Stat Papers (2012) 53:205–218
DOI 10.1007/s00362-010-0329-2

REGULAR ARTICLE

Bayesian spatial regression models with closed skew normal correlated errors and missing observations

Omid Karimi · Mohsen Mohammadzadeh

Received: 22 June 2009 / Revised: 14 April 2010 / Published online: 19 May 2010

Abstract This paper is concerned with Bayesian estimation of a spatial regression model with skew non-Gaussian errors. The regression parameters are estimated by using a closed skew normal (CSN) distribution, which is closed under conditioning and linear combination. The proposed model captures skewness in the response variable. Sometimes, we may encounter missing observations in the response variable, accordingly we model and predict the missing observations by a Bayesian approach using Gibbs sampling methods. Next, a simulation study is performed to asses our model validity. Also, the proposed model in this work is applied to CO data from Tehran, the capital city of Iran. Then, the accuracy of the CSN and Gaussian models is compared by cross validation criterion.

Keywords Bayesian prediction · Closed skew normal · Missing observations · MCMC

Mathematics Subject Classification (2000) 62H11 · 91B72

1 Introduction

The analysis of spatial data is of interest in many scientific fields including agriculture, biology, geology and geography. One important object in the literature are regression models for spatial data. Anselin (1988, 1990) proposed approaches to incorporate spatial structural into regression models and Gschlößl and Czado (2008) presented Modelling count data with overdispersion and spatial effects. Shin and Sarkar (1994)

O. Karimi · M. Mohammadzadeh (✉)
Department of Statistics, Tarbiat Modares University, Tehran, Iran
e-mail: mohsen_m@modares.ac.ir

obtained the maximum likelihood estimation of the parameters in regression models with autocorrelated errors and Oh et al. (2002) exhibited Bayesian analysis for these models. Most of the theories in spatial regression analysis assume that the data are a realization of a Gaussian random field, where the errors are assumed to be normal. However, this assumption is not true for some of applications, for example the rainfall data in Kim and Mallick (2004), the seismic data in Karimi et al. (2010) and the air pollution data in Sect. 5.

When the distribution of the data is non-Gaussian but has many similar properties as a normal distribution, a skew normal distribution (Azzalini 1985) can be used to model their skewness. Recently in the literature, skew errors have been used in regression models; for example, see Sahu et al. (2003), Arellano-Valle et al. (2005), Bolfarinea and Lachos (2007). The closed skew normal (CSN) distribution (Dominguez-Molina et al. 2003), which is an extension of the multivariate skew normal distribution, has the advantage of being closed under marginalization and conditioning. Allard and Naveau (2007) studied a new spatial skew random field model in a spatial context. They used the class of CSN distributions and presented approaches for spatial interpolation. A random vector $\boldsymbol{Y}$ distributed according to the multivariate CSN with parameters; $\boldsymbol{\mu}$, Σ, D, $\boldsymbol{\nu}$, Δ, denoted by $\boldsymbol{Y} \sim CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$ has pdf given by

$$f_{p,q}(\boldsymbol{y}|\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = [\Phi_q(0; \boldsymbol{\nu}, \Delta + D\Sigma D')]^{-1} \times \phi_p(\boldsymbol{y}; \boldsymbol{\mu}, \Sigma)\Phi_q(D(\boldsymbol{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta), \quad (1)$$

where $\phi_p(\cdot; \boldsymbol{\mu}, \Sigma)$ is the p-dimensional normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, elements of $q \times p$ matrix D are skewness parameters and $\Phi_q(\cdot; \boldsymbol{\nu}, \Delta)$ is the q-dimensional normal cdf with mean $\boldsymbol{\nu}$ and covariance matrix Δ. When D is a zero matrix, the density of Eq. 1 reduces to the multivariate normal density. The multivariate skew normal density of Azzalini and Dalla Valle (1996) can be obtained by letting $q = 1$, $\boldsymbol{\nu} = \boldsymbol{0}$ and $\Delta = I$ in Eq. 1. In this paper, we apply a Bayesian approach of spatial regression models, when the errors are distributed as CSN. Since missing values occur in many real data, we consider missing observations for the response variable in the spatial regression model. Oh et al. (2002) presented Bayesian prediction for missing observation of a Gaussian random field. We extend their method for a general random field called closed skew Gaussian (CSG) random field, which includes Gaussian and skew Gaussian random fields. Skew Gaussian random field has been defined by Kim and Mallick (2004). Now, we define a set of random variables $\{Z(s), s \in U \subseteq R^d\}$ to be a CSG random field if for any finite integer m the random vector $(Z(s_1), \ldots, Z(s_m))'$ has a multivariate CSN distribution. The CSG random field has been completely defined in Karimi and Mohammadzadeh (2010) and a discrete form of this random field was applied to seismic data in Karimi et al. (2010).

