

Spatial models for non-Gaussian data with covariates measurement error

V. Tadayon^{1,2} | M. Torabi^{*3}

¹Department of Statistics, Shahid Chamran University of Ahvaz, Ahvaz, Iran.

²Department of Statistics, Higher Education Center of Eghlid, Eghlid, Iran.

Email: VahidTadayon24@gmail.com

³Departments of Community Health Sciences and Statistics, University of Manitoba, Winnipeg, Manitoba, Canada.

Email: Mahmoud.Torabi@umanitoba.ca

Correspondence

*M. Torabi,

Department of Community Health Sciences, University of Manitoba, Winnipeg, Manitoba, Canada, R3E 0W3.

Email: Mahmoud.Torabi@umanitoba.ca

Summary

Spatial models have been widely used in the public health set-up. In the case of continuous outcomes, the traditional approaches to model spatial data are based on the Gaussian distribution. This assumption might be overly restrictive to represent the data. The real data could be highly non-Gaussian and may show features like heavy tails and/or skewness. In spatial data modeling, it is also commonly assumed that the covariates are observed without errors, but for various reasons such as measurement techniques or instruments used, uncertainty is inherent in spatial (especially geostatistics) data and so these data are susceptible to measurement error in the covariates of interest. In this paper, we introduce a general class of spatial models with covariates measurement error that can account for both heavy tails, skewness, and also uncertainty of the covariates. A likelihood method, which leads to maximum likelihood estimation approach, is used for the inference through Monte Carlo Expectation-Maximization algorithm. The predictive distribution at non-sampled sites is approximated based on Markov chain Monte Carlo algorithm. The proposed approach is evaluated through a simulation study and also by a real application (particulate matters dataset).

KEYWORDS:

MCEM Algorithm, Measurement Error, Spatial Modeling, Unified Skew Gaussian.

1 | INTRODUCTION

One of the main difficulty faced by spatial data analysis (especially geostatistical data) is to adjust for non-Gaussianity feature. This feature may manifest itself in the exploratory data analysis by demonstrating heavy tail

or skewness. Heavy tails can be caused by special cases which have extreme values compared to their neighboring observation, while asymmetry in the distribution of the response (as another non-Gaussianity's symptoms) may be caused by skewness. The presence of skewness arises in many studies including the spatial prediction problems. Again, in such a setting, to describe the characteristics, normality may not be a realistic assumption. In addition, a real data may contain the presence of outliers and skewness simultaneously. Some examples of such data can be found in [Zareifard and Khaledi \(2013\)](#), [Tadayon and Khaledi \(2015\)](#) and [Tadayon \(2017\)](#). A widely used approach to take data sets with non-Gaussian characteristics into account is to find some non-linear transformation (e.g., a log transformation or a transformation from the Box-Cox family) for the data to satisfy the assumption of normality. However, in general, an appropriate transformation may not exist or may be difficult to find. In addition to the problem of choosing an appropriate transformation, the transformed variables are sometimes difficult to interpret ([Kim & Mallick 2004](#)). By applying a transformation, the mean field will also affect the resultant covariance structure ([Wallin & Bolin 2015](#)). Evidently the situation is also exacerbated in multivariate settings with several spatial response variables since it is hard to provide a joint transformation of the entire process into a Gaussian process. In addition, the back-transformed fitted model produces severely biased estimates ([Cressie 1993](#); [Miller 1984](#)). In the recent years, authors have developed more suitable theoretical strategy to handle some of the potential weaknesses associated with the transformation method.

Modeling of non-Gaussian geostatistical data based on a process with fat-tailed finite-dimensional distributions was studied by [Palacios and Steel \(2006\)](#). They introduced a Gaussian-log-Gaussian (GLG) model that accommodates non-Gaussian tail behavior in space. Their proposed model which is based on a scale mixing of a Gaussian process, leads to the Gaussian model as a limiting case. They also adopted the Matérn class for the spatial correlation structure ([Stein 2012](#)) and used a Markov chain Monte Carlo (MCMC) algorithm to perform Bayesian inference. More details on the GLG model can be found in [Steel and Fuentes \(2010\)](#). Modeling of spatial data with heteroscedasticity in space or time has been studied by [Fonseca and Steel \(2011\)](#) wherein the finite dimensional distributions have heavier tails than the normal distribution. They extended the ideas in [Palacios and Steel \(2006\)](#)

to processes in space and time and also considered similar mixing in the nugget effect component to accommodate individual outliers. Indeed, in their model, the non-Gaussian behavior was induced by correlated mixing in the spatio-temporal process both in time and space and uncorrelated mixing in the nugget effect. However, this approach is suitable only for symmetric heavy tail distributions and will fail to handle the skewed data. Recently, [Bueno, Fonseca, and Schmidt \(2017\)](#) studied a GLG spatio-temporal model based on spatially varying covariates.

When the observed data have a skew distribution, again an initial strategy is to make data transformation ([De-Oliveira, Kedem, & Short 1997](#)). The issues associated with the transformation method has received considerable attention in the literature. This led to introduce more suitable theoretical strategies based on the skew-normal (SN) distribution (see e.g., [Azzalini 1985](#); [Azzalini and Capitanio 1999](#); [Azzalini and Dalla-Valle 1996](#); [Azzalini 2013](#)) which resulted to numerous developments that sometimes confusing which class of SN model needs to be used. With this view in mind, [Arellano-Valle and Azzalini \(2006\)](#) introduced a SN model and named it *unified skew-normal* (SUN) which also includes the normal density and has very similar properties as normal density. We will review this model in more detail in [Section 2](#).

On the other hand, in order to circumvent the problem of departure from normality in the spatial domains, many authors concentrated on the introduction of stationary stochastic processes, whose finite-dimensional marginal distributions are multivariate skew-normal. To handle skewed spatial data, [Kim and Mallick \(2004\)](#) defined a skew Gaussian random field based on the multivariate skew normal distribution. With the aim of introducing a spatial skewed Gaussian process, [Allard and Naveau \(2007\)](#) used the multivariate closed SN distribution which also has closed forms under marginalization and conditioning. Another example of a stationary stochastic process in the spatial domain with univariate SN marginal distributions was given by [Zhang and El-Shaarawi \(2010\)](#), exploiting one of the stochastic characterizations of the SN distribution. In particular, the authors used $X_0(s)$ as a stationary Gaussian process with standardized marginals in place of normal random variable X_0 in the multivariate extension of skew-normal distribution defined by [Azzalini and Dalla-Valle \(1996\)](#) and considered an extension of the SN distribution to a stationary process. For more details about the skew-Gaussian spatial random fields, we refer interested readers to [Genton and Zhang \(2012\)](#).

The unified skew GLG model which recently proposed by [Zareifard and Khaledi \(2013\)](#) offers a more flexible class of sampling spatial models to account for both skewness and heavy tails which are two pervasive features of the spatial data. Since the likelihood function involves analytically intractable integrals and direct maximization of the marginal likelihood is numerically difficult, they developed a likelihood-based approach for the inference and a stochastic approximation version of EM algorithm for estimating the model parameters. [Tadayon \(2017\)](#) used this model to take censored spatial data into account in analyzing rainfall records.

In spatial data modeling, it is commonly assumed that the covariates have been observed precisely, but for various reasons such as the measurement techniques or instruments used, uncertainty is inherent and so these data are susceptible to measurement error in the covariate of interest. For example in environmental studies, the pollution level, e.g., nitrite concentration in groundwater or precipitation and wind speed are measured with error. In such a setting, the results provided by ignoring the measurement error can be biased and unreliable. In the recent years, some authors have suggested to accommodate measurement error in the spatial context. [Li, Tang, and Lin \(2009\)](#) proposed a new class of linear mixed models for spatial data in the presence of covariate measurement errors. In particular, they showed that the regression estimates obtained from naive use of a proxy variable are attenuated while the naive estimators of the variance components are inflated. The authors also, by taking measurement error into account, developed a structural modeling approach to obtain the maximum likelihood estimation (MLE) of model parameters and studied its large sample properties. [Le-Gallo and Fingleton \(2012\)](#) considered the case of cross-sectional spatial regression models with measurement errors in the explanatory variables and demonstrated that measurement error in an independent variable as one reason why ordinary least squares estimates may not be consistent. [Militino, Ugarte, Iribas, and Lizarraga-Garcia \(2013\)](#) studied to address the quality of the Global Positioning System (GPS) measurements using a likelihood-based approach for the analysis of positional errors based on a spatial linear mixed model. In line with what observed in [Li et al. \(2009\)](#), [Huque, Bondell, and Ryan \(2014\)](#) showed that the presence of covariate measurement error can lead to significant sensitivity of parameter estimation to the choice of spatial correlation structure. However, their approaches require correct specification of the true covariate measurement error variance. In contrast, [Huque, Bondell, Carroll, and Ryan \(2016\)](#) presented an

approach which is robust in the sense that it neither assumes that the covariate measurement error is known nor depends on any particular kind of spatial correlation structure. In fact, they developed a semi-parametric framework to obtain a consistent estimate of the true regression coefficients when covariates are measured with error. Alexeeff, Carroll, and Coull (2016) proposed a new spatial simulation extrapolation (SIMEX) procedure and derived the biases induced by estimation error and by model misspecification in the exposure model. For more details in this subject, see Gryparis, Coull, and Schwartz (2007), Szpiro, Sheppard, and Lumley (2011), Sheppard et al. (2012) and Thomas (2013).

The aim of this paper is to provide a more flexible modeling of spatial data which can be used for non-Gaussian data wherein the covariates of interest cannot be observed precisely. In particular, our model provides flexibility in capturing the effects of skewness and heavy tails behavior of the data and simultaneously facilitates representing and taking fuller account of the susceptibility of measurement error in covariates. As measurement error (ME) is commonly due solely to instrument or laboratory-analysis error in geostatistical data, the classical measurement error model appears appropriate for this situation as we would expect the surrogate measure to be randomly distributed around the true value. The unobserved covariates can be treated as fixed parameters (called functional ME) or random variables (called structural ME). In this paper, we focus on structural ME which may lead to more efficiency (Buzas, Stefanski, & Tosteson 2014) in the context of geostatistical data. In this paper, a likelihood-based approach is used for the inference. In a nutshell, our major contribution on this paper is to account for the uncertainty of covariates measured with error in the context of spatial models with non-Gaussian data.

The rest of the paper is organized as follows. In Section 2, a class of spatial model for non-Gaussian data with structural ME covariates is proposed. We use a likelihood-based approach, which leads to the MLE, to estimate the model parameters and corresponding variances (Section 3). Spatial prediction is studied in Section 4. Performance of the proposed approach is evaluated through a simulation study (Section 5) and also by a real data application (particulate matters dataset from Iran) (Section 6). Concluding remarks are given in Section 7. Technical details are deferred to the Appendix.

2 | MODEL FORMULATION

Following [Zareifard and Khaledi \(2013\)](#), let $Y(s_i)$ be the response variable at location s_i in the spatial region. We assume that for $i = 1, 2, \dots, n$,

$$Y(s_i) = \beta_0 + X'(s_i) \boldsymbol{\beta}_x + \frac{W(s_i)}{\sqrt{\lambda(s_i)}} + \varepsilon(s_i), \quad (1)$$

where β_0 represents the intercept term, $\boldsymbol{\beta}_x = (\beta_1, \beta_2, \dots, \beta_k)'$ are unknown parameters with corresponding $X(\cdot)$ as a random vector of error-prone covariates, $\varepsilon(\cdot)$ denotes an uncorrelated Gaussian process with zero mean and known variance σ^2 to avoid overparameterization and identifiability problems, and $\lambda(\cdot)$ is a log-Gaussian stochastic process. The $\mathbf{W}(\cdot) = (W(s_1), \dots, W(s_n))'$ is a SUN process with the skewness parameter α and the scale parameter γ which is denoted by

$$SUN_{n,n}(\mathbf{0}_n, \mathbf{0}_n, [\gamma^2 + \alpha^2] C_\theta, C_\theta, \alpha C_\theta), \quad \alpha \in \mathfrak{R}, \gamma \in \mathfrak{R}^+, \quad (2)$$

where C_θ is a spatial correlation matrix. More precisely, [Arellano-Valle and Azzalini \(2006\)](#) introduced a SUN process as follows. Suppose that \mathcal{G}_0 and \mathcal{G}_1 be two random vectors of dimensions m and n , respectively, such that

$$\begin{bmatrix} \mathcal{G}_0 \\ \mathcal{G}_1 \end{bmatrix} \sim N_{m+n} \left(\begin{bmatrix} \varphi_0 \\ \varphi_1 \end{bmatrix}, \begin{bmatrix} \Gamma & \Delta' \\ \Delta & \Omega \end{bmatrix} \right), \quad (3)$$

where $\varphi_0 \in \mathfrak{R}^m$ and $\varphi_1 \in \mathfrak{R}^n$ are mean vectors, $\Gamma \in \mathfrak{R}^{m \times m}$ and $\Omega \in \mathfrak{R}^{n \times n}$ are variance-covariance matrices and $\Delta = (\delta_1, \dots, \delta_n)' \in \mathfrak{R}^{n \times m}$. Then, the n -dimensional random vector \mathcal{G} is said to have the multivariate SUN distribution if $\mathcal{G} \stackrel{d}{=} \mathcal{G}_1 | \mathcal{G}_0 > \mathbf{0}_m$, where $\mathbf{0}_m$ is a $m \times 1$ vector of 0's. This distribution which has the density

$$f_{SUN_{n,m}}(g) = \phi_n(\varphi_1, \Omega) \Phi_m(\varphi_0 + \Delta' \Omega^{-1}(g - \varphi_1), \Gamma - \Delta' \Omega^{-1} \Delta) / \Phi_m(\varphi_0, \Gamma), \quad (4)$$

is denoted by $SUN_{n,m}(\varphi_1, \varphi_0, \Omega, \Gamma, \Delta)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ represent normal density and normal cumulative distribution function, respectively. The density in (4) reduces to the n -variate normal density function $\phi_n(g; \varphi_1, \Omega)$ when $\Delta = \mathbf{0}_{n \times m}$, noting that $\mathbf{0}_{n \times m}$ is a $n \times m$ matrix of zeros. [Arellano-Valle and Azzalini \(2006\)](#) presented

another form of genesis for distribution (4), using a convolution mechanism instead of conditioning in which a $SUN_{n,n}(\mathbf{0}_n, \mathbf{0}_n, [\gamma^2 + \alpha^2]V, V, \alpha V)$ distribution can be written as $\gamma T_0 + \alpha T_1$, $\gamma \in \mathfrak{R}^+$, $\alpha \in \mathfrak{R}$, where $T_0 \sim N_n(\mathbf{0}_n, V)$ and T_1 is a n -variate normal distribution with mean $\mathbf{0}_n$ and variance V which is truncated below at the point $\mathbf{0}_n$. It is obvious that $\alpha = 0$ leads to the n -variate normal distribution.

