data are plotted in Figure 9. The histogram suggests that the data have a right-skewed distribution. Furthermore, it is clear from the robust empirical semivariograms that there exists a strong spatial correlation in the data. To deal with departures from normality, a common method in spatial statistics is to use a single transformation (e.g., square-root or log) for the continuous response variable, and then assume a Gaussian latent process. However, in addition to the problem of choosing appropriate transformation, the transformed variables are difficult to interpret. Moreover, it is hard to offer a joint transformation of the entire process into a Gaussian process. Consequently we focus on a more suitable strategy based on assuming that the data arise from a spatial latent process being skewed. Thus we employ the GLGCF model to fit to the data. We also fitted three models GF, LGF and a skew factor model with spatially independent components (non-spatial GLGCF). In order to compare four models, we computed the Bayes factors between models. The Bayes factor in favor of the GLGCF versus the GF, LGF and non-spatial GLGCF models was evaluated as 7.4×10^4, 8.2×10^3 and 9.6×10^4 respectively, which indicates an overwhelming support for the GLGCF model. In particular, the non-spatial model performs poorly. The estimation results of model fitting are displayed in Table 5. The estimate of the factor loading is positive, indicating that the linear relationship between soil pollution condition and the esophageal cancer mortality rate is positive. Hence soil pollution influences human health which might be explained by the fact that the different elements in the soil can easily be absorbed by plants. In addition, the water quality for daily use may be influenced by present elements. The estimates of skewness parameters also indicate that two less flexible models, i.e. GF and LGF, fall short of a sound when fitting to the data.

Figure 10 also displays the prediction map of the spatial latent process in the study region. This figure is important for a better understanding of spatial distribution of

Parameters	ν_1	ν_2	ω	τ^2	κ	α	σ_1^2	σ_2^2	θ_1	θ_2
1st Qu	-0.72	29.84	0.58	6.58	3.19	0.91	1.81	5.27	2.46	1.81
Median	-0.13	30.91	0.62	7.91	3.79	1.79	2.15	6.74	2.94	2.20
Mean	-0.15	30.83	0.65	7.87	3.86	1.88	2.26	7.93	2.86	2.15
3rd Qu	0.48	31.68	0.71	8.61	4.47	2.37	2.58	9.62	3.41	2.32
S.D	0.93	1.17	0.21	1.27	0.98	1.83	0.67	3.70	0.55	0.37

Table 5: Posterior results under fitting the GLGCF model.

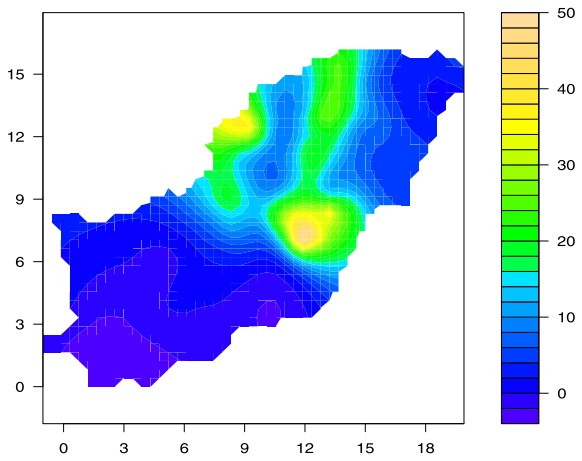

Figure 10: The prediction map of actual soil pollution.

actual soil pollution in the province. From the map, we see that $\hat{H}(\mathbf{s})$ is a spatially continuous process. Further, the subregions with high concentration are clearly specified. For example, the condition in the eastern region of the province may be of concern.

7 Discussion and further directions

In most spatial latent variable models, it is assumed that the underlying latent process is Gaussian, which may not be a valid assumption. Although the skew-normal distributions family has some desirable properties, the skew-Gaussian spatial process demonstrates certain limitations since the family is capable of capturing only mild amounts of skewness. Having this in mind along with some other difficulties concerning the utilization of the skew-normal family in a spatial context, this paper has aimed to tackle skewness by introducing a Gaussian-log Gaussian convolution process. Our model in which the log-normal distribution is used, can evidently capture a greater amount of skewness. Next, a spatial factor model relying on the skew process was introduced for the analysis of spatial data with mixed ordinal and continuous outcomes which share a spatial latent

process. The Gaussian factor model is a special case of the general construction. We presented a Bayesian approach for the inference. Then, to obtain simulations from the posterior distribution, a Gibbs framework was developed. Through simulation studies we demonstrated that the GLGCF model is favored by the generated data and leads to a more precise prediction. This is a direct consequence of the fact that this model has great flexibility in analyzing skewed spatial data. Beside this, the applied example concerning soil pollution illustrates our methodology and its utility in practice. According to our findings, the Bayes factor noticeably favors the GLGCF model.