This paper is organized as follows. In Sect. 2, the spatial regression model with autocorrelated errors is introduced. The Bayesian estimates of the parameters are obtained by MCMC algorithms in Sect. 3. Also the Bayesian prediction of each missing observation is presented. In Sect. 4, a simulation study is performed to check the validity of our model. The model is applied on a real data set in Sect. 5. Finally, a brief discussion is given in the last section.

2 The model

Basu and Reinsel (1994) considered the spatial regression model

$$y_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + z_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{2}$$

on a regular grids $\{(i, j) : i = 1, \ldots, m, \quad j = 1, \ldots, n\}$, where y_{ij}s are response variables, $\boldsymbol{x}_{ij}$ is a vector of r dimensional explanatory variables, $\boldsymbol{\beta} \in R^r$ is regression coefficients and z_{ij}s are correlated errors following a first order ARMA model. Oh et al. (2002) considered the regression model of Eq. 2 with multiplicative AR(1) errors

$$z_{ij} = \theta_1 z_{i-1,j} + \theta_2 z_{i,j-1} - \theta_1\theta_2 z_{i-1,j-1} + \varepsilon_{ij}, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n, \tag{3}$$

where ε_{ij}s are iid normal errors having mean zero and variance σ^2 and $|\theta_1| < 1$, $|\theta_2| < 1$. They presented Bayesian estimation of the model parameters when the data are Gaussian. But their method can not be used for skew or other non-normal data. Therefore, we propose the CSN model that is more general than the Gaussian model by modeling the skewness of the data. Let $\boldsymbol{Y} = (y_{11}, y_{12}, \ldots, y_{mn})'$, $X = (\boldsymbol{x}_{11}, \boldsymbol{x}_{12}, \ldots, \boldsymbol{x}_{mn})'$, $\boldsymbol{z} = (z_{11}, z_{12}, \ldots, z_{mn})'$, $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \varepsilon_{12}, \ldots, \varepsilon_{mn})'$ and $\boldsymbol{z}_0 = (z_{00}, z_{01}, \ldots, z_{m0})'$. Then Eqs. 2 and 3 can be written as

$$\boldsymbol{Y} = X\boldsymbol{\beta} + \boldsymbol{z}, \tag{4}$$

$$\boldsymbol{z} = B\boldsymbol{z} + A\boldsymbol{z}_0 + \boldsymbol{\varepsilon}, \tag{5}$$

where $\boldsymbol{z}_0$ denotes the vector of unobserved initial values of z_{ij}, B is a $mn \times mn$ lower triangular matrix and A is a $mn \times (m + n + 1)$ upper triangular matrix with elements zeros and functions of θ_1 and θ_2, (for examples of A and B see Appendix A). A Eq. 5 can be written as $(I - B)\boldsymbol{z} = A\boldsymbol{z}_0 + \boldsymbol{\varepsilon}$, where $(I - B)$ is a nonsingular matrix with determinant 1. Therefore, $\boldsymbol{z} = (I - B)^{-1}A\boldsymbol{z}_0 + (I - B)^{-1}\boldsymbol{\varepsilon}$. Let $W = (I - B)^{-1}$, thus $\boldsymbol{z} = WA\boldsymbol{z}_0 + W\boldsymbol{\varepsilon}$.

Proposition 1 *In the spatial regression model as in Eq. 2 with correlated errors of Eq. 3,*

(i) *let $\boldsymbol{\varepsilon}$ be an additive model which includes $\boldsymbol{v} \sim N_{mn}(\boldsymbol{0}, I_{mn})$ and $\boldsymbol{u} \sim N_q^{\nu}(\boldsymbol{0}, \Delta + D\Sigma D')$, then*

$$\boldsymbol{\varepsilon} = \boldsymbol{\mu} + K\boldsymbol{v} + G\boldsymbol{u} \sim CSN_{mn,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta),$$

where $\boldsymbol{\mu}$ is constant, $K = (\Sigma^{-1} + D'\Delta^{-1}D)^{-\frac{1}{2}}$, $G = \Sigma D'(\Delta + D\Sigma D')^{-1}$ and $N_q^{\nu}(\cdot, \cdot)$ is a normal distribution truncated below at $\boldsymbol{\nu}$. Thus

$$\boldsymbol{Y} \sim CSN_{mn,q}(\boldsymbol{\mu}_y, \Sigma_y, DW^{-1}, \boldsymbol{\nu}, \Delta),$$

where $\boldsymbol{\mu}_y = X\boldsymbol{\beta} + WA\boldsymbol{z}_0 + W\boldsymbol{\mu}$ and $\Sigma_y = W\Sigma W'$.