Since small values of the mixing variables $\lambda(s_i)$ lead to locate the corresponding observation $Y(s_i)$ in a region with larger observational variance relative to the rest, similar to Palacios and Steel (2006), we call these observations as *outliers*. Furthermore, we suppose that the random field $\boldsymbol{\psi}(\cdot) \equiv \log \lambda(\cdot) = (\log \lambda(s_1), \dots, \log \lambda(s_n))'$ is Gaussian with mean $-\frac{\nu}{2}\mathbf{1}_n$ ($\mathbf{1}_n$ is a $n \times 1$ vector of ones) and covariance matrix νC_θ , wherein the scalar parameter ν is a non-negative value. Regarding the spatial correlation matrix C_θ , it is possible to consider different correlation matrices for $\mathbf{W}(\cdot)$ and $\boldsymbol{\psi}(\cdot)$, however, for the purpose of the model complexity reduction, we correlate the elements of $\boldsymbol{\psi}(\cdot)$ through the same correlation matrix as $\mathbf{W}(\cdot)$ which is adopted as an isotropic exponential correlation structure with the correlation function $C_\theta(\|s_i - s_j\|) = C_\theta(\|h\|) = \exp\{-\theta \|h\|\}$ and $\theta > 0$. Further, this approach also prevents from the identifiability issue as choosing different correlation structure increases the model parameters and also model cannot separately identify different correlation.

In the presence of ME one is unable to observe $X_i = X(s_i)$, ($i = 1, \dots, n$), but $Z_i = Z(s_i)$ is observed as surrogate for the X_i through $Z_i = X_i + U_i$, where X_i has a normal distribution with mean $\boldsymbol{\mu}_x$ and variance-covariance matrix $\sigma_x^2 I_k$, $U_i = U(s_i)$ is a random vector from normal distribution with mean $\mathbf{0}_k$ and variance-covariance matrix $\sigma_u^2 I_k$ with I_k as the identity matrix of dimension k , and σ_u^2 is known (Hoque & Torabi 2018; Torabi 2012; Torabi, Datta, & Rao 2009). We can then write

$$(X'_i, U'_i, \varepsilon_i)' \sim N\left([\boldsymbol{\mu}'_x, \mathbf{0}'_k, 0]', \text{diag}(\sigma_x^2 I_k, \sigma_u^2 I_k, \sigma^2)\right), \quad i = 1, \dots, n, \quad (5)$$

where $\text{diag}(\sigma_x^2 I_k, \sigma_u^2 I_k, \sigma^2)$ is a diagonal matrix with the given elements on the diagonal. We also suppose that the random vector $(X'_i, U'_i, \varepsilon_i)'$ in (5) is independent of the other model components, namely, $\boldsymbol{\lambda}$ and \mathbf{W} . Thus, if $\sigma_z^2 = \sigma_x^2 + \sigma_u^2$, then $Z_i \sim N_k(\boldsymbol{\mu}_x, \sigma_z^2 I_k)$ which, using the properties of the SUN distribution (Gupta, Aziz, & Ning 2013), can be rewritten as $Z_i \sim SUN_{k,k}(\boldsymbol{\mu}_x, \mathbf{0}_k, \sigma_z^2 I_k, I_k, \mathbf{0}_{k \times k})$.

Note that if one assumes $\lambda(s_i) = \lambda \sim \text{Gamma}\left(\frac{\nu}{2}, \frac{\nu}{2}\right) = \chi_\nu^2/\nu$, then the equation (2) can be written as $\mathbf{W}(\cdot) \sim \text{SN}_n(\alpha; \mathbf{0}_n, [\gamma^2 + \alpha^2] C_\theta)$. In such a setting, the distribution of $\mathbf{Y} = (Y_1, \dots, Y_n)'$ given $\mathbf{X}_{n \times k} = (X_1, \dots, X_k)$ will be skew Student-t (SSt) with the skewness parameter α , mean $\beta_0 \mathbf{1}_n + \mathbf{X} \boldsymbol{\beta}_x$, variance $[\gamma^2 + \alpha^2] C_\theta + \sigma^2 I_n$, and ν degrees of freedom as $\mathbf{Y} | \mathbf{X} \sim \text{SSt}(\alpha, \nu; \beta_0 \mathbf{1}_n + \mathbf{X} \boldsymbol{\beta}_x, [\gamma^2 + \alpha^2] C_\theta + \sigma^2 I_n)$. Although the SSt model allows for variance inflation, it increases the kurtosis for the process in every location and does not allow for individual changes in variability (Palacios & Steel 2006). However, the proposed model allows for spatial variability in $\lambda(s_i)$ and hence is more flexible.

3 | MAXIMUM LIKELIHOOD ESTIMATION

Following the model (1), the likelihood of the observed data (\mathbf{Y}, \mathbf{Z}) , for $\mathbf{Z}_{n \times k} = (Z_1, \dots, Z_k)$ given \mathbf{X}, λ can be written as

$$L(\mathbf{Y}, \mathbf{Z} | \mathbf{X}, \lambda) \propto L(\mathbf{Y} | \mathbf{X}, \lambda) L(\mathbf{Z} | \mathbf{X}, \lambda), \quad (6)$$

where $L(\mathbf{Y}, \mathbf{Z} | \lambda) = \int L(\mathbf{Y} | \mathbf{X}, \lambda) L(\mathbf{Z} | \mathbf{X}, \lambda) L(\mathbf{X} | \lambda) d\mathbf{X}$. Since we intend to implement the maximum likelihood approach, we first obtain the distributions of $\mathbf{Y} | \mathbf{X}, \lambda$ and $\mathbf{Z} | \mathbf{X}, \lambda$ as follows:

- $\mathbf{Y} | \mathbf{X}, \lambda$. It can be easily seen that if $\Lambda = \text{diag}(\lambda(s_1), \dots, \lambda(s_n))$, then by (2) we have

$$\Lambda^{-\frac{1}{2}} \mathbf{W} | \lambda \sim \text{SUN}_{n,n}(\mathbf{0}_n, \mathbf{0}_n, [\gamma^2 + \alpha^2] \Lambda^{-\frac{1}{2}} C_\theta \Lambda^{-\frac{1}{2}}, C_\theta, \alpha \Lambda^{-\frac{1}{2}} C_\theta),$$

and also $\boldsymbol{\varepsilon} | \lambda \sim \text{SUN}_{n,n}(\mathbf{0}_n, \mathbf{0}_n, \sigma^2 I_n, I_n, \mathbf{0}_{n \times n})$. Hence, from (1) and again by using the properties of the SUN distribution, we can get

$$\Lambda^{-\frac{1}{2}} \mathbf{W} + \boldsymbol{\varepsilon} | \lambda \sim \text{SUN}_{n,2n}(\mathbf{0}_n, \mathbf{0}_{2n}, [\gamma^2 + \alpha^2] \Lambda^{-\frac{1}{2}} C_\theta \Lambda^{-\frac{1}{2}} + \sigma^2 I_n, C_\theta \oplus I_n, \left[\alpha \Lambda^{-\frac{1}{2}} C_\theta, \mathbf{0}_{n \times n} \right]),$$

where $C_\theta \oplus I_n = \begin{pmatrix} C_\theta & \mathbf{0}_{n \times n} \\ \mathbf{0}_{n \times n} & I_n \end{pmatrix}$. So if $\mathbf{X}_* = (\mathbf{1}_n, X_1, \dots, X_k)$ and $\boldsymbol{\beta} = (\beta_0, \boldsymbol{\beta}'_x)'$, then

$$\mathbf{Y}|\mathbf{X}, \lambda \sim SUN_{n,2n} \left(\mathbf{X}_* \boldsymbol{\beta}, \mathbf{0}_{2n}, [\gamma^2 + \alpha^2] \Lambda^{-\frac{1}{2}} C_\theta \Lambda^{-\frac{1}{2}} + \sigma^2 I_n, C_\theta \oplus I_n, \left[\alpha \Lambda^{-\frac{1}{2}} C_\theta, \mathbf{0}_{n \times n} \right] \right). \quad (7)$$

- $\mathbf{Z}|\mathbf{X}, \lambda$. Similarly, one can show

$$\mathbf{Z}|\mathbf{X}, \lambda \sim SUN_{k,k} \left(\mathbf{X}, \mathbf{0}_k, \sigma_u^2 I_k, I_k, \mathbf{0}_{k \times k} \right). \quad \blacksquare$$

For brevity, we can write $\Omega = [\gamma^2 + \alpha^2] \Lambda^{-\frac{1}{2}} C_\theta \Lambda^{-\frac{1}{2}} + \sigma^2 I_n$, $\Gamma = C_\theta \oplus I_n$ and $\Delta = \left[\alpha \Lambda^{-\frac{1}{2}} C_\theta, \mathbf{0}_{n \times n} \right]$. We also regard \mathbf{y} as a realization of \mathbf{Y} . Therefore, after some simplification, we can write

$$\begin{aligned} f_{\mathbf{Y},\mathbf{Z}|\mathbf{X},\lambda} &\propto \phi_n(\mathbf{X}_* \boldsymbol{\beta}, \Omega) \Phi_{2n}(\Delta' \Omega^{-1} \mathbf{y}, \Gamma - \Delta' \Omega^{-1} \Delta) \phi_n(\mathbf{X}_* \boldsymbol{\beta}, \sigma_u^2 I_n) / \Phi_{2n}(\mathbf{0}_{2n}, \Gamma) \\ &\propto \frac{\Phi_{2n}(\Delta' \Omega^{-1} \mathbf{y}, \Gamma - \Delta' \Omega^{-1} \Delta)}{\Phi_{2n}(\mathbf{0}_{2n}, \Gamma)} \exp \left\{ -\frac{1}{2} (\mathbf{y} - \Sigma \boldsymbol{\mu}_y) \Sigma^{-1} (\mathbf{y} - \Sigma \boldsymbol{\mu}_y) \right\}, \end{aligned} \quad (8)$$

where $\Sigma^{-1} = \Omega^{-1} + \sigma_u^{-2} I_n$ and $\boldsymbol{\mu}_y = \Sigma^{-1} \mathbf{X}_* \boldsymbol{\beta}$. In other words, we have

$$f_{\mathbf{Y},\mathbf{Z}|\mathbf{X},\lambda} \propto \phi_n(\Sigma \boldsymbol{\mu}_y, \Sigma) \Phi_{2n}(\Delta' \Omega^{-1} \mathbf{y}, \Gamma - \Delta' \Omega^{-1} \Delta) / \Phi_{2n}(\mathbf{0}_{2n}, \Gamma). \quad (9)$$

Hence, the likelihood function of $\eta = (\beta_0, \boldsymbol{\beta}_x, \boldsymbol{\mu}_x, \sigma_x^2, \gamma, \alpha, \nu, \theta)$ given the observed sample ($\mathbf{Y} = \mathbf{y}, \mathbf{Z} = \mathbf{z}$) is given by $L(\eta|\mathbf{y}, \mathbf{Z}) = \int f_{\mathbf{Y},\mathbf{Z}|\mathbf{X},\lambda} dP_x dP_\lambda$, where P_x and P_λ show the distribution measure of \mathbf{X} and λ , respectively. However, we can not maximize the likelihood function of observed data directly as the $L(\eta|\mathbf{y}, \mathbf{Z})$ is analytically intractable. Hence, we consider a version of the EM algorithm (Monte Carlo EM) to estimate the model parameters η . One can show that under some mild regularity conditions, the MCEM converges to the maximum likelihood estimate (Neath et al. 2013). This method presents a modification of the EM algorithm where the expectation in the *E-step* is computed numerically through Monte Carlo simulation.