Sampling from the full conditional distributions of some latent variables involves taking consecutive inverses of $n \times n$ covariance matrices. This problem will be more critical when n is large. Because of the conditional independence property in Gaussian Markov random fields (GMRFs), the mentioned problem is less critical for GMRFs since they involve sparse precision matrices. Hence, we can use numerical methods for them, which are much quicker than general dense matrix calculations (Rue and Held, 2005). Based on this idea, Lindgren et al. (2011) addressed the big n problem in Gaussian random fields (GRFs) by constructing an explicit link between GRFs with the Matérn covariance function and GMRFs. Using their approach, the GMRF representation can be constructed explicitly for the latent components $\mathbf{H}$ and $\mathbf{U}$ of the GLGCF model, and then it is possible to use numerical methods for sparse matrices. This formulation can greatly simplify the implementation of the proposed approach because it allows us to use any of the computational tools available for the latent Gaussian processes.

The proposed factor model is appealing for analyzing the mixed continuous and ordinal response variables, but a point is inherently raised for discussion when at least one of the continuous response variables is symmetric. In such a case, we can employ the idea proposed here in a different way. Specifically, the problem suggests that the responses may measure something different from the considered Gaussian latent variables, therefore another latent variable would be necessary to regulate skewness. This leads to a spatial skew factor model for $\mathbf{T}(\mathbf{s}) = (T_1(\mathbf{s}), \cdots, T_J(\mathbf{s}))'$ as

$$\mathbf{T}(\mathbf{s}) = \boldsymbol{\mu}(\mathbf{s}) + \boldsymbol{\Lambda}\mathbf{w}(\mathbf{s}) + \boldsymbol{\rho}(\mathbf{s}), \tag{22}$$

where $\boldsymbol{\Lambda} = (\boldsymbol{\lambda}_1, \cdots, \boldsymbol{\lambda}_r, \boldsymbol{\alpha})$ is a $J \times (r+1)$ matrix with the last column $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_J)'$, $\mathbf{w}(\mathbf{s}) = (Z_1(\mathbf{s}), \cdots, Z_r(\mathbf{s}), \exp(U(\mathbf{s})))'$ and $\boldsymbol{\rho}(\mathbf{s})$ is an $J \times 1$ vector of measurement errors distributed as $N(\mathbf{0}, \boldsymbol{\Psi})$. The Gaussian processes $Z_1(\mathbf{s}), \cdots, Z_r(\mathbf{s}), U(\mathbf{s})$ are independent, each with a parametric covariance function as described in Section 2. The model given in (22) as an extension of the spatial Gaussian factor model (see e.g., Wang and Wall, 2003; Ren and Banerjee, 2013) seems flexible in effectively capturing skewness in spatial data. However, exploring this issue is beyond the current scope of this paper and will be pursued elsewhere.

Furthermore, in order to extend the proposed method here for a general multivariate setting, consider a case where we aim to model only two skew processes $H_1(\mathbf{s})$ and $H_2(\mathbf{s})$ (without loss of generality, the means of two processes are taken to be zero). Routinely, we may assume two correlated GLGC processes for them as

$$H_1(\mathbf{s}) \quad = \quad \alpha_1 \exp\left(U_1(\mathbf{s})\right) + Z_1(\mathbf{s}),$$

$$H_2(\mathbf{s}) \quad = \quad \alpha_2 \exp\left(U_2(\mathbf{s})\right) + Z_2(\mathbf{s}),$$

where $\mathbf{U}(\mathbf{s}) = (U_1(\mathbf{s}), U_2(\mathbf{s}))'$ and $\mathbf{Z}(\mathbf{s}) = (Z_1(\mathbf{s}), Z_2(\mathbf{s}))'$ are two independent bivariate Gaussian processes so that each follows a linear model of coregionalization (LMC). Specifically, $\mathbf{U}(\mathbf{s}) = \mathbf{A}_1\mathbf{W}_1(\mathbf{s})$ and $\mathbf{Z}(\mathbf{s}) = \mathbf{A}_2\mathbf{W}_2(\mathbf{s})$ where for $k = 1, 2$, $\mathbf{A}_k$ is the coregionalization matrix, and $\mathbf{W}_k(\mathbf{s}) = (W_{k_1}(\mathbf{s}), W_{k_2}(\mathbf{s}))'$ where $W_{k_1}(\mathbf{s})$ and $W_{k_2}(\mathbf{s})$ are independent Gaussian processes with zero means, unit variances, and correlation functions $\rho_{k_1}(h)$ and $\rho_{k_2}(h)$ respectively. As an alternative way to extend the univariate skew model, we may first consider an LMC model for $\mathbf{H}(\mathbf{s}) = (H_1(\mathbf{s}), H_2(\mathbf{s}))'$, i.e., $\mathbf{H}(\mathbf{s}) = \mathbf{A}\mathbf{W}(\mathbf{s})$ where $\mathbf{W}(\mathbf{s}) = (W_1(\mathbf{s}), W_2(\mathbf{s}))'$. We then assume that $W_1(\mathbf{s})$ and $W_2(\mathbf{s})$ are two independent GLGC processes. In such a situation the LMC covariance structure which is of the form $\mathbf{A}\mathbf{\Sigma}_{\mathbf{s},\mathbf{s}'}\mathbf{A}'$ where $\mathbf{\Sigma}_{\mathbf{s},\mathbf{s}'} = Cov(\mathbf{W}(\mathbf{s}), \mathbf{W}(\mathbf{s}'))$, is preserved. As we have pointed out, we can extend the univariate skew model to a multivariate case in at least two different ways. It would be interesting if the properties and the flexibility of these models were to be explored and compared in a further work.