(ii) *Define* $\boldsymbol{Y}^* = (\boldsymbol{Y}'_{obs}, \boldsymbol{Y}'_{mis})' = Q\boldsymbol{Y}$ *to separate missing values* ($\boldsymbol{Y}_{mis}$) *from observed values* ($\boldsymbol{Y}_{obs}$) *in the vector* $\boldsymbol{Y}$, *where* Q *is an orthogonal matrix, that reorders the elements of* $\boldsymbol{Y}$. *Thus*

$$\boldsymbol{Y}_{mis}|\boldsymbol{y}_{obs} \sim CSN_{N_{mis},q}(\boldsymbol{\mu}_{mis} + \Sigma_{mo}\Sigma_{oo}^{-1}(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs}), \times \Sigma_{mm.o}, D_{mis}, \boldsymbol{\nu}_m, \Delta), \tag{6}$$

where N_{mis} *is dimension of missing values,* $\boldsymbol{\nu}_m = \boldsymbol{\nu} - D^*(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs})$, $D^* = D_{obs} + D_{mis}\Sigma_{mo}\Sigma_{oo}^{-1}$, $\Sigma_{mm.o} = \Sigma_{mm} - \Sigma_{mo}\Sigma_{oo}^{-1}\Sigma_{om}$ *and* $\boldsymbol{\mu}_{mis}$, $\boldsymbol{\mu}_{obs}$, Σ_{oo}, Σ_{mo}, Σ_{om}, Σ_{mm}, D_{mis}, D_{obs} *come from the partitions*

$$Q\boldsymbol{\mu}_y = \begin{pmatrix} \boldsymbol{\mu}_{obs} \\ \boldsymbol{\mu}_{mis} \end{pmatrix}, \quad Q\Sigma_y Q' = \begin{pmatrix} \Sigma_{oo} & \Sigma_{om} \\ \Sigma_{mo} & \Sigma_{mm} \end{pmatrix}, \quad DW^{-1}Q' = \begin{pmatrix} D_{obs} & D_{mis} \end{pmatrix}.$$

Proof see Appendix B. □

Now, by using the conditional distribution of Eq. 6, the adequate predictor of the missing values under square-error loss is given by

$$E(\boldsymbol{Y}_{mis}|\boldsymbol{y}_{obs}) = \boldsymbol{\mu}_{mis} + \Sigma_{mo}\Sigma_{oo}^{-1}(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs}) + \Sigma_{mm.o}D'_{mis}\Psi,$$

where

$$\Psi = \frac{\Phi_q^*(\boldsymbol{0}; \boldsymbol{\nu} - D^*(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs}), \Delta + D_{mis}\Sigma_{mm.o}D'_{mis})}{\Phi_q(\boldsymbol{0}; \boldsymbol{\nu} - D^*(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs}), \Delta + D_{mis}\Sigma_{mm.o}D'_{mis})},$$

and for any positive definite matrix; Ω, $\Phi_q^*(\boldsymbol{s}; \boldsymbol{\nu}, \Omega) = \nabla_s \Phi_q(\boldsymbol{s}; \boldsymbol{\nu}, \Omega)$, where $\nabla_s = (\frac{\partial}{\partial s_1}, \ldots, \frac{\partial}{\partial s_q})'$ is the gradient operator. Expectation of CSN distribution has been defined by Dominguez-Molina et al. (2003).

3 Bayesian spatial regression model

In this section, we present the Bayesian estimation of the model parameters and Bayesian prediction of the missing values by using MCMC algorithms. In the model of Eq. 5, we assume that $\boldsymbol{\varepsilon} \sim CSN_{mn,q}(\boldsymbol{0}, \sigma^2 I_{mn}, D, \boldsymbol{0}, I_q)$ thus according to Proposition 1

$$\boldsymbol{Y} \sim CSN_{mn,q}(\boldsymbol{\mu}_y, \sigma^2 WW', DW^{-1}, 0, I_q), \tag{7}$$

where $\boldsymbol{\mu}_y = X\boldsymbol{\beta} + WAz_0$. We define two cases for matrix skewness of D to reduce the dimension of skewness parameters as follows:

(i) The skewness can be fixed throughout spatial regions. Therefore, $D = \alpha J_{q\times mn}$, where α is an unknown scalar skewness parameter and $J_{q\times mn}$ is a $q \times mn$ matrix with unit elements. This case is approximately similar to the model of Kim and Mallick (2004).