The basis idea of the algorithm is to use the log-likelihood of the complete data that has not been observed due to the presence of the latent variables \mathbf{X} , \mathbf{W} and λ . With regard to the definition of the *SUN* distribution in Section 1, we can rewrite \mathbf{W} as a sum of two independent random vectors, i.e., $\mathbf{W} \stackrel{d}{=} T_0 + T_1$, where $T_0 \sim N_n(\mathbf{0}_n, \gamma^2 C_\theta)$,

T_1 is n -variate normal distribution with mean $\mathbf{0}_n$ and variance $\alpha^2 C_\theta$ which is truncated below at the point $\mathbf{0}_n$ (that is followed by right-skewed responses, i.e., $\alpha > 0$). Hence, the model (1) can be formulated in a matrix form as

$$\mathbf{Y} = \beta_0 \mathbf{1}_n + \mathbf{X}\boldsymbol{\beta}_x + \Lambda^{-\frac{1}{2}} (T_0 + T_1) + \boldsymbol{\varepsilon}, \quad \mathbf{Z} = \mathbf{X} + \mathbf{U}, \quad \mathbf{U}_{n \times k} = (\mathbf{u}_1, \dots, \mathbf{u}_k). \quad (10)$$

It must be noted that although for now we chose the positive sign for α , it can be visually determined by the direction of skewness. This approach works well if the direction of skewness is clear. When the direction of skewness is not obvious, there is only a limited interest in modeling the dataset with a skew-normal random field. Henceforth, the complete data will be $(\mathbf{Y}, \mathbf{Z}, \mathbf{X}, T_0, T_1, \lambda)$, where (\mathbf{Y}, \mathbf{Z}) are observed data and $(\mathbf{X}, T_0, T_1, \lambda)$ are latent variables. The complete data log-likelihood function (see Appendix A for details) is given by

$$\begin{aligned} \ell(\eta) &= \log f_{\mathbf{Y}, \mathbf{Z} | \mathbf{X}, T_0, T_1, \lambda}(\mathbf{y}, \mathbf{Z}; \boldsymbol{\beta}) + \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\mu}_x, \sigma_x^2) \\ &\quad + \log f_{T_0}(\mathbf{t}_0; \gamma^2, \theta) + \log f_{T_1}(\mathbf{t}_1; \alpha^2, \theta) + \log f_\lambda(\lambda; \nu, \theta). \end{aligned} \quad (11)$$

To implement the MCEM algorithm, let η^t be the current (step t) best guess at the MLE $\hat{\eta}$. The *E-step* is to compute the Q -function defined by

$$Q(\eta | \eta^t) = E \{ \ell(\eta) | \mathbf{y}, \mathbf{Z}, \eta^t \} = \int \ell(\eta) f(\mathbf{X}, \mathbf{t}_0, \mathbf{t}_1, \lambda | \mathbf{y}, \mathbf{Z}, \eta^t) d\mathbf{X} d\mathbf{t}_0 d\mathbf{t}_1 d\lambda, \quad (12)$$

using

$$\mathbf{X} | \mathbf{Z} \sim N_k(a\mathbf{Z} + [1-a]\boldsymbol{\mu}_x, \sigma_x^2 [1-a] I_k), \quad a = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2}. \quad (13)$$

The *M-step* is to maximize the Q with respect to η to obtain $\eta^{t+1} = \arg \max_{\eta \in \Theta} Q(\eta | \eta^t)$, where Θ is the parameter space. Due to the complexity of the conditional expectations in (12), we replace them by corresponding Monte Carlo approximations. In particular, we can write

$$Q(\eta | \eta^t) \approx \frac{1}{L} \sum_{l=1}^L \ell(\eta^t; \mathbf{y}, \mathbf{Z}, \mathbf{X}^{(l)}, \mathbf{t}_0^{(l)}, \mathbf{t}_1^{(l)}, \lambda^{(l)}),$$

and then an optimization procedure can be employed to maximize $Q(\eta|\eta^t)$ with respect to η . The two-step process is repeated until convergence occurs. At the t^{th} iteration of the MCEM algorithm, we use the notation \mathbb{E}_i^t for $i = 1, 2, \dots, 8$ to show the approximation of the i^{th} conditional expectation in (12), i.e., if $\left\{ \left(\mathbf{X}^{(l)}, T_0^{(l)}, T_0^{(l)}, \lambda^{(l)} \right) \right\}_{l=1}^L$ are samples from the joint distribution $f_{\mathbf{X}, T_0, T_1, \lambda | \mathbf{y}, \mathbf{Z}}$, then the Monte Carlo approximation of $E[g(\mathbf{X}, T_0, T_1, \lambda | \mathbf{y}, \mathbf{Z})]$ is given by

$$\mathbb{E}^t = L^{-1} \sum_{l=1}^L g\left(\mathbf{y}, \mathbf{Z}, \mathbf{X}^{(l)}, T_0^{(l)}, T_0^{(l)}, \lambda^{(l)}; \eta^t\right).$$

For details regarding the updates of the model parameters, see the Appendix B. As the conditional expectations cannot be evaluated in closed forms, a MCMC algorithm can be used to generate samples from the joint distribution $f(\mathbf{X}, T_0, T_1, \lambda | \mathbf{y}, \mathbf{Z}; \eta^t)$. To that end, we explore the full conditional distributions as:

- $\mathbf{X} | \mathbf{y}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1, \lambda$:

$$\begin{aligned} f(\mathbf{X} | \mathbf{y}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1, \lambda; \eta^t) &\propto f(\mathbf{Y} | \mathbf{X}, \mathbf{t}_0, \mathbf{t}_1, \lambda; \eta^t) f(\mathbf{X} | \mathbf{Z}; \eta^t) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{b}_1 - \mathbf{X}\beta_x)' (\mathbf{b}_1 - \mathbf{X}\beta_x) - \frac{1}{2\sigma_x^2(1-a)} (\mathbf{X} - \mu_{x|z})' (\mathbf{X} - \mu_{x|z}) \right\} \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2\sigma_x^2(1-a)} \left[\sigma_x^2(1-a) \beta_x' \mathbf{X}' \mathbf{X} \beta_x \right. \right. \\ &\quad \left. \left. + \sigma^2 \mathbf{X}' \mathbf{X} - 2\sigma^2 \mathbf{X}' \mu_{x|z} - 2\sigma_x^2(1-a) \beta_x' \mathbf{X}' \mathbf{b}_1 \right] \right\}, \end{aligned} \quad (14)$$

where $\mu_{x|z}$ shows the mean of the distribution of $\mathbf{X} | \mathbf{Z}$ in (13) and $\mathbf{b}_1 = \mathbf{y} - \beta_0 \mathbf{1}_n - \Lambda^{-\frac{1}{2}} (T_0 + T_1)$. This full conditional distribution has a nonstandard form, so a Metropolis-Hastings step or sampling importance resampling (SIR) method can be used (see Appendix C for details).

- $T_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_1, \lambda$:

$$\begin{aligned} f(T_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_1, \lambda; \eta^t) &\propto f(\mathbf{Y} | \mathbf{X}, \mathbf{t}_0, \mathbf{t}_1, \lambda; \eta^t) f(T_0; \eta^t) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{b}_2 - \Lambda^{-\frac{1}{2}} T_0)' (\mathbf{b}_2 - \Lambda^{-\frac{1}{2}} T_0) - \frac{1}{2\gamma^2} T_0' C_\theta^{-1} T_0 \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[T_0' \left(\frac{1}{\sigma^2} \Lambda^{-1} + \frac{1}{\gamma^2} C_\theta^{-1} \right) T_0 - 2 \frac{1}{\sigma^2} T_0' \Lambda^{-\frac{1}{2}} \mathbf{b}_2 \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(T_0 - \frac{1}{\sigma^2} \mathbf{B}_2^{-1} \Lambda^{-\frac{1}{2}} \mathbf{b}_2 \right)' \mathbf{B}_2 \left(T_0 - \frac{1}{\sigma^2} \mathbf{B}_2^{-1} \Lambda^{-\frac{1}{2}} \mathbf{b}_2 \right) \right\}, \end{aligned}$$

and so $T_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_1, \lambda \sim N_n \left(\frac{1}{\sigma^2} B_2^{-1} \Lambda^{-\frac{1}{2}} \mathbf{b}_2, B_2^{-1} \right)$ where $\mathbf{b}_2 = \mathbf{y} - \beta_0 \mathbf{1}_n - \mathbf{X} \boldsymbol{\beta}_x - \Lambda^{-\frac{1}{2}} T_1$ and $B_2 = \frac{1}{\sigma^2} \Lambda^{-1} + \frac{1}{\gamma^2} C_\theta^{-1}$.

• $T_1 | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \lambda$:

$$\begin{aligned} f(T_1 | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \lambda; \eta^t) &\propto f(\mathbf{Y} | \mathbf{X}, \mathbf{t}_0, \mathbf{t}_1, \lambda; \eta^t) f(T_1; \eta^t) \\ &\propto \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{b}_3 - \Lambda^{-\frac{1}{2}} T_1)' (\mathbf{b}_3 - \Lambda^{-\frac{1}{2}} T_1) - \frac{1}{2\alpha^2} T_1' C_\theta^{-1} T_1 \right\} I_{\{\mathfrak{R}_{\geq 0}^n\}}(T_1) \\ &\propto \exp \left\{ -\frac{1}{2} \left[T_1' \left(\frac{1}{\sigma^2} \Lambda^{-1} + \frac{1}{\alpha^2} C_\theta^{-1} \right) T_1 - 2 \frac{1}{\sigma^2} T_1' \Lambda^{-\frac{1}{2}} \mathbf{b}_3 \right] \right\} I_{\{\mathfrak{R}_{\geq 0}^n\}}(T_1) \\ &\propto \exp \left\{ -\frac{1}{2} \left(T_1 - \frac{1}{\sigma^2} B_3^{-1} \Lambda^{-\frac{1}{2}} \mathbf{b}_3 \right)' B_3 \left(T_1 - \frac{1}{\sigma^2} B_3^{-1} \Lambda^{-\frac{1}{2}} \mathbf{b}_3 \right) \right\} I_{\{\mathfrak{R}_{\geq 0}^n\}}(T_1), \end{aligned}$$

where $\mathbf{b}_3 = \mathbf{y} - \beta_0 \mathbf{1}_n - \mathbf{X} \boldsymbol{\beta}_x - \Lambda^{-\frac{1}{2}} T_0$, $B_3 = \frac{1}{\sigma^2} \Lambda^{-1} + \frac{1}{\alpha^2} C_\theta^{-1}$ and $I_{\{\cdot\}}(\cdot)$ denotes the indicator function. Thus

$$T_1 | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \lambda \sim TN_n \left(\mathbf{0}_n; \frac{1}{\sigma^2} B_3^{-1} \Lambda^{-\frac{1}{2}} \mathbf{b}_3, B_3^{-1} \right).$$

• $\boldsymbol{\psi} | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1$:

If $\boldsymbol{\psi}_{-i}$ denotes the vector $\boldsymbol{\psi}$ without ψ_i , then

$$f(\psi_i | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1, \boldsymbol{\psi}_{-i}; \eta^t) \propto f(\mathbf{Y} | \mathbf{X}, \mathbf{t}_0, \mathbf{t}_1, \boldsymbol{\psi}; \eta^t) f(\psi_i | \boldsymbol{\psi}_{-i}; \eta^t), \quad (15)$$

where the first term i.e., the likelihood contribution, is proportional to the product of normal probability density function truncated below at the point zero as

$$\exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{b}_4 - \Lambda^{-\frac{1}{2}} T)' (\mathbf{b}_4 - \Lambda^{-\frac{1}{2}} T) \right\} \prod_i I_{\{\mathfrak{R}^+\}}(\psi_i), \quad (16)$$

$$T = T_0 + T_1, \quad \mathbf{b}_4 = \mathbf{y} - \beta_0 \mathbf{1}_n - \mathbf{X} \boldsymbol{\beta}_x, \quad \Lambda = \text{diag}(e^{\psi_1}, \dots, e^{\psi_n}),$$

and the second term is

$$\psi_i | \boldsymbol{\psi}_{-i}; \eta^t \sim N \left(-\frac{\nu}{2} + c_{\theta_i} C_{\theta_{-i}}^{-1} \left[\boldsymbol{\psi}_{-i} - \frac{\nu}{2} \mathbf{1}_{n-1} \right], \nu - c_{\theta_i} C_{\theta_{-i}}^{-1} c_{\theta_i} \right), \quad (17)$$

where c_{θ_i} denotes the i^{th} row of C_θ after omitting the i^{th} component and $C_{\theta_{-i}}$ shows C_θ after omitting c_{θ_i} and c_{θ_i} (i.e., the i^{th} row and the i^{th} column of C_θ). Obtaining a closed form for the full conditional (15) is not possible as it

does not define a standard probability distribution. Details of the MCMC algorithm are presented in the Appendix D.

We also provide the standard errors of the ML estimates. The standard errors of the ML estimates are obtained by inverting the Fisher information matrix, i.e.,

$$\widehat{Cov}(\hat{\eta}) \approx \left[\frac{1}{L} \sum_{l=1}^L S_l(\hat{\eta}) S_l'(\hat{\eta}) \right]^{-1}, \quad (18)$$

where $S_l(\hat{\eta}) = \frac{\partial}{\partial \eta} \ell(\eta; \mathbf{y}, \mathbf{Z}, \mathbf{X}^{(l)}, \mathbf{t}_0^{(l)}, \mathbf{t}_1^{(l)}, \lambda^{(l)}) \Big|_{\eta=\hat{\eta}}$. Then by taking the square root of the diagonal elements of $\widehat{Cov}(\hat{\eta})$, the standard errors of the MLEs are obtained.

4 | PREDICTION

The prediction of a generic unobserved location, e.g., s_0 , is of primary interest in the context of spatial applications.