Supplementary Material

Supplementary Material of "Modeling Skewed Spatial Data Using a Convolution of Gaussian and Log-Gaussian Processes" (DOI: 10.1214/17-BA1064SUPP; .pdf).

References

Agarwal, D. K., Gelfand, A. E. (2005). Slice sampling for simulation based fitting of spatial data models. *Statistics and Computing*, 15(1): 61–69. MR2137218. doi: https://doi.org/10.1007/s11222-005-4790-z. 533, 546

Aigner, D., Lovell, K., Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, 6: 21–37. MR0448782. doi: https://doi.org/10.1016/0304-4076(77)90052-5. 532

Albert, J., Chib, S. (1997). Bayesian methods for cumulative, sequential, and two-step ordinal data regression models. Technical report. Department of Mathematics and Statistics, Bowling Green State University. 545

Allard, D., Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics: Theory and Methods*, 36: 1821–1834. MR2409813. doi: https://doi.org/10.1080/03610920601126290. 532

Arellano-Valle, R. B., Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33: 561–574. MR2298065. doi: https://doi.org/10.1111/j.1467-9469.2006.00503.x. 532

Asmussen, S., Jensen, J. L., Rojas-Nandayapa, L. (2016). On the Laplace transform of the log-normal distribution. *Methodology and Computing in Applied Probability*, 18: 441–458. MR3488586. doi: https://doi.org/10.1007/s11009-014-9430-7. 537

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, 12: 171–178. MR0808153. 532

Azzalini, A., Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Association. Series B*, 61: 579–602. MR1707862. doi: https://doi.org/10.1111/1467-9868.00194. 532

Barouch, E. Kaufman, G. M., Glasser, M. L. (1986). On sums of log-normal random variables. *Studies in Applied Mathematics*, 75: 37–55. MR0850438. doi: https://doi.org/10.1002/sapm198675137. 537

Berrocal, V. J., Gelfand, A. E., Holland, D. M. (2010). A bivariate spatio-temporal downscaler under space and time misalignment. *Annals of Applied Statistics*, 4: 1942–1975. MR2829942. doi: https://doi.org/10.1214/10-AOAS351. 550

Chagneau, P., Mortier, F., Picard, N., Bacro, J. N. (2010). A hierarchical Bayesian model for spatial prediction of multivariate non-Gaussian random fields. *Biometrics*, 67: 97–105. MR2898821. doi: https://doi.org/10.1111/j.1541-0420.2010.01415.x. 531

Chen, M.-H. (2004). Skewed link models for categorical response data. Chap. 8, pages 131–152 of: Genton, M. G. (ed.), *Skew-elliptical Distributions and their Applications: A Journey Beyond Normality*. Boca Raton, FL: Chapman and Hall/CRC. MR2155327. 542

Chen, M.-H., Dey, D. K., Shao, Q.-M. (1999). A new skewed link model for dichotomous quantal response data. *Journal of the American Statistical Association*, 94: 1172–1186. MR1731481. doi: https://doi.org/10.2307/2669933. 542

Gelfand, A. E. (2010). Misaligned spatial data: The change of support problem. In: Gelfand, AE.; Diggle, P.; Guttorp, P.; Fuentes, M., editors. *Handbook of Spatial Statistics*. Boca Raton, FL: CRC Press, 517–539. MR2730964. doi: https://doi.org/10.1201/9781420072884-c29. 533

Gelman, A., Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7: 457–511. 548

Greene, W. H. (1990). A Gamma distributed stochastic frontier model. *Journal of Econometrics*, 46: 141–164. MR1080869. doi: https://doi.org/10.1016/0304-4076(90)90052-U. 532