(ii) Different values of skewness can be considered in every grid of the spatial region. Thus, D can be defined as $D = \left(\alpha_1 J_1', \ldots, \alpha_k J_k'\right)'$, where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_k)'$ is a vector of unknown skewness parameters and J_i's are $q_i \times mn$ with unit elements and $\sum_{i=1}^{k} q_i = q$. We consider the case (ii) which is more general than the case (i). This model has been presented by Karimi and Mohammadzadeh (2010).

Therefore the model parameters are $\boldsymbol{\eta} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}', \boldsymbol{\theta}')'$, where $\boldsymbol{\beta}$ is a vector of regression coefficients and $\boldsymbol{\theta} = (\theta_1, \theta_2)'$ is a vector of spatial correlation parameters. The Bayesian model requires us to adopt prior distributions for all the unknown parameters. We consider prior independence as $\pi(\boldsymbol{\eta}) = \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\theta}) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\theta})$, thus the posterior distribution is proportional to $f(\boldsymbol{y}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}) = f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\theta})\pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\theta})$. Liseo and Loperfido (2006) discussed the choice of prior for skew-normal distribution and obtained reference and Jeffreys priors for the scalar skew-normal distribution. They showed that Jeffreys and reference priors are proper for the skewness of the multivariate skew-normal distribution. We adopt proper priors for all the unknown parameters, to insure that the posterior distribution is proper. For the common prior distributions, $\boldsymbol{\beta} \sim N_r(\boldsymbol{\beta}_0, \Sigma_0)$, $\sigma^2 \sim IG(\lambda_0, \sigma_0)$, $\boldsymbol{\alpha} \sim N_k(\boldsymbol{\alpha}_0, \tau_0^2 I_k)$ and $\theta_i \sim N(\theta_{0i}, \psi_i^2) I(|\theta_i| < 1)$, $i = 1, 2$, the posterior distribution is given by

$$
\begin{aligned}
\pi(\boldsymbol{\eta}|\boldsymbol{y}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\theta})\pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\boldsymbol{\alpha})\pi(\boldsymbol{\theta}) \\
&\propto [\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2 DD')]^{-1} \phi_{mn}(\boldsymbol{y}; \boldsymbol{\mu}_y, \sigma^2 WW') \\
&\quad \Phi_q(DW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y); 0, I_q) \\
&\quad \times \phi_r(\boldsymbol{\beta}; \boldsymbol{\beta}_0, \Sigma_0) \frac{\sigma_0^{\lambda_0}}{\Gamma(\lambda_0)} \left(\frac{1}{\sigma^2}\right)^{\lambda_0+1} \exp\left\{-\frac{\sigma_0}{\sigma^2}\right\} \\
&\quad \phi_k(\boldsymbol{\alpha}; \boldsymbol{\alpha}_0, \tau_0^2 I_k)\phi(\theta_1; \theta_{01}, \psi_1^2) \\
&\quad \times \phi(\theta_2; \theta_{02}, \psi_2^2) I(|\theta_1| < 1) I(|\theta_2| < 1).
\end{aligned} \tag{8}
$$

Since this is a complicated distribution, we use MCMC techniques to obtain quantities of the posterior distribution. To use the Gibbs sampling algorithm, the full conditional posterior distributions are obtained as follows:

$$
\begin{aligned}
\pi(\sigma^2|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\theta}) &\propto IG(a_\sigma, b_\sigma)[\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2 DD')]^{-1}, \\
\pi(\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\theta}) &\sim CSN_{r,q}(\boldsymbol{\mu}_\beta, \Sigma_\beta, -DW^{-1}X, \boldsymbol{\nu}_\beta, I_q), \\
\pi(\boldsymbol{\alpha}|\boldsymbol{y}, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\theta}) &\propto \phi_k(\boldsymbol{\alpha}; \boldsymbol{\alpha}_0, \tau_0^2 I_k)\Phi_q(DW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y); 0, I_q) \\
&\quad \times [\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2 DD')]^{-1}, \\
\pi(\theta_i|\boldsymbol{y}, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\alpha}, \theta_j) &\propto \phi_{mn}(\boldsymbol{y}; \boldsymbol{\mu}_y, \sigma^2 WW')\phi(\theta_i; \theta_{0i}, \psi_i^2) \\
&\quad \Phi_q(DW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y); 0, I_q) \\
&\quad \times I(|\theta_i| < 1), \quad i = 1, 2\,, \ j = 1, 2\,, \ i \neq j,
\end{aligned}
$$