A suitable prediction of the unobserved location can be made using the plug-in approach. In what follows, we consider the prediction of response variable at locations s_0, \dots, s_{0_p} , say $\mathbf{Y}_0 = (Y(s_0), \dots, Y(s_{0_p}))'$, which requires the joint distribution of $(\mathbf{W}', \mathbf{W}_0)'$ where \mathbf{W}_0 (and also T_{1_0} , $\boldsymbol{\psi}_0$ and \mathbf{X}_0 that are appeared below) can be defined as in \mathbf{Y}_0 . Following (3), we can write $(\mathbf{W}', \mathbf{W}_0)' \stackrel{d}{=} \left[(T_1', T_{1_0}')' \mid T_0 > \mathbf{0}_n \right]$ where

$$\begin{bmatrix} T_0 \\ T_1 \\ T_{1_0} \end{bmatrix} \sim N_{2n+p} \left(\mathbf{0}_{2n+p}, \begin{bmatrix} C_\theta^{\mathcal{O}\mathcal{O}} & \Delta^\dagger' \\ \Delta^\dagger & \Sigma^\dagger \end{bmatrix} \right), \quad \Delta^\dagger = \alpha \begin{bmatrix} C_\theta^{\mathcal{O}\mathcal{O}} \\ C_\theta^{\mathcal{P}\mathcal{O}} \end{bmatrix},$$

with $C_\theta^{\mathcal{O}\mathcal{O}} = [C_\theta(\|s_i - s_j\|)]_{n \times n}$, $C_\theta^{\mathcal{P}\mathcal{O}} = [C_\theta(\|s_{0_i} - s_j\|)]_{p \times n}$, $\Sigma^\dagger = [\gamma^2 + \alpha^2] C_\theta^\dagger$ and

$$C_\theta^\dagger = \begin{bmatrix} C_\theta^{\mathcal{O}\mathcal{O}} & C_\theta^{\mathcal{O}\mathcal{P}} \\ C_\theta^{\mathcal{P}\mathcal{O}} & C_\theta^{\mathcal{P}\mathcal{P}} \end{bmatrix}, \quad C_\theta^{\mathcal{P}\mathcal{P}} = [C_\theta(\|s_{0_i} - s_{0_j}\|)]_{p \times p}.$$

Thus, $(\mathbf{W}', \mathbf{W}'_0)' \sim SUN_{n+p, n}(\mathbf{0}_{n+p}, \mathbf{0}_n, \Sigma^\dagger, C_\theta^{\mathcal{O}\mathcal{O}}, \Delta^\dagger)$ and then the joint distribution of $(\mathbf{Y}', \mathbf{Y}'_0)'$ given $(\mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \boldsymbol{\psi}_0)$ is

$$SUN_{n+p, 2n} \left(\boldsymbol{\mu}_{y_{y_0}}, \mathbf{0}_{2n}, \Lambda^{\dagger-\frac{1}{2}} \Sigma^\dagger \Lambda^{\dagger-\frac{1}{2}} + \sigma^2 I_{n+p}, C_\theta^{\mathcal{O}\mathcal{O}} \oplus I_n, \left[\Lambda^{\dagger-\frac{1}{2}} \Delta^\dagger, \mathbf{0}_{(n+p) \times (n+p)} \right] \right), \quad (19)$$

where $\boldsymbol{\mu}_{y_{y_0}} = \beta_0 \mathbf{1}_{n+p} + \mathbf{X}_{\mathcal{O}P} \boldsymbol{\beta}_x$, $\mathbf{X}_{\mathcal{O}P} = (\mathbf{X}', \mathbf{X}'_0)'$ and $\Lambda^\dagger = \Lambda \oplus \exp\{\boldsymbol{\psi}_0\}$. The predictive distribution is

$$P(\mathbf{Y}_0 | \mathbf{y}, \mathbf{Z}) = \int P(\mathbf{Y}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \boldsymbol{\psi}_0, \mathbf{Z}) P(\boldsymbol{\psi}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \mathbf{Z}) \\ P(\boldsymbol{\psi} | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \mathbf{Z}) P(\mathbf{X}_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}) P(\mathbf{X} | \mathbf{y}, \mathbf{Z}) d\mathbf{X} d\mathbf{X}_0 d\boldsymbol{\psi} d\boldsymbol{\psi}_0, \quad (20)$$

which is approximated by Monte Carlo methods. In what follows, we determine the five terms as required in (20):

i) $\mathbf{Y}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \boldsymbol{\psi}_0, \mathbf{Z}$:

$$\mathbf{Y}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \boldsymbol{\psi}_0, \mathbf{Z} \sim SUN_{p, 2n} \left(\boldsymbol{\mu}_{y_{0|}}, \alpha C_\theta^{\mathcal{O}\mathcal{O}} \Lambda^{-\frac{1}{2}} \boldsymbol{\Omega} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_x), \Sigma_{y_{0|}}, \Gamma_{y_{0|}}, \Delta_{y_{0|}} \right),$$

where

$$\boldsymbol{\mu}_{y_{0|}} = \beta_0 \mathbf{1}_n + \mathbf{X}_0 \boldsymbol{\beta}_x + [\gamma^2 + \alpha^2] \Lambda^{\dagger-\frac{1}{2}} C_\theta^{\mathcal{P}\mathcal{O}} \Lambda^{-\frac{1}{2}} \boldsymbol{\Omega}^{-1}, \\ \Sigma_{y_{0|}} = [\gamma^2 + \alpha^2] \Lambda^{\dagger-\frac{1}{2}} C_\theta^{\mathcal{P}\mathcal{P}} \Lambda^{\dagger-\frac{1}{2}} \boldsymbol{\Omega}^{-1} + \sigma^2 I_p - [\gamma^2 + \alpha^2]^2 \Lambda^{\dagger-\frac{1}{2}} C_\theta^{\mathcal{P}\mathcal{O}} \Lambda^{-\frac{1}{2}} \boldsymbol{\Omega}^{-1} \Lambda^{-\frac{1}{2}} C_\theta^{\mathcal{O}\mathcal{P}} \Lambda^{\dagger-\frac{1}{2}}, \\ \Gamma_{y_{0|}} = C_\theta^{\mathcal{O}\mathcal{O}} - \alpha^2 C_\theta^{\mathcal{O}\mathcal{O}} \Lambda^{-\frac{1}{2}} \boldsymbol{\Omega}^{-1} \Lambda^{-\frac{1}{2}} C_\theta^{\mathcal{O}\mathcal{O}}, \\ \Delta_{y_{0|}} = \Lambda^{\dagger-\frac{1}{2}} C_\theta^{\mathcal{P}\mathcal{O}} - [\gamma^2 + \alpha^2] \Lambda^{\dagger-\frac{1}{2}} C_\theta^{\mathcal{P}\mathcal{O}} \Lambda^{-\frac{1}{2}} \boldsymbol{\Omega}^{-1} \Lambda^{-\frac{1}{2}} C_\theta^{\mathcal{O}\mathcal{O}}.$$

ii) $\boldsymbol{\psi}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \mathbf{Z}$:

$$P(\boldsymbol{\psi}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \mathbf{Z}) = P(\boldsymbol{\psi}_0 | \boldsymbol{\psi}) \\ = N_p \left(-\frac{\nu}{2} \mathbf{1}_p + C_\theta^{\mathcal{P}\mathcal{O}} C_\theta^{\mathcal{O}\mathcal{O}^{-1}} \left[\boldsymbol{\psi} + \frac{\nu}{2} \mathbf{1}_n \right], \nu \left[C_\theta^{\mathcal{P}\mathcal{P}} - C_\theta^{\mathcal{P}\mathcal{O}} C_\theta^{\mathcal{O}\mathcal{O}^{-1}} C_\theta^{\mathcal{O}\mathcal{P}} \right] \right).$$

iii) The third term is $P(\boldsymbol{\psi} | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \mathbf{Z}) = P(\boldsymbol{\psi} | \mathbf{y})$ and so, as mentioned before, it can be sampled using MCMC algorithms.

iv) $\mathbf{X}_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}$:

$$\begin{aligned}
 P(\mathbf{X}_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}) &\propto P(\mathbf{X}_0, \mathbf{X} | \mathbf{y}, \mathbf{Z}) \\
 &\propto P(\mathbf{Y} | \mathbf{X}_0, \mathbf{X}, \mathbf{Z}) P(\mathbf{X}_0, \mathbf{X} | \mathbf{Z}) \\
 &\propto P(\mathbf{Y} | \mathbf{X}_0, \mathbf{X}) P(\mathbf{X}_0, \mathbf{X} | \mathbf{Z}) \\
 &\propto \left[\int P(\mathbf{Y} | \mathbf{X}_0, \mathbf{X}, \boldsymbol{\psi}) P(\boldsymbol{\psi}) d\boldsymbol{\psi} \right] P(\mathbf{X}_0, \mathbf{X} | \mathbf{Z}),
 \end{aligned}$$

where $\mathbf{Y} | \mathbf{X}_0, \mathbf{X}, \boldsymbol{\psi}$ has a SUN distribution as in (7) and $\mathbf{X}_0, \mathbf{X} | \mathbf{Z}$ has a normal distribution as in (13). So, a Metropolis-Hastings step will be used to obtain $P(\mathbf{X}_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z})$.

v) The last term in (20) can be defined as

$$\begin{aligned}
 P(\mathbf{X} | \mathbf{y}, \mathbf{Z}) &\propto P(\mathbf{Y} | \mathbf{X}, \mathbf{Z}) P(\mathbf{X} | \mathbf{Z}) \\
 &\propto P(\mathbf{Y} | \mathbf{X}) P(\mathbf{X} | \mathbf{Z}),
 \end{aligned} \tag{21}$$

where $P(\mathbf{X} | \mathbf{Z})$ is as in (13). Unfortunately, the distribution $P(\mathbf{Y} | \mathbf{X})$ does not have a closed form. Therefore, (21) does not have a closed form and can be sampled by MCMC algorithms or SIR method. Now, sampling from the predictive distribution (20) is straightforward; for each posterior drawing of (21), we generate a drawing from $\mathbf{X}_0 | \mathbf{y}, \mathbf{X}, \mathbf{Z}$ then $\boldsymbol{\psi} | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \mathbf{Z}$ and $\boldsymbol{\psi}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \mathbf{Z}$ step by step. Finally, using sampling from density $\mathbf{Y}_0 | \mathbf{y}, \mathbf{X}, \mathbf{X}_0, \boldsymbol{\psi}, \boldsymbol{\psi}_0, \mathbf{Z}$, we can obtain a realization from the predictive distribution. So, we generate L draws (after a burn-in period) from $\mathbf{Y}_0 | \cdot$ and then the spatial predictor and prediction variance are given by

$$\widehat{\mathbf{Y}}_0 = L^{-1} \sum_{l=1}^L \mathbf{y}_{0_l}, \quad \widehat{Var}(\widehat{\mathbf{Y}}_0) = L^{-1} \sum_{l=1}^L (\mathbf{y}_{0_l} - \widehat{\mathbf{Y}}_0)^2,$$

respectively.

5 | SIMULATION STUDY

We now evaluate performance of the proposed approach by conducting some simulation studies. In particular, our goals are: (i) to examine the effect of sample size (the number of spatial locations) on model performance, (ii) to evaluate how the inference (or, more precisely, prediction) behaves with changing σ_u^2 , (iii) to assess the influence of misspecified ME by using different distribution of ME for the model fitting, and (iv) to compare the performance of our model with its competitors. All computations were performed using freely available **R software**. To achieve our first goal, we did three simulations each with $\mathcal{R} = 1000$ generated data sets which contain $n = 225, 325$ and 425 observations, respectively, where the sites uniformly distributed over the region $[0, 100] \times [0, 100]$. Moreover, in these simulations, subsets of size $200, 300$ and 400 , respectively, were randomly selected and used for model training, and the remaining 25 samples (in each scenario) were used for prediction. The data were simulated from the model in (10) with $k = 1$ and the following are presumed parameters: $\beta_0 = 2, \beta_1 = 3.5, \sigma^2 = 0.7, \gamma = 1.75, \alpha = -2.5, \nu = 0.85, \mu_x = 0, \sigma_x^2 = 1, \sigma_u^2 = 0.6$ and $\theta = 0.03$. Thus, for instance, for a small and a large distance, say $\|h_1\| = 3$ and $\|h_2\| = 120$, respectively, we have $C(\|h_1\|) \approx 0.91$ and $C(\|h_2\|) \approx 0.02$ which show the approximations of the maximum and the minimum dependencies based on the presented exponential correlation function. Thus, we considered a simple linear regression as $\beta_0 + \beta_1 X_i^{(r)}$, where $X_i^{(r)} \sim N(0, 1)$ is not directly observed but we observe instead $Z_i^{(r)} = X_i^{(r)} + U_i^{(r)}$, where $U_i^{(r)} \sim N(0, \sigma_u^2 = 0.6), r = 1, \dots, \mathcal{R}$. Our observed data set in each simulation run are $\left\{ \left(Y_i^{(r)}, Z_i^{(r)} \right); i = 1, \dots, n; r = 1, \dots, \mathcal{R} \right\}$. We use the bias criterion as $Bias(\hat{\eta}) = \mathcal{R}^{-1} \sum_{r=1}^{\mathcal{R}} (\hat{\eta}^{(r)} - \eta)$ for a parameter η to assess performance of $\hat{\eta}$ of the proposed methodology. In addition, the accuracy of the estimators is quantified by model-based variance of the estimators through the equation (18) where $L = 100$ and also comparing with the empirical variance of each estimation as $MSE(\hat{\eta}) = \mathcal{R}^{-1} \sum_{r=1}^{\mathcal{R}} (\hat{\eta}^{(r)} - \bar{\eta})^2$ where $\bar{\eta} = \mathcal{R}^{-1} \sum_{r=1}^{\mathcal{R}} \hat{\eta}^{(r)}$. We also compare performance of the proposed approach with the naive approach which ignores the measurement error in covariate ($\sigma_u^2 = 0$) and with the stationary Gaussian ME model. Table 1, which summarizes the results, confirms better performance of the proposed model compared to the naive approach and the Gaussian ME model. This table also reports the precision of model prediction by taking the average of the

square root of the mean squared prediction errors (ARPE), where the mean squared prediction error (MSPE) of the predicted response is $MSP E_{y_i} = \mathcal{R}^{-1} \sum_{r=1}^{\mathcal{R}} \left(\hat{y}_i^{(r)} - y_i \right)^2$ for $i = 1, \dots, 25$, and also the variance of square root of the mean squared prediction errors (VRPE). As might be expected, the most discernible differences between the model-based and empirical variances occur for the naive approach and also we can see larger variances based on the naive approach. Moreover, in this table, the CPU times, in hours, were recorded for a single run of the algorithm on an iMac with Intel Core i7 2.93GHz processor and 8GB RAM. It is obvious that the CPU times for the proposed model intensively increase as the sample size increases (see Conclusion section for more discussion on this). However, in the era of high-performance computing, the accuracy of the result is more important than the computing time.