Higgs, M. D., Hoeting, J. A. (2010). A clipped latent variable model for spatially correlated ordered categorical data. *Computational Statistics and Data Analysis*, 54 (8): 1999–2011. MR2640303. doi: https://doi.org/10.1016/j.csda.2010.02.024. 531, 546

Hogan, J. W., Tchernis, R. (2004). Bayesian factor analysis for spatially correlated data, with application to summarizing area-Level material deprivation from census data. *Journal of the American Statistical Association*, 99: 314–324. MR2109313. doi: https://doi.org/10.1198/016214504000000296. 541

Johnson, N. L., Kotz, S., Balakrishnan, N. (1994). *Continuous Univariate Distributions*. New York: John Wiley, 2 ed. MR1299979. 537

Kass, R. E., Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90: 773–795. MR3363402. doi: https://doi.org/10.1080/01621459.1995.10476572. 541

Kim, H., Mallick, B. K. (2004). A Bayesian prediction using the skew-Gaussian processes. *Journal of Statistical Planning and Inference*, 120: 85–101. MR2026484. doi: https://doi.org/10.1016/S0378-3758(02)00501-3. 532

Lindgren, F., Rue, H., Lindstrom, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Association. Series B*, 73: 423–498. MR2853727. doi: https://doi.org/10.1111/j.1467-9868.2011.00777.x. 553

Neal, R. (2003). Slice sampling. *Annals of Statistics*, 31: 705–767. MR1994729. doi: https://doi.org/10.1214/aos/1056562461. 533, 546

Newton, M. A., Raftery, A. E. (1994). Approximate Bayesian inference by the weighted likelihood Bootstrap (with discussion). *Journal of the Royal Statistical Society. Series B*, 3: 3–48. MR1257793. 541

Recta, V., Haran, M., Rosenberger, J. L. (2012). A two-stage model for incidence and prevalence in point-level spatial count data. *Environmetrics*, 23(2): 162–174. MR2897343. doi: https://doi.org/10.1002/env.1129. 533

Ren, Q., Banerjee, S. (2013). Hierarchical factor models for large spatially misaligned data: A low-rank predictive process approach. *Biometrics*, 69: 19–30. MR3058048. doi: https://doi.org/10.1111/j.1541-0420.2012.01832.x. 553

Roshandel, G., Sadjadi, A., Aarabi, M., Keshtkar, A., Sedaghat, S. M., Nouraie, S. M., Semnani, S., Malekzadeh, R. (2012). Cancer incidence in Golestan province: report of an ongoing population-based cancer registry in Iran, 2004–2008. *Archives of Iranian Medicine*, 15(4): 196–200. 550

Royle, J., Berliner, L. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, 4(1): 29–56. MR1812239. doi: https://doi.org/10.2307/1400420. 533

Rue, H., Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman and Hall-CRC Press. MR2130347. doi: https://doi.org/10.1201/9780203492024. 553

Schliep, E. M., Hoeting, J. A. (2013). Multilevel latent Gaussian process model for mixed discrete and continuous multivariate response data. *Journal of Agricultural, Biological, and Environmental Statistics*, 18: 492–513. MR3142597. doi: https://doi.org/10.1007/s13253-013-0136-z. 531, 533, 541, 542

Silver, J., Ritchie, M. E., Smyth, G. K. (2009). Microarray background correction: maximum likelihood estimation for the normal-exponential convolution model. *Biostatistics*, 10: 352–363. 532

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory of Kriging*. New York: Springer-Verlag. MR1697409. doi: https://doi.org/10.1007/978-1-4612-1494-6. 540

Tanner, M. A., Wong, W. H. (1987). The calculation of posterior distributions by data augmentation (with discussion). *Journal of the American Statistical Association*, 82: 528–550. MR0898357. 533

Wang, F., Wall, M. M. (2003). Generalized common spatial factor model. *Biostatistics*, 4: 569–582. 553

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, 41: 434–449. MR0067450. doi: https://doi.org/10.1093/biomet/41.3-4.434. 540

Zareifard, H., 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. MR2993870. doi: https://doi.org/10.1016/j.jmva.2012.07.003. 532, 534

Zareifard, H., Khaledi, M. J., Rivaz, F. and Vahidi-Asl, M. Q. (2017). Supplementary Material of "Modeling Skewed Spatial Data Using a Convolution of Gaussian and Log-Gaussian Processes". *Bayesian Analysis*. doi: https://doi.org/10.1214/17-BA1064SUPP. 545, 547

Zhang, H., El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, 21: 33–47. MR2842222. doi: https://doi.org/10.1002/env.982. 532, 534

Acknowledgments

The Editor, Associated Editor and two referees are gratefully acknowledged. Their very much precise comments and constructive suggestions have substantially improved the manuscript.