where

$$\begin{aligned} a_\sigma &= \frac{mn}{2} + \lambda_0, \quad b_\sigma = \frac{1}{2}\{(\boldsymbol{y} - \boldsymbol{\mu}_y)'(WW')^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y) + 2\sigma_0\}, \\ \Sigma_\beta &= (X'(\sigma^2 WW')^{-1}X + \Sigma_0^{-1})^{-1}, \\ \boldsymbol{\mu}_\beta &= \Sigma_\beta[X'(\sigma^2 WW')^{-1}(\boldsymbol{y} - WAz_0) + \Sigma_0^{-1}\boldsymbol{\beta}_0], \end{aligned}$$

and $\boldsymbol{\nu}_\beta = DW^{-1}(WAz_0 + X\boldsymbol{\mu}_\beta - \boldsymbol{y})$ (see Appendix C). According to Proposition 1, the full conditional posterior distribution of missing values is as

$$\boldsymbol{Y}_{mis}|\boldsymbol{\eta}, \boldsymbol{y}_{obs} \sim CSN_{N_{mis},q}(\boldsymbol{\mu}_{mis} + \Sigma_{mo}\Sigma_{oo}^{-1}(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs}), \Sigma_{mm.o}, D_{mis}, \boldsymbol{\nu}_m, \Delta).$$

Since the full conditionals of σ^2, $\boldsymbol{\alpha}$ and θ_is have no closed form, we use the Metropolis-Hastings algorithm for generating samples from them. The full conditional of σ^2 can be written as $\pi(\cdot) \propto f(\cdot)\Phi_q(\cdots)^{-1}$, where $f(\cdot)$ is a density that can be sampled, so the probability of move in the Metropolis-Hastings algorithm is reduced to $r(x, y) = \min(\frac{\Phi_q(x)}{\Phi_q(y)}, 1)$ (see Kim and Mallick (2004)). Therefore, $IG(a_\sigma, b_\sigma)$ can be used as a proposal distribution for σ^2.

Proposition 2 *Let the skewness parameter of the Bayesian CSN model (8) has the special form $D = \alpha J$, where α is a scalar and J is a matrix with unit elements. Then the full conditional of α is proportional to*

$$CSN_{1,q}(\alpha_0, \tau_0^2, D_\alpha, \boldsymbol{\nu}_\alpha, I_q)[\Phi_q(\boldsymbol{0}; \boldsymbol{0}, I_q + \sigma^2\alpha^2 JJ')]^{-1},$$

where $D_\alpha = JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$ and $\boldsymbol{\nu}_\alpha = -JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)\alpha_0$.

Proof see Appendix D. □

According to Proposition 2, the CSN distribution $CSN_{1,q}(\alpha_0, \tau_0^2, D_\alpha, \boldsymbol{\nu}_\alpha, I_q)$ can be used as proposal distribution for α. For the case of k skewness parameters given by $D = \left(\alpha_1 J_1', \ldots, \alpha_k J_k'\right)'$, the full conditional of $\boldsymbol{\alpha}$ has no special form. To determine a proposal distribution for $\boldsymbol{\alpha}$, the CSN distribution $CSN_{k,q}(\boldsymbol{\alpha}_0, \tau_0^2 I_k, D_\alpha, \boldsymbol{\nu}_\alpha, I_q)$ can be used, which is close to the kernel density of the full conditional of $\boldsymbol{\alpha}$. The truncated normal distribution in interval $(-1, 1)$ is considered as proposal distributions for both θ_1 and θ_2. We apply Gibbs sampler and Metropolis-Hastings together to generate a sample from the posterior distribution. Then the parameters can be estimated by $\hat{\eta} = \frac{1}{M-burn}\sum_{i=burn+1}^{M} \eta_{(i)}$ and the missing values can be predicted by $\hat{Y}_{mis} = \frac{1}{M-burn}\sum_{i=burn+1}^{M} Y_{mis(i)}$ where M is number of generated samples, $burn$ is burn-in values in the MCMC algorithms, $\eta_{(i)}$ and $Y_{mis(i)}$ are ith generated sample.