To explore the effect of σ_u^2 on prediction (as one of the main goals of spatial studies), we again carried out the prediction based on our model but this time we fixed $n = 325$ and instead considered different values for σ_u^2 as $\sigma_u^2 = 0.4$ and 0.9 (since it is expected that $\sigma_u^2 < \sigma_x^2 = 1$). Results are presented in Table 2 where the values based on $\sigma_u^2 = 0.6$ are taken from Table 1. In a nutshell, although the biases behave differently for some model parameters for different values of σ_u^2 , but the corresponding model-based variances are very similar.

[Tables 1 and 2 about here]

To examine the sensitivity of our model when we have misspecification in the distribution of the covariate ME, we assumed two different distributions for U_i in the simulation (based on $n = 225$): t -distribution with 2 degrees of freedom i.e., $t(2)$, and the skew normal distribution $SN(0, 1; 2)$ with zero mean, variance one, and skewness parameter 2 (with density $2\phi(x)\Phi(2x)$, where $\phi(\cdot)$ and $\Phi(\cdot)$ respectively denote the standard normal density and cumulative function); noting that we generated the data sets from the above scenarios but used normal distribution for U_i in model fitting. Table 3 shows that under the misspecified ME distribution, the proposed model still provides relatively satisfactory estimators, even for a not so large sample size, however, with larger biases and variances. The last goal of this section is to compare our model performance with its competitors such as Gaussian, skew Gaussian and the GLG models through Akaike information criterion (AIC) which is a popular criterion for the model assessment. The results are summarized in Table 4. We also provide the ARPE for each model. Based on

the ARPE and AIC, the unified skew GLG model has the best performance which has the lowest ARPE and AIC among the models.

[Tables 3 and 4 about here]

Finally, our interest is to evaluate the discrepancy between the predictions of new locations and the actual values of those locations. To that end, we considered another simulation, based on $n = 400$, where a subset of size 300 uniformly selected from the region $[0, 100] \times [0, 100]$ and the remaining 100 locations will be considered on a regular 10×10 lattice with ten units between nearest neighbours as $\{5, 15, 25, 35, 45, 55, 65, 75, 85, 95\} \times \{5, 15, 25, 35, 45, 55, 65, 75, 85, 95\}$, to compare the discrepancy between the true spatial surface and the predicted spatial surface. Panel (a) of Figure 1 shows that our model prediction works very well, however, to explore more precisely, we provided a variogram for the difference between the true and predicted spatial surface (panel (b)). As we expected, the variogram is spatially unstructured and close to zero.

[Figure 1 about here]

6 | PM DATA APPLICATION

Environmental air pollution encompasses various particulate matters. The increased ambient particulate matter (PM) from industrialization and urbanization is highly associated with morbidity and mortality worldwide, presenting one of the most severe environmental pollution problems. PMs include the harmful suspended mixture of both solid and liquid particles. They are often separated into three classifications: “coarse”, “fine” and “ultrafine” particles. Coarse particles have a diameter of between 10 micrometer (μm) and $2.5\mu m$ and settle relatively quickly whereas fine (0.1 to $2.5\mu m$ in diameter) and ultrafine ($< 0.1\mu m$ in diameter) particles remain in suspension for longer period of time. To put things into perspective, human hair has a diameter of $50 - 70\mu m$ and a grain of sand has a diameter of $90\mu m$. Since an excess of PM_{10} concentration in any area causes serious environmental pollution, the detection of areas with high PM_{10} concentration is an important problem and this is a greatest concern of

recent studies that link PM_{10} exposure to the premature death of people who already have heart and lung disease, especially the elderly (Titi, Dweirj, & Tarawneh 2015).

This section consists of an illustrative application of our methodology to investigate the spatial variations of PM_{10} concentration in Iran. The study area contains the 72 most populous cities in Iran with the annual maximum daily PM_{10} concentration. The data set collected from a monitoring network composed of 72 stations in 2011 and can be downloaded from the [Iran Meteorological Organization data centre](#) website. The distance of the nearest and the farthest data point are 14.1km and 1624.3km, respectively. The data contains the annual maximum PM_{10} concentration (in $\mu g/m^3$), which is used as the response variable in our study, spatial coordinates in terms of latitude and longitude, and mean wind speed (WS, in km/h) as a covariate in analyzing PM_{10} concentration (see, e.g., Cameletti, Ignaccolo, & Bande 2011). The mean wind speed has been obtained by averaging the corresponding values while the PM_{10} concentrations was in the highest level at each location. Moreover, a review of the literature confirms that measuring wind speed is associated with ME. Namely, a professional and well calibrated rotation anemometers such as cup anemometers and rotation vane anemometers which are apparently the most usual instruments for measuring the wind speed, has at least a ME around 1% ([Danish Wind Turbine Manufacturers Association](#)). Knowledge of the ME variance was empirically determined for wind speed by the Iran Meteorological Organization data centre as $\sigma_u^2 \approx 0.7$. Figure 2 shows a schematic description of the region and the location of monitoring stations as well as the histogram of the maximum PM_{10} concentration. As a result of simple exploratory data analysis, the histogram induces a non-Gaussian feature but to explore more precisely this feature, further investigation is needed. Therefore, a simple mean function ($\beta_0 + \beta_1 WS$) was fitted and several goals were pursued: the normal Q-Q plots of the response and residuals, the kernel estimate of residuals density, and the Shapiro-Wilk test for normality assumption of the response and residuals. The p -values obtained from the Shapiro-Wilk normality tests for the response and residuals were 0.0091 and 3.66×10^{-6} , respectively, which indicate the non-normality of response and residuals. Figure 3 provides the Q-Q plots of response and residuals, and histogram of residuals which all clearly show deviation from the normality assumption.

[Figures 2, 3 and 4 about here]

We also use the H-scatter plot as a tool for outlier detection and spatial continuity. In particular, an H-scatter plot shows all possible pairs of data values whose locations are separated by a certain distance h in a particular direction. When the spatial correlations are strong with closer distance of the sampling sites or the data is highly homogenous, the monitoring sites tend to assemble to a straight line with the angle of 45° (which shows the stationarity of the process as well). If the spatial correlation between two samples decreases or the relationship between two variables weakens, the shape of the cloud of points will spread out displaying a characteristic butterfly-wing shape. The H-scatter plot of the maximum PM_{10} concentration is shown in Figure 4 (left panel). The points are located far from the cross-line for small distance $\|h\|$ with a strong variability which can be treated as outliers in the data set. Moreover, structural geostatistical description of the response variability was revealed by the omnidirectional experimental semi-variogram map as shown in Figure 4 (right panel). In this map, a class of distance and a direction, which can be converted into a grid cell representing the vertex of the vector whose origin is at the center of the grid and whose norm equals the distance between the two points and direction equals the direction along which the two points, are aligned. This indicates that we need to analyze the spatial continuity of the maximum PM_{10} in all the directions of space. The isotropic feature is clearly identified by this figure as well.

We also need the value of σ^2 which is often obtained from other sources. However, since we do not have access to this information, our practical advice is to analyze the data based on some different values of σ^2 to evaluate different scenarios by AIC (which is the most popular criterion for model assessment in the literature) and come up with the final/good model which fits well for the dataset. Table 5 indicates that the proposed model presents a better fit based on $\sigma^2 = 0.65$ than the other values. Table 6 displays the model parameters estimate and corresponding standard error for the proposed and naive (ignoring measurement error) approaches, and also based on the ME models GLG, skew Gaussian, and Gaussian. As might be expected from the simulation, the proposed model presents a better fit for the data compared to the naive model based on the AIC model comparison.

[Table 5 and 6 about here]

Finally, the contour map corresponding to the predictive mean under the unified skew GLG case is shown in Figure 5.

7 | CONCLUSION

In this paper, we have developed a modeling approach to account for the covariates measurement error in non-Gaussian spatial data to also capture both skewness and heavy tails in the response data. Moreover, we equated heavy tails with outliers/large values and explained that the large values may be due to the large variance. Although, this could be a property of a non-stationary Gaussian process, and not necessarily indicates non-Gaussian distribution, the exploratory data analysis of the maximum PM_{10} concentration implied that the response was stationary (since the monitoring sites tend to assemble to a straight line with the angle of 45° in the H-scatterplot). An expectation-maximization algorithm was developed to compute the maximum likelihood estimate of model parameters. A simulation study was conducted to show outperformance of the proposed model compared to the naive model which ignores the measurement error in covariates. A real data application regarding the maximum PM_{10} concentration with wind speed as a covariate measured with error was also analyzed.

It is clear that without some kind of information regarding the magnitude of measurement error, measurement error models will not be identifiable. Practically speaking, there are two possibilities: (i) assuming that measurement error variance is known or can be estimated using other sources (such as a validation data set, see [Carroll, Ruppert, Stefanski, and Crainiceanu 2006](#)); (ii) assumptions are made regarding measurement error process. However, [Huque et al. \(2016\)](#) presented an approach which contrasts with the assumption of existence of validation data. Our paper has used the first approach. One of the additional assumptions required by our approach is that the correlation structure $\mathbf{W}(\cdot)$ and $\boldsymbol{\psi}(\cdot)$ are equal but one can choose different spatial correlation structures and address the identifiability issue. In addition, the focus of this work was structural measurement error. One can also extend the proposed model in this study to the functional measurement error. Moreover, in the last decade, there has been a growing interest in terms of modeling spatial big data. A core difficulty of analyzing spatial big data is in inverting the $n \times n$ covariance matrix (see [Heaton et al. 2017](#) as an introductory overview of the spatial big n problems and several proposed methods). Another natural extension of this work is to scale the proposed model to

big data. Another direction of this work is to extend the proposed spatial model to the spatio-temporal model. We have planned to study these approaches in our future studies.

SUPPLEMENTARY MATERIALS

The supplementary materials contain R codes and corresponding "readme" files for the simulation study and real data application conducted in this paper.

ACKNOWLEDGEMENTS

We would like to thank the Associate Editor and two referees for constructive comments and suggestions, which led to an improved version of the manuscript. The research of M. Torabi was supported by the Natural Sciences and Engineering Research Council of Canada. Partial financial support of V. Tadayon's research by the Ministry of Science, Research, and Technology of Iran is gratefully appreciated. This research was done when V. Tadayon visited M.Torabi as a visiting Ph.D. student..

APPENDIX

A: Log-Likelihood Function

We provide details of equation (11). To that end, we can write

$$\begin{aligned}
 \ell(\eta) &= \log f_{\mathbf{Y}, \mathbf{Z} | \mathbf{X}, T_0, T_1, \lambda}(\mathbf{y}, \mathbf{Z}; \boldsymbol{\beta}) + \log f_{\mathbf{X}}(\mathbf{X}; \boldsymbol{\mu}_x, \sigma_x^2) \\
 &\quad + \log f_{T_0}(\mathbf{t}_0; \gamma^2, \theta) + \log f_{T_1}(\mathbf{t}_1; \alpha^2, \theta) + \log f_{\lambda}(\lambda; \nu, \theta) \\
 &= -\frac{1}{2} \|\mathbf{y} - \beta_0 \mathbf{1}_n - \mathbf{X}\boldsymbol{\beta}_x - \mathbf{W}_{\lambda}\|^2 - \frac{n}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} \|\mathbf{Z} - \mathbf{X}\|^2 \\
 &\quad - \frac{n}{2} \log \sigma_x^2 - \frac{1}{2\sigma_x^2} \|\mathbf{X} - \boldsymbol{\mu}_x\|^2 \\
 &\quad - \frac{n}{2} \log \gamma^2 - \frac{1}{2} \log |C_{\theta}| - \frac{1}{2\gamma^2} \mathbf{t}'_0 C_{\theta}^{-1} \mathbf{t}_0
 \end{aligned}$$

$$\begin{aligned}
& -\frac{n}{2} \log \alpha^2 - \frac{1}{2\alpha^2} \mathbf{t}_1' C_\theta^{-1} \mathbf{t}_1 - \log \int_{\mathfrak{R}_+^n} \exp \left\{ -\frac{1}{2} \mathbf{t}_1' C_\theta^{-1} \mathbf{t}_1 \right\} d\mathbf{t}_1 \\
& -\frac{n}{2} \log \nu - \frac{1}{2} \log |C_\theta| - \frac{1}{2\nu} \left(\boldsymbol{\psi} + \frac{\nu}{2} \mathbf{1}_n \right)' C_\theta^{-1} \left(\boldsymbol{\psi} + \frac{\nu}{2} \mathbf{1}_n \right),
\end{aligned}$$

where $\|\cdot\|$ denotes the square norm, $|\cdot|$ denotes the determinant and $\mathbf{W}_\lambda = \Lambda^{-\frac{1}{2}} (T_0 + T_1)$. Further, equation (12)