4 Simulation study

In this section, first we present a method for generating samples from a CSN distribution. Then the Bayesian estimates of the model parameters and the Bayesian

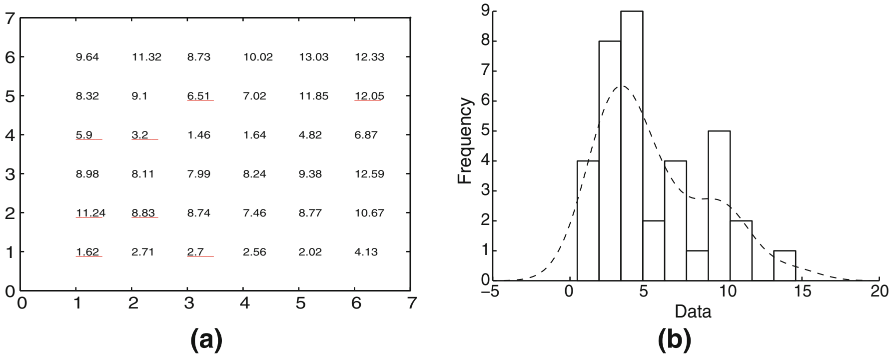


Fig. 1 **a** Simulated data and their locations, missing values are *underlined*; **b** Histogram of the simulated data

prediction of the missing values are obtained in a simulation example to assess the validity of the model. Also, we compare the proposal model with Gaussian model to show improvement in results when the data have some skewness in their distribution.

To generate a sample from distribution $CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$ the following steps can be repeated: (i) Generate $\boldsymbol{u}$ from distribution $N_q(\boldsymbol{\nu}, \Delta)$ and $\boldsymbol{y}$ from distribution $N_q(\boldsymbol{\mu}, \Sigma)$ (ii) Compute $\boldsymbol{w} = \boldsymbol{u} - D(\boldsymbol{y} - \boldsymbol{\mu})$. (iii) If $\boldsymbol{w} \leq \boldsymbol{0}$ accept $\boldsymbol{y}$ as a sample from $CSN_{p,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$ otherwise repeat steps.

Example A sample of $\boldsymbol{Y}$ is simulated in a regular grid 6×6 by the spatial regression model of Eq. 2 where

$$\begin{cases} \boldsymbol{x}'_{ij} = (1, i)\ ,\ \boldsymbol{\beta}' = (\beta_0, \beta_1) \text{ with } \beta_0 = 2 \text{ and } \beta_1 = 1, \\ z \text{ from a AR(1) model as } z{=}WAz_0 + W\boldsymbol{\epsilon} \text{ with parameters } \theta_1 = 0.9 \text{ and } \theta_2 = 0.7, \\ \boldsymbol{\epsilon}\ \sim CSN_{mn,q}(0, \sigma^2 I_{mn}, D, 0, I_q), \\ \sigma^2 = 2,\ D' = (\alpha_1 \mathbf{1}',\ \alpha_2 \mathbf{1}'),\ mn = 36,\ q = 2,\ \alpha_1 = 2 \text{ and } \alpha_2 = 1. \end{cases}$$

and z_0 is a vector of initial values which is generated from $N(0, 1)$. We assume that $\boldsymbol{y}$ has 8 missing values as completely at random. The simulated data and their locations are shown in Fig. 1a. The missing values $\boldsymbol{y}_{mis}{=}(Y_{11}, Y_{12}, Y_{14}, Y_{22}, Y_{24}, Y_{31}, Y_{35}, Y_{65})$ are specified by underline. The histogram of the data in Fig. 1b shows some skewness in the distribution of the data. However, the number of the data is small and the histogram is not a strong evidence that the data depart from the Gaussian distribution. But, the CSG random field is more general than Gaussian random field. Therefore, we consider a CSG random field for modeling the data. The common priors in Sect. 3 are adapted with large variances for Bayesian analysis. The Bayesian estimates of the model parameters $\boldsymbol{\eta} = (\boldsymbol{\beta}', \sigma^2, \boldsymbol{\alpha}', \boldsymbol{\theta}')'$ are calculated by Metropolis-Hastings and Gibbs sampler algorithms. For bivariate and trivariate distributions, the multivariate normal cumulative distribution function is calculated basic on methods developed by Genz (2004). The algorithms have converged in 10,000 iterations by looking at running means over the MCMC iterations and the burn-in is chosen as 6,000 iterations for the all parameters. To compare the CSN model with a basic model, we also obtain

Table 1 Bayesian estimates (posterior means) and standard errors (S.E.) of the parameters for CSN and Gaussian models for 5000 simulated data set presented in Fig. 1

Model		σ^2	β_0	β_1	θ_1	θ_2	α_1	α_2
	True val.	2	2	1	0.9	0.7	2	1
CSN	Bay. est.	2.0394	2.5229	0.9854	0.9130	0.6421	2.1445	1.3804
	S.E.	0.0095	0.0232	0.0063	0.0009	0.0021	0.0203	0.0184
Gaussian	Bay. est.	3.0715	3.3425	1.0505	0.8668	0.5883	–	–
	S.E.	0.0191	0.0261	0.0094	0.0019	0.0041	–	–