can be written as

$$\begin{aligned}
Q(n|\eta') &= E \{ \ell(\eta) | \mathbf{y}, \mathbf{Z}, \eta' \} \\
&= -\frac{1}{2} [\mathbf{y}'\mathbf{y} - 2\beta_0 \mathbf{y}'\mathbf{1}_n - 2\mathbf{y}' E_1(\mathbf{X}|\mathbf{Z}) \boldsymbol{\beta}_x - 2\mathbf{y}' E_2(\mathbf{W}_\lambda | \mathbf{y}) \\
&\quad + \beta_0^2 \mathbf{1}_n' \mathbf{1}_n + 2\beta_0 \mathbf{1}_n' E_1(\mathbf{X}|\mathbf{Z}) \boldsymbol{\beta}_x + 2\beta_0 \mathbf{1}_n' E_2(\mathbf{W}_\lambda | \mathbf{y}) \\
&\quad + \boldsymbol{\beta}_x' E_3(\mathbf{X}'\mathbf{X}|\mathbf{Z}) \boldsymbol{\beta}_x + 2 [E_2(\mathbf{W}_\lambda | \mathbf{y})]' E_1(\mathbf{X}|\mathbf{Z}) \boldsymbol{\beta}_x + E_4(\mathbf{W}_\lambda' \mathbf{W}_\lambda | \mathbf{y})] \\
&\quad - \frac{n}{2} \log \sigma_u^2 - \frac{1}{2\sigma_u^2} \{ \mathbf{Z}'\mathbf{Z} - 2\mathbf{Z}'\mathbf{X} + \mathbf{X}'\mathbf{X} \} \\
&\quad - \frac{n}{2} \log \sigma_x^2 - \frac{1}{2\sigma_x^2} \{ E_3(\mathbf{X}'\mathbf{X}|\mathbf{Z}) - 2 [E_1(\mathbf{X}|\mathbf{Z})]' \boldsymbol{\mu}_x + \boldsymbol{\mu}_x' \boldsymbol{\mu}_x \} \\
&\quad - \frac{n}{2} \log \gamma^2 - \frac{1}{2} \log |C_\theta| - \frac{1}{2\gamma^2} \text{tr} [C_\theta^{-1} E_5(T_0 T_0' | \mathbf{y})] \\
&\quad - \frac{n}{2} \log \alpha^2 - \frac{1}{2\alpha^2} \text{tr} \{ C_\theta^{-1} E_6(T_1 T_1' | \mathbf{y}) \} - \log \int_{\mathfrak{R}_+^n} \exp \left\{ -\frac{1}{2} \mathbf{t}_1' C_\theta^{-1} \mathbf{t}_1 \right\} d\mathbf{t}_1 \\
&\quad - \frac{n}{2} \log \nu - \frac{1}{2} \log |C_\theta| - \frac{1}{2\nu} \left[\text{tr} \{ C_\theta^{-1} E_7(\boldsymbol{\psi} \boldsymbol{\psi}' | \mathbf{y}) \} + \nu \mathbf{1}_n' C_\theta^{-1} E_8(\boldsymbol{\psi} | \mathbf{y}) + \frac{\nu^2}{4} \mathbf{1}_n' C_\theta^{-1} \mathbf{1}_n \right], \tag{22}
\end{aligned}$$

in which the last three lines of (22) are derived from equation (13), the role of the expected value of a quadratic form, and also

$$E_3(\mathbf{X}'\mathbf{X}|\mathbf{Z}) = \sigma_x^2 [1 - a] I_k + E_1(\mathbf{X}'|\mathbf{Z}) E_1(\mathbf{X}|\mathbf{Z}).$$

B: Updates of the Parameters in the MCEM Algorithm

To update the MCEM estimates at the stage $t + 1$ of the algorithm, we have the following equations:

$$\boldsymbol{\beta}_x^{t+1} = [\mathbb{E}_3^{t+1}]^{-1} [\mathbb{E}_1^{t+1}]' (\mathbf{y} - \beta_0^t \mathbf{1}_n - \mathbb{E}_2^{t+1}),$$

$$\begin{aligned}
\beta_0^{t+1} &= \frac{1}{n} \mathbf{1}'_n (\mathbf{y} - \mathbb{E}_1^{t+1} \beta'_x - \mathbb{E}_2^{t+1}), \\
\sigma_x^{2^{t+1}} &= \frac{1}{n} \left[\mathbb{E}_3^{t+1} - 2 [\mathbb{E}_1^{t+1}]' \mu'_x + \mu'^t_x \mu'_x \right], \\
\gamma^{2^{t+1}} &= \frac{1}{n} \text{tr} (C_{\theta^t}^{-1} \mathbb{E}_5^{t+1}), \\
\alpha^{2^{t+1}} &= \frac{1}{n} \text{tr} (C_{\theta^t}^{-1} \mathbb{E}_6^{t+1}), \\
\nu^{t+1} &= 2 [\mathbf{1}'_n C_{\theta^t}^{-1} \mathbf{1}_n]^{-1} \left[-n + \sqrt{n^2 + \text{tr} [C_{\theta^t}^{-1} \mathbb{E}_7^{t+1}] \mathbf{1}'_n C_{\theta^t}^{-1} \mathbf{1}_n} \right], \\
\mu_x^{t+1'} &= \sigma_x^{2^t} \left(1 - \frac{\sigma_x^{2^t}}{\sigma_x^{2^t} + \sigma_u^2} \right) \left[\mathbf{y}' - \beta_0^t \mathbf{1}'_n - \left(1 - \frac{\sigma_x^{2^t}}{\sigma_x^{2^t} + \sigma_u^2} \right) [\mathbb{E}_2^{t+1}]' \right] \mathbf{1}'_k \beta_x^t + \mathbb{E}_1^{t+1} \mathbf{1}_k, \\
\theta^{t+1} &= \arg \min_{\theta} \left\{ \log |C_{\theta^t}| + \frac{1}{2\gamma^{2^t}} \text{tr} [C_{\theta^t}^{-1} \mathbb{E}_5^{t+1}] \right. \\
&\quad \left. + \frac{1}{2\alpha^{2^t}} \text{tr} [C_{\theta^t}^{-1} \mathbb{E}_6^{t+1}] + \log \int_{\mathfrak{R}_+^n} \exp \left\{ -\frac{1}{2} \mathbf{t}'_1 C_{\theta^t}^{-1} \mathbf{t}_1 \right\} d\mathbf{t}_1 \right. \\
&\quad \left. + \frac{1}{2\nu^{t+1}} \text{tr} [C_{\theta^t}^{-1} \mathbb{E}_7^{t+1}] + \frac{1}{2} \mathbf{1}'_n C_{\theta^t}^{-1} \mathbb{E}_8^{t+1} + \frac{\nu^t}{8} \mathbf{1}'_n C_{\theta^t}^{-1} \mathbf{1}_n \right\}.
\end{aligned}$$

C: Sampling from Conditional Distribution $\mathbf{X} \mid \mathbf{y}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1, \lambda$

Choosing a Metropolis-Hastings step to draw samples from (14) consists of accepting the produced value \mathbf{X}^* from the candidate generator $q(\mathbf{X}^*)$ at the k^{th} iteration with probability $\min \{1, r_k\}$ where $r_k = \frac{f(\mathbf{X}^* | \text{data}) q(\mathbf{X}^{(k)})}{f(\mathbf{X}^{(k)} | \text{data}) q(\mathbf{X}^*)}$ and $f(\cdot | \text{data})$ is defined by (14) and the candidate generator q can be chosen as a n-variate normal distribution where its parameters can be simply chosen as the sample mean and sample variance of $\mathbf{X}^{(k)}$. Similarly, in the SIR approach, we may generate (say, L) approximate samples from (14) as follows:

- Draw samples $\{\mathbf{X}^{(i)}\}_{i=1}^L$ from the proposal distribution $q(\mathbf{X})$,
- Calculate importance weights $\omega_i = f(\mathbf{X}^{(i)} | \text{data}) / q(\mathbf{X}^{(i)})$,
- Normalize the weights as $p_i = \omega_i / \sum_i \omega_i$,
- Resample with replacement from $\{\mathbf{X}^{(i)}\}_{i=1}^L$ with sample probabilities p_i ,

where the proposal distribution $q(\mathbf{X})$ can be chosen same as the first approach.

D: Sampling from Conditional Distribution $\psi \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1$

Since (15) does not have a standard probability distribution, we use the Metropolis-Hastings algorithm to generate a sample from it. To that end, we approximate (16) by log-normal distribution. If $\mathfrak{S}_i = \lambda_i^{-\frac{1}{2}}$, $\mathbf{t}_i = \mathbf{t}_0 + \mathbf{t}_1$ and $h_i = \sigma^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)$ where $\text{sign}(\cdot)$ denotes the sign function, then

$$\begin{aligned}
 E(\mathfrak{S}_i) &= \frac{\int_0^\infty \mathfrak{S}_i \exp \left\{ -\frac{1}{2} \left[\frac{\mathfrak{S}_i - |\mathbf{t}_i|^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)}{\sigma |\mathbf{t}_i|^{-1}} \right]^2 \right\} d\mathfrak{S}_i}{\int_0^\infty \exp \left\{ -\frac{1}{2} \left[\frac{\mathfrak{S}_i - |\mathbf{t}_i|^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)}{\sigma |\mathbf{t}_i|^{-1}} \right]^2 \right\} d\mathfrak{S}_i} \\
 &= \sigma |\mathbf{t}_i|^{-1} \left[\frac{\int_0^\infty \frac{\mathfrak{S}_i - |\mathbf{t}_i|^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)}{\sigma |\mathbf{t}_i|^{-1}} \exp \left\{ -\frac{1}{2} \left[\frac{\mathfrak{S}_i - |\mathbf{t}_i|^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)}{\sigma |\mathbf{t}_i|^{-1}} \right]^2 \right\} d\mathfrak{S}_i}{\int_0^\infty \exp \left\{ -\frac{1}{2} \left[\frac{\mathfrak{S}_i - |\mathbf{t}_i|^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)}{\sigma |\mathbf{t}_i|^{-1}} \right]^2 \right\} d\mathfrak{S}_i} + h_i \right] \\
 &= \sigma |\mathbf{t}_i|^{-1} \left[\frac{\int_{-h_i}^\infty \mathfrak{S}_i^* \exp \left\{ -\frac{1}{2} \mathfrak{S}_i^{*2} \right\} d\mathfrak{S}_i^*}{\int_{-h_i}^\infty \exp \left\{ -\frac{1}{2} \mathfrak{S}_i^{*2} \right\} d\mathfrak{S}_i^*} + h_i \right] \\
 &= \sigma |\mathbf{t}_i|^{-1} [\delta(h_i) + h_i],
 \end{aligned}$$

where

$$\mathfrak{S}_i^* = \frac{\mathfrak{S}_i - |\mathbf{t}_i|^{-1} \mathbf{b}_4 \text{sign}(\mathbf{t}_i)}{\sigma |\mathbf{t}_i|^{-1}}, \quad \delta(h_i) = \frac{\int_{-h_i}^\infty \mathfrak{S}_i^* \exp \left\{ -\frac{1}{2} \mathfrak{S}_i^{*2} \right\} d\mathfrak{S}_i^*}{\int_{-h_i}^\infty \exp \left\{ -\frac{1}{2} \mathfrak{S}_i^{*2} \right\} d\mathfrak{S}_i^*}.$$

In a similar way, one can easily see that $E(\mathfrak{S}_i^2) = \sigma^2 \mathbf{t}_i^{-2} [1 + h_i \delta(h_i) + h_i^2]$. On the other hand, we know that if

$\lambda_i \sim LN(m_i, s_i^2)$, then

$$E\left(\lambda_i^{-\frac{1}{2}}\right) = e^{-\frac{1}{2}m_i + \frac{1}{8}s_i^2}, \quad E(\lambda_i^{-1}) = e^{-m_i + \frac{1}{2}s_i^2}.$$

Therefore, by solving the system of following linear equations:

$$e^{-\frac{1}{2}m_i + \frac{1}{8}s_i^2} = \sigma |\mathbf{t}_i|^{-1} [\delta(h_i) + h_i]$$

$$e^{-m_i + \frac{1}{2}s_i^2} = \sigma^2 \mathbf{t}_i^{-2} [1 + h_i \delta(h_i) + h_i^2]$$

in terms of m_i and s_i^2 , we obtain, as an approximating distribution of the likelihood contribution to ψ_i as

$$N(m_i, s_i^2), \quad m_i = \ln \frac{t_i^2 [1 + h_i \delta(h_i) + h_i^2]}{\sigma^2 [\delta(h_i) + h_i]^4}, \quad s_i^2 = 4 \ln \frac{1 + h_i \delta(h_i) + h_i^2}{[\delta(h_i) + h_i]^2}. \quad (23)$$

A suitable candidate generator can be constructed by combining (17) and (23) as

$$\begin{aligned} q(\psi_i | \cdot) &\propto \exp \left\{ -\frac{1}{2} \sum_{j=1}^n \frac{(\psi_j - m_j)^2}{s_j^2} \right\} \exp \left\{ -\frac{1}{2} \frac{(\psi_i - m_{i*})^2}{s_{i*}^2} \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left[\frac{\psi_i^2 - 2m_i \psi_i + m_i^2}{s_i^2} + \frac{\psi_i^2 - 2m_{i*} \psi_i + m_{i*}^2}{s_{i*}^2} \right] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} \left(\frac{s_i^2 + s_{i*}^2}{s_i^2 s_{i*}^2} \right) \left[\psi_i^2 - 2\psi_i \left(\frac{s_i^2 m_{i*} + s_{i*}^2 m_i}{s_i^2 + s_{i*}^2} \right) \right] \right\}, \end{aligned}$$

i.e.,

$$q(\psi_i | \mathbf{y}, \mathbf{X}, \mathbf{Z}, \mathbf{t}_0, \mathbf{t}_1, \boldsymbol{\psi}_{-i}; \eta^t) \stackrel{d}{=} N \left(\frac{s_i^2 m_{i*} + s_{i*}^2 m_i}{s_i^2 + s_{i*}^2}, \frac{s_i^2 s_{i*}^2}{s_i^2 + s_{i*}^2} \right),$$

where m_{i*} and s_{i*}^2 represent the mean and variance of (17).

References

- Alexeeff, S. E., Carroll, R. J., & Coull, B. (2016). Spatial measurement error and correction by spatial SIMEX in linear regression models when using predicted air pollution exposures. *Biostatistics*, 17(2), 377–389.
- Allard, D., & Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics & Theory and Methods*, 36(9), 1821–1834.
- Arellano-Valle, R. ., & Azzalini, A. (2006). On the unification of families of skew-normal distributions. *Scandinavian Journal of Statistics*, 33(3), 561–574.
- Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, 171–178.
- Azzalini, A. (2013). *The skew-normal and related families* (Vol. 3). Cambridge University Press.

- Azzalini, A., & Capitanio, A. (1999). Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Association*, *61*, 579–602.
- Azzalini, A., & Dalla-Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika*, *83*, 715–726.
- Bueno, R. S., Fonseca, T. C., & Schmidt, A. M. (2017). Accounting for covariate information in the scale component of spatio-temporal mixing models. *Spatial Statistics*, *22*, 196–218.
- Buzas, J. S., Stefanski, L. A., & Tosteson, T. D. (2014). Measurement error. *Handbook of epidemiology*, 1241–1282.
- Cameletti, M., Ignaccolo, R., & Bande, S. (2011). Comparing spatio-temporal models for particulate matter in Piemonte. *Environmetrics*, *22*(8), 985–996.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., & Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: A modern perspective*. Chapman and Hall/CRC.
- Cressie, N. (1993). *Statistics for spatial data*. New York: Wiley.
- De-Oliveira, V., Kedem, B., & Short, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association*, *92*(440), 1422–1433.
- Fonseca, T. C. O., & Steel, M. F. J. (2011). Non-Gaussian spatiotemporal modelling through scale mixing. *Biometrika*, *98*(4), 761–774.
- Genton, M. G., & Zhang, H. (2012). Identifiability problems in some non-Gaussian spatial random fields. *Chilean Journal of Statistics*, *3*(2), 171–179.
- Gryparis, A., Coull, B. A., & Schwartz, J. (2007). Controlling for confounding in the presence of measurement error in hierarchical models: A Bayesian approach. *Journal of Exposure Science and Environmental Epidemiology*, *17*, S20–S28.

- Gupta, A. K., Aziz, M. A., & Ning, W. (2013). On some properties of the unified skew normal distribution. *Journal of Statistical Theory and Practice*, 7(3), 480–495.
- Heaton, M., Datta, A., Finley, A., Furrer, R., Guhaniyogi, R., Gerber, F., ... others (2017). Methods for analyzing large spatial data: A review and comparison. *arXiv preprint arXiv:1710.05013*.
- Hoque, M. E., & Torabi, M. (2018). Modeling the random effects covariance matrix for longitudinal data with covariates measurement error. *Statistics in medicine*, <https://doi.org/10.1002/sim.7908>.
- Huque, M. H., Bondell, H. D., Carroll, R. J., & Ryan, L. M. (2016). Spatial regression with covariate measurement error: A semiparametric approach. *Biometrics*, 72(3), 678–686.
- Huque, M. H., Bondell, H. D., & Ryan, L. (2014). On the impact of covariate measurement error on spatial regression modelling. *Environmetrics*, 25(8), 560–570.
- Kim, H., & Mallick, B. K. (2004). A Bayesian prediction using the skew Gaussian distribution. *Journal of Statistical Planning and Inference*, 120(1), 85–101.
- Le-Gallo, J., & Fingleton, B. (2012). Measurement errors in a spatial context. *Regional Science and Urban Economics*, 42(1), 114–125.
- Li, Y., Tang, H., & Lin, X. (2009). Spatial linear mixed models with covariate measurement errors. *Statistica Sinica*, 19(3), 1077–1093.
- Militino, A., Ugarte, M., Iribas, J., & Lizarraga-Garcia, E. (2013). Mapping GPS positional errors using spatial linear mixed models. *Journal of Geodesy*, 87(7), 675–685.
- Miller, D. M. (1984). Reducing transformation bias in curve fitting. *The American Statistician*, 38(2), 124–126.
- Neath, R. C., et al. (2013). On convergence properties of the Monte Carlo EM algorithm. In *Advances in modern statistical theory and applications: A festschrift in honor of morris L. Eaton* (pp. 43–62). Institute of Mathematical Statistics.

- Palacios, M. B., & Steel, M. F. J. (2006). Non-Gaussian Bayesian geostatistical modeling. *Journal of the American Statistical Association*, *101*(474), 604–618.
- Sheppard, L., Burnett, R. T., Szpiro, A. A., Kim, S. Y., Jerrett, M., Pope III, C. A., & Brunekreef, B. (2012). Confounding and exposure measurement error in air pollution epidemiology. *Air Quality, Atmosphere & Health*, *5*(2), 203–216.
- Steel, M. F., & Fuentes, M. (2010). Non-Gaussian and nonparametric models for continuous spatial data. *Handbook of Spatial Statistics*, 149–167.
- Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media.
- Szpiro, A. A., Sheppard, L., & Lumley, T. (2011). Efficient measurement error correction with spatially misaligned data. *Biostatistics*, *12*(4), 610–623.
- Tadayon, V. (2017). Bayesian analysis of censored spatial data based on a non-Gaussian model. *Journal of Statistical Research of Iran*, *13*(2), 155–180.
- Tadayon, V., & Khaledi, M. J. (2015). Bayesian analysis of skew Gaussian spatial models based on censored data. *Communications in Statistics-Simulation and Computation*, *44*(9), 2431–2441.
- Thomas, D. C. (2013). Measurement error in spatial exposure models: Study design implications. *Environmetrics*, *24*(8), 518–520.
- Titi, A., Dweirj, M., & Tarawneh, K. (2015). Environmental effects of the open cast mining a case study: Irbid Area, North Jordan. *American Journal of Industrial and Business Management*, *5*(06), 404–423.
- Torabi, M. (2012). Small area estimation using survey weights under a nested error linear regression model with structural measurement error. *Journal of Multivariate Analysis*, *109*, 52–60.
- Torabi, M., Datta, G. S., & Rao, J. (2009). Empirical Bayes estimation of small area means under a nested error linear regression model with measurement errors in the covariates. *Scandinavian Journal of Statistics*, *36*(2), 355–369.

Wallin, J., & Bolin, D. (2015). Geostatistical modelling using non-Gaussian Matérn fields. *Scandinavian Journal of Statistics*, 42(3), 872–890.

Zareifard, H., & Khaledi, M. J. (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., & El-Shaarawi, A. (2010). On spatial skew-Gaussian processes and applications. *Environmetrics*, 21(1), 33–47.

TABLE 1 Bias value, model-based variance and empirical variance of the estimated parameters for the proposed and naive approaches and also for the Gaussian ME model based on three different simulations with $n = 225, 325$ and 425 . ARPE: the average of the square root of the mean squared prediction errors; VRPE: the variance of square root of the mean squared prediction errors; CPU: the CPU times, in hours, for a single run of the algorithm.

Parameter	True value	Proposed Model ($\sigma_u^2 = 0.6$)				Naive Model				Gaussian Model			
		Bias		Variance		Bias		Variance		Bias		Variance	
		Model-Based	Empirical	Model-Based	Empirical	Model-Based	Empirical	Model-Based	Empirical	Model-Based	Empirical	Model-Based	Empirical
$n = 225$													
β_0	2	-0.012	0.095	0.077	1.610	1.133	1.016	2.715	2.511	2.621	2.715	2.511	2.621
β_1	3.5	0.205	0.112	0.123	1.864	1.469	1.601	3.112	2.747	2.850	3.112	2.747	2.850
μ_x	0	0.115	0.095	0.078	1.750	0.985	1.207	-2.217	1.551	1.472	-2.217	1.551	1.472
σ_x^2	1	0.293	0.402	0.373	1.973	1.891	1.625	2.340	2.012	1.879	2.340	2.012	1.879
γ	1.75	0.224	0.110	0.193	-1.414	1.981	2.233	1.697	1.953	2.011	1.697	1.953	2.011
α	-2.5	-0.313	0.227	0.211	-1.123	2.264	2.081	-	-	-	-	-	-
ν	0.85	-0.325	0.201	0.209	1.102	1.781	1.683	-	-	-	-	-	-
θ	0.03	0.254	0.262	0.295	0.809	1.203	1.475	1.231	1.330	1.441	1.231	1.330	1.441
ARPE (VRPE)		0.64 (0.122)		4.88 (2.856)		9.56 (3.733)		4.2		9.56 (3.733)		4.2	
CPU		11.5		10.1		10.1		4.2		10.1		4.2	
$n = 325$													
β_0	2	0.011	0.096	0.075	-1.413	0.901	0.925	2.101	2.042	1.988	2.101	2.042	1.988
β_1	3.5	0.189	0.104	0.101	1.868	1.470	1.443	-2.406	2.119	2.226	-2.406	2.119	2.226
μ_x	0	0.111	0.095	0.073	1.746	0.964	0.996	-1.923	1.490	1.603	-1.923	1.490	1.603
σ_x^2	1	-0.288	0.400	0.374	1.966	1.784	1.425	1.965	2.131	2.214	1.965	2.131	2.214
γ	1.75	0.213	0.098	0.180	-1.40	1.983	2.208	1.613	1.899	1.938	1.613	1.899	1.938
α	-2.5	0.270	0.230	0.209	-1.119	2.356	2.080	-	-	-	-	-	-
ν	0.85	0.312	0.196	0.198	1.087	1.780	1.684	-	-	-	-	-	-
θ	0.03	0.250	0.260	0.288	0.796	1.192	1.450	1.118	1.315	1.297	1.118	1.315	1.297
ARPE (VRPE)		0.51 (0.096)		4.72 (2.822)		8.71 (3.421)		6.3		8.71 (3.421)		6.3	
CPU		25.2		23.8		23.8		6.3		23.8		6.3	
$n = 425$													
β_0	2	-0.008	0.090	0.071	1.320	0.902	0.920	-1.823	1.847	1.925	-1.823	1.847	1.925
β_1	3.5	0.166	0.101	0.103	-1.80	1.451	1.545	-2.057	1.921	1.780	-2.057	1.921	1.780
μ_x	0	0.109	0.091	0.070	1.746	0.960	0.990	1.996	1.375	1.249	1.996	1.375	1.249
σ_x^2	1	0.286	0.395	0.370	1.960	1.778	1.403	1.721	1.930	1.956	1.721	1.930	1.956
γ	1.75	0.206	0.096	0.179	-1.397	1.861	1.976	1.431	1.870	1.944	1.431	1.870	1.944
α	-2.5	0.245	0.219	0.198	-1.107	2.155	2.067	-	-	-	-	-	-
ν	0.85	-0.309	0.196	0.191	1.068	1.772	1.680	-	-	-	-	-	-
θ	0.03	0.248	0.260	0.284	0.790	1.187	1.443	1.020	1.126	1.098	1.020	1.126	1.098
ARPE (VRPE)		0.43 (0.072)		4.41 (2.371)		8.22 (3.245)		11.7		8.22 (3.245)		11.7	
CPU		41.4		40.1		40.1		11.7		40.1		11.7	

TABLE 2 Bias value, model-based variance of the estimated parameters and the average of the square root of the mean squared prediction errors (ARPE) based on the proposed model with $n = 325$. The numbers in the parentheses denote the variances; VRPE: the variance of square root of the mean squared prediction errors.

	β_0	β_1	μ_x	σ_x^2	γ	α	ν	θ	ARPE (VRPE)
$\sigma_u^2 = 0.4$	0.01 (0.09)	0.17 (0.11)	-0.10 (0.10)	0.32 (0.50)	0.24 (0.11)	0.35 (0.20)	-0.20 (0.15)	0.25 (0.27)	0.79 (0.090)
$\sigma_u^2 = 0.6$	0.01 (0.09)	0.18 (0.10)	0.11 (0.09)	-0.28 (0.40)	0.21 (0.09)	0.27 (0.23)	0.31 (0.19)	0.25 (0.26)	0.51 (0.096)
$\sigma_u^2 = 0.9$	0.03 (0.12)	-0.14 (0.10)	0.11 (0.13)	0.53 (0.47)	-0.20 (0.11)	-0.30 (0.22)	-0.27 (0.10)	0.30 (0.29)	0.98 (0.122)

TABLE 3 Bias value, model-based variance of the estimated parameters and the average of the square root of the mean squared prediction errors (ARPE) based on the proposed model with $n = 225$ and misspecification in the distribution of the covariate MEs. The numbers in the parentheses denote the variances; $t(2)$ denotes the t -distribution with 2 degrees of freedom; SN denotes the skew normal distribution $SN(0, 1; 2)$ and N denotes the normal distribution $N(0, 0.6)$. The values of the last row are taken from Table 1. VRPE: the variance of square root of the mean squared prediction errors.

	β_0	β_1	μ_x	σ_x^2	γ	α	ν	θ	ARPE (VRPE)
$t(2)$	0.19 (0.13)	0.23 (0.20)	0.26 (0.17)	0.36 (0.53)	-0.40 (0.25)	0.42 (0.30)	0.45 (0.27)	0.33 (0.31)	1.67 (0.28)
SN	-0.40 (0.22)	-0.30 (0.25)	0.23 (0.16)	0.62 (0.51)	0.54 (0.37)	-0.55 (0.30)	0.41 (0.34)	0.47 (0.35)	2.09 (0.40)
N	-0.01 (0.09)	0.20 (0.11)	0.11 (0.09)	0.29 (0.40)	0.22 (0.11)	-0.31 (0.22)	-0.32 (0.20)	0.25 (0.26)	0.64 (0.12)

TABLE 4 Bias value and model-based variance of the estimated parameters based on the various ME models: Gaussian, skew Gaussian, GLG, and unified skew GLG. Results are based on $\mathcal{R} = 1000$ simulations each with a sample of size $n = 125$ (a subset of size 100 was randomly selected and used for model training, and the remaining 25 samples were used for prediction). ARPE: the average of the square root of the mean squared prediction errors; AIC: Akaike information criterion.