Table 2 MSE and Bias for CSN and Gaussian models averaged for 5000 simulated data sets

Model		σ^2	β_0	β_1	θ_1	θ_2	α_1	α_2
CSN	MSE	0.3000	1.2884	0.1305	0.0080	0.0204	0.0248	0.1351
	Bias	0.0897	−0.1128	0.0968	−0.0293	−0.0220	0.0814	0.3329
Gaussian	MSE	0.7718	1.8909	0.1815	0.0088	0.0226	–	–
	Bias	−0.8167	0.1477	0.1586	−0.0466	−0.0562	–	–

Bayesian estimation of the parameters for Gaussian model. The results of the two models are summarized in Table 1. Bayesian estimates of the CSN model are closer to the true values than the Gaussian model. Mean square error (MSE) and bias of the parameters were obtained for the two models in Table 2 averaged over 5,000 simulated data sets with different seed numbers. The results of MSE and bias for all parameters shows adequacy of our proposed model. The algorithms have converged and burn-in values can be consider 6,000 iterations.

The Bayesian prediction of the missing values for the two models are presented in Table 3 as well as standard errors of the predictions. The prediction results of the CSN model are closer to the true values than the Gaussian model. Moreover, the prediction MSE (PMSE) were obtained for the two models in Table 4 for 5,000 simulated data sets. The PMSE values of the CSN model are less than the Gaussian model. Therefore, they show the adequacy of our proposed model to predict missing observations.

5 Application to data

Air pollution still is a major issue in many cities, and Tehran is one of the most air polluted cities among all. In this section, we consider CO (Carbon monoxide) one of the basic substances of air pollution. The air pollution data were collected from six air stations of Tehran on 1st March 2007 at 12:00 PM by Air Quality Control Company (AQCC). Measure unit of the CO values are ppm. To predict CO values from six air stations in Tehran, we divide this city to a regular grid 16×16. In this case, we only have observations in six locations and the other locations are considered as missing values. Figure 2a shows locations of the air stations and points of the missing values on Tehran map. Tehran has been divided to 22 regions (lines in Fig. 2a) by municipality

Table 3 Bayesian prediction and standard error of the missing values for CSN and Gaussian models for 5,000 simulated data set presented in Fig. 1

Model	y_{mis}	Y_{13}	Y_{21}	Y_{22}	Y_{23}	Y_{25}	Y_{33}	Y_{41}	Y_{45}
	True val.	1.62	11.24	5.9	8.83	3.2	2.7	6.51	12.05
CSN	Bay. est.	2.2412	10.698	5.6849	9.6593	4.1965	2.5801	6.5352	12.207
	S. E.	0.0125	0.0127	0.0123	0.0241	0.0121	0.0105	0.0121	0.0227
Gaussian	Bay. est.	2.7368	10.072	6.6975	9.8755	5.1770	3.8092	7.1899	13.520
	S. E.	0.0137	0.0143	0.0127	0.0330	0.0132	0.0112	0.0115	0.0244

Bayesian predictions are mean of generated samples from the full conditionals of the missing values after removing burn-in values

Table 4 MSE prediction and Bias of the missing values for CSN and Gaussian models averaged for 5000 simulated data sets

Model	y_{mis}	Y_{13}	Y_{21}	Y_{22}	Y_{23}	Y_{25}	Y_{33}	Y_{41}	Y_{45}
CSN	PMSE	1.1983	1.3141	0.9337	1.1284	1.0472	0.7926	0.8260	1.3974
	Bias	0.0706	−0.0582	0.0321	−0.1006	0.0603	0.0498	−0.0213	−0.0166
Gaussian	PMSE	1.2030	1.4265	1.3230	1.1307	1.0815	0.8527	0.9145	1.4662
	Bias	0.1019	−0.0634	0.0391	−0.1132	0.0112	0.0921	−0.0380	−0.1029

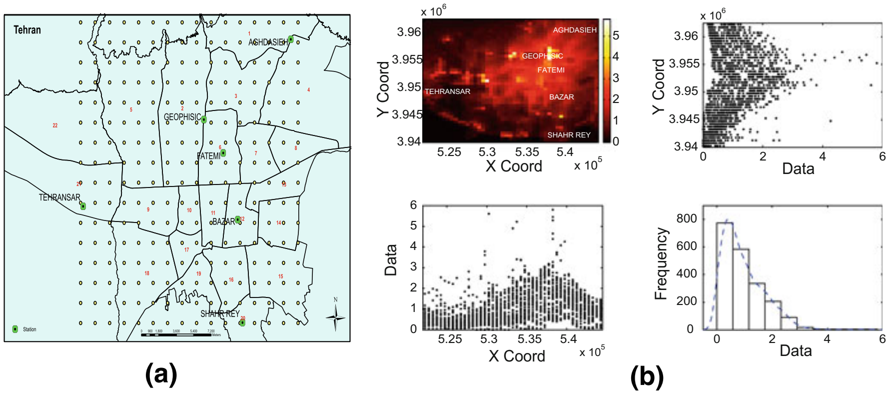


Fig. 2 **a** The CO values of the six air stations located in a regular grid 16 × 16 on Tehran map, where *yellow points* are location of missing values. *Lines* and *red numbers* specify 22 regions divided by municipality of Tehran. **b** Scatter plots of the CO modeling data and their histogram. Measure unit of the CO values are ppm

of Tehran. There is a physical modeling in AQCC based on physiological conditions to predict the air pollution data in Tehran. There is no stochastic statistical model in the physical modeling. The histograms of the CO data show that their distributions are skewed (Fig. 2b). Scatter plots in Fig. 2b illustrate that the data have trends in x and y directions proportional to bivariate normal distribution. So, we fitted a trend to the data as $g(y_{ij}) = 10^{8.5}\phi_2((i, j); coord_{Fatemi}, 10^7 I_2)$ where $\phi_2(\cdot, \boldsymbol{\mu}, \Sigma)$ is bivariate

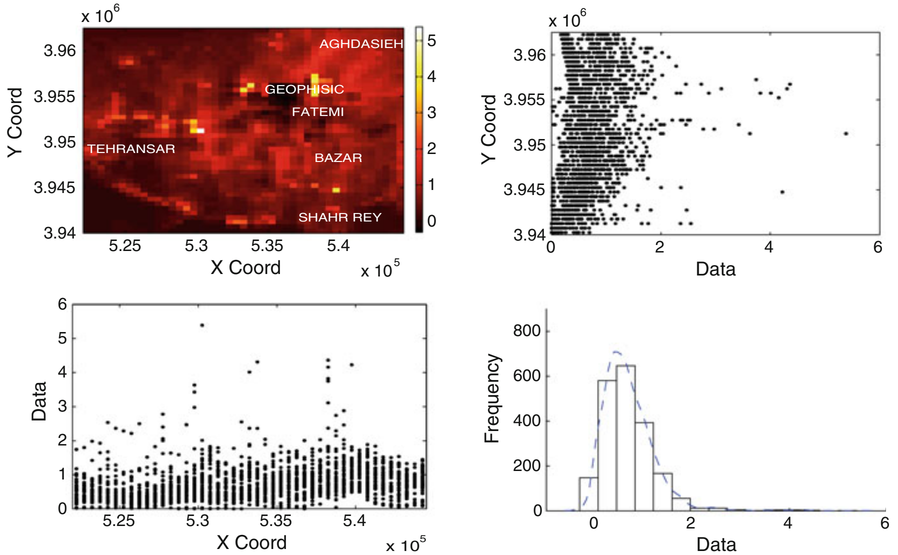


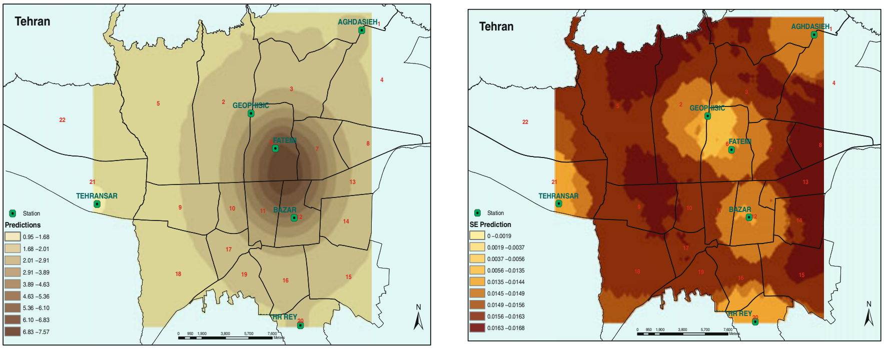
Fig. 3 Scatter plots of the CO modeling data after removing the trend

Table 5 Bayes estimates of the parameters

Parameter	σ^2	β_0	α_1	α_2	θ_1	θ_2
Bayesian Est.	0.6411	1.9742	3.0380	1.1753	−0.1045	0.1443
S. E.	0.0031	0.0023	0.0162	0.0164	0.0027	0.0030

normal density with mean $\boldsymbol{\mu}$ and covariance Σ