Parameter	True value	Gaussian		Skew Gaussian		GLG		Unified Skew GLG	
		Bias	variance	Bias	variance	Bias	variance	Bias	variance
β_0	2	-1.845	1.380	1.509	1.019	-1.418	1.175	-0.033	0.089
β_1	3.5	2.361	1.815	-1.926	1.743	1.653	1.351	0.272	0.147
μ_x	0	-3.739	3.391	-2.952	2.839	2.014	2.193	0.153	0.116
σ_x^2	1	2.162	2.052	1.903	1.410	1.835	1.544	0.329	0.417
γ	1.75	3.567	2.136	2.421	1.815	2.250	1.960	0.258	0.209
α	-2.5	-	-	1.855	2.079	-	-	-0.349	0.221
ν	0.85	-	-	-	-	-0.642	0.566	-0.305	0.238
θ	0.03	3.245	2.349	2.519	1.830	2.042	1.761	0.298	0.324
ARPE (VRPE)		11.957 (7.431)		8.539 (4.246)		7.265 (3.541)		0.960 (0.137)	
AIC		992.3		845.5		784.1		367.4	

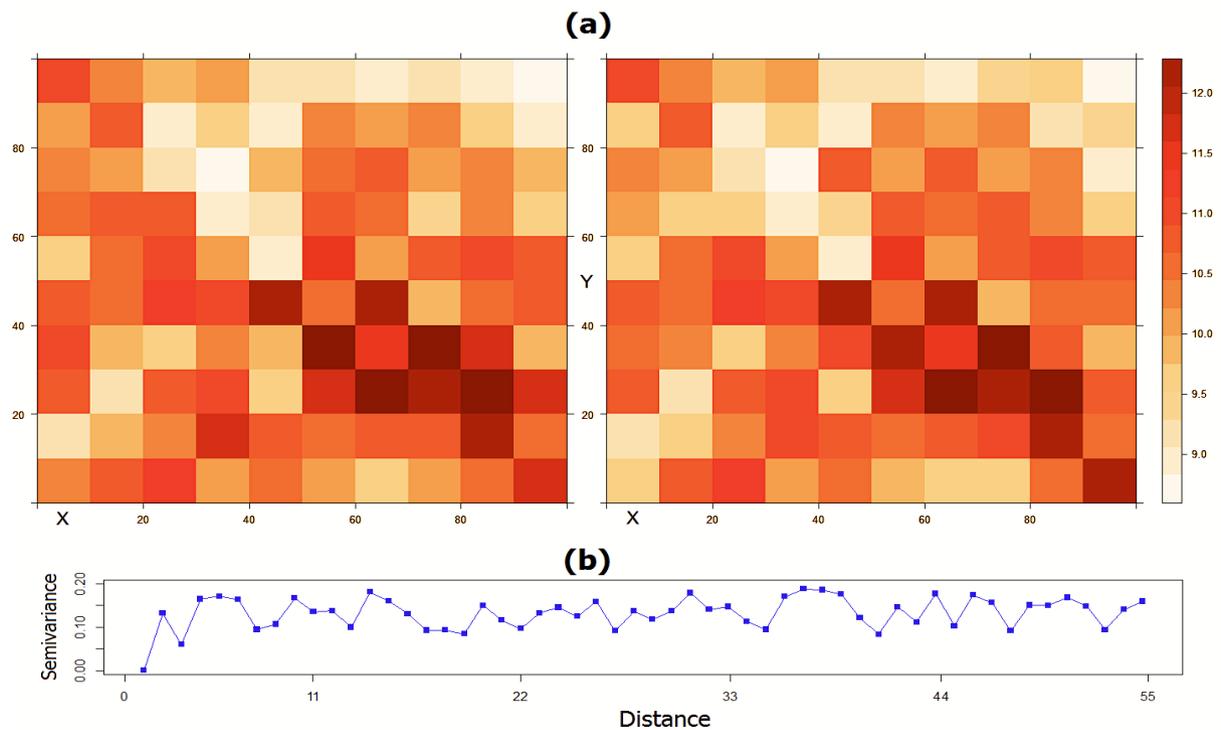


FIGURE 1 (a): Left panel shows the true spatial surface and right panel shows the predicted spatial surface. (b): Variogram of the difference between the true and predicted spatial surface.

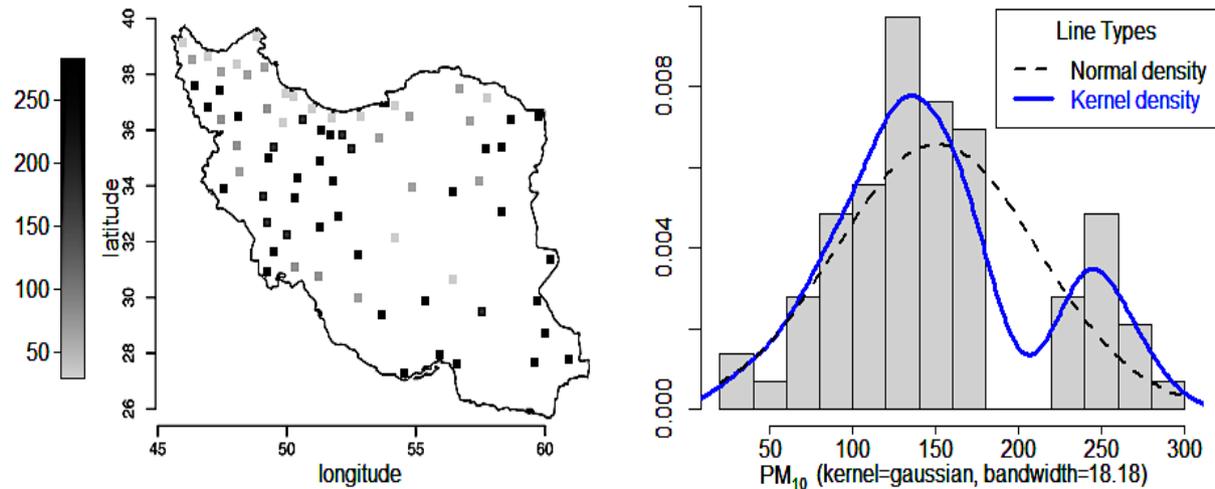


FIGURE 2 Left panel shows the spatial location of the synoptic weather stations in Iran; the coordinates (latitude and longitude) are in decimal degrees. Right panel displays the histogram, normal density, and kernel density estimates of the maximum PM_{10} concentration.

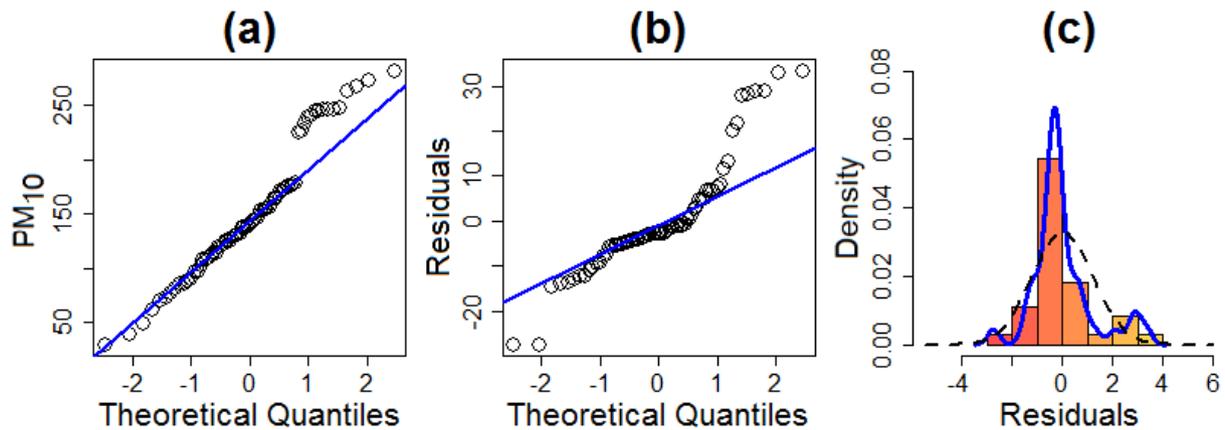


FIGURE 3 Panels (a) and (b) show the normal Q-Q plots of the maximum PM_{10} and the residuals, respectively. Panel (c) represents the normal residuals density estimate where the solid line is the kernel density estimate (with kernel=Gaussian and bandwidth=2.48) and the dashed line indicates normal density.

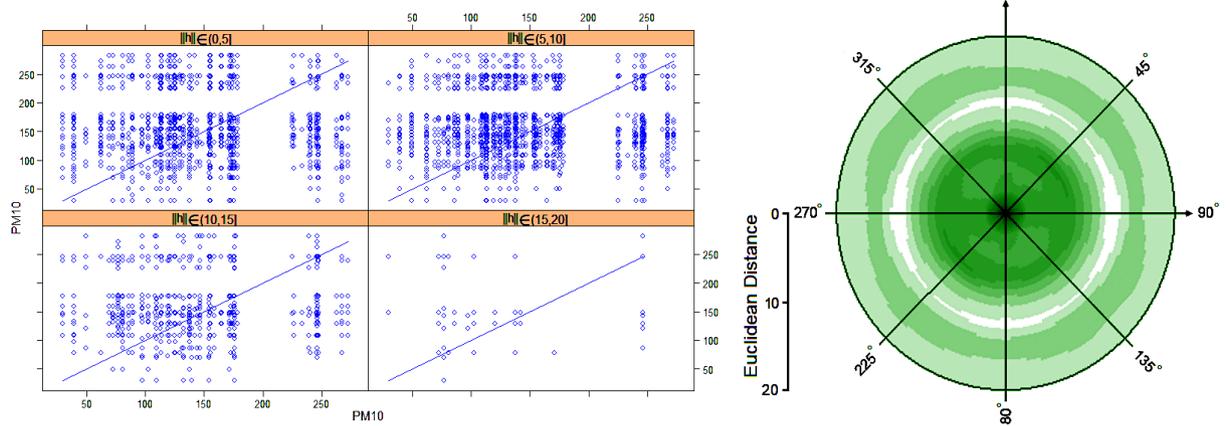


FIGURE 4 Left panel shows the H-scatterplot corresponding to the Euclidean distance $\|h\|$. Right panel shows the omnidirectional experimental semi-variogram of the maximum PM_{10} concentration.

TABLE 5 Results of fitting proposed model to the PM_{10} data with different values for σ^2 based on AIC.

σ^2	0.1	0.25	0.45	0.65	0.95	1.15	1.3
AIC	942.1	917.3	874.3	832.6	923.2	1012.0	1134.3

TABLE 6 The estimated value (EVal) and the estimated standard error (ESE) of parameters for the analysis of the maximum PM_{10} concentration using proposed model (USGLG) and its competitors: naive, GLG, skew Gaussian, and Gaussian. The model comparison is done using AIC.

Parameter	USGLG		USGLG (naive)		GLG ME		Skew Gaussian ME		Gaussian ME	
	EVal	ESE	EVal	ESE	EVal	ESE	EVal	ESE	EVal	ESE
β_0	16.38	0.51	18.42	1.65	14.34	2.17	19.10	2.26	19.02	3.31
β_1	3.64	0.45	2.97	1.26	4.87	2.03	2.21	2.44	2.47	3.05
μ_x	32.14	0.78	36.22	1.81	40.57	3.46	38.85	4.11	46.71	6.23
σ_x	9.67	0.83	14.31	2.15	15.52	2.58	15.21	2.63	19.30	5.12
γ	1.84	0.80	1.03	2.31	3.25	2.81	3.07	2.66	4.61	3.98
α	2.77	0.85	3.16	2.11	-	-	3.69	2.84	-	-
ν	0.82	0.96	0.94	1.35	2.31	1.41	-	-	-	-
θ	0.15	0.50	0.34	0.89	1.18	1.32	1.25	1.74	2.23	2.12
AIC	833.4		1450.2		1732.4		1736.7		1983.1	

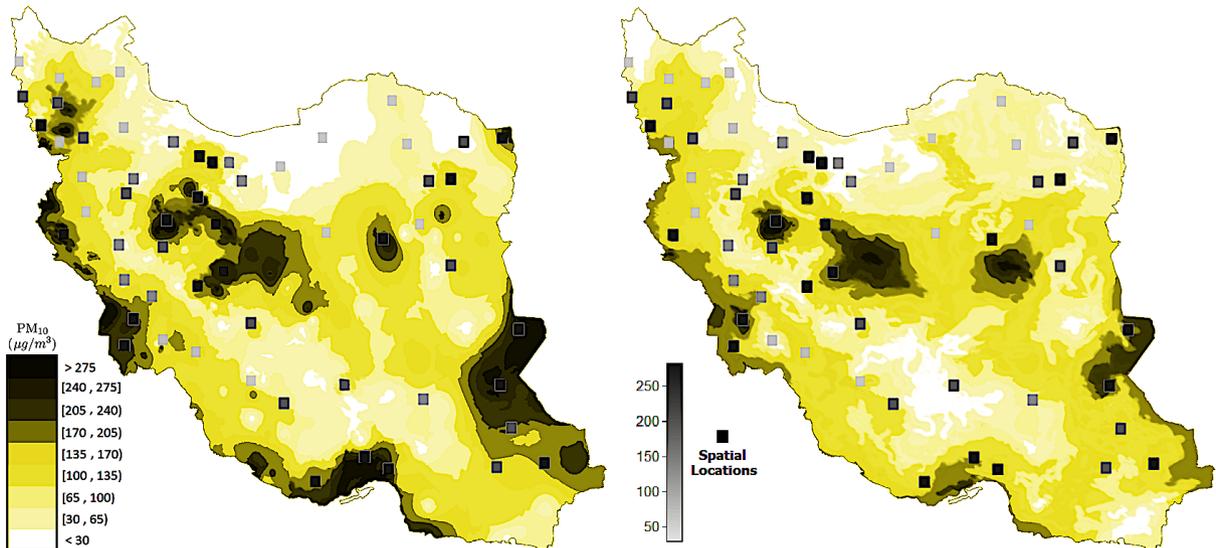


FIGURE 5 Map of predicted maximum PM_{10} concentration based on the proposed model (left panel) and Gaussian ME model (right panel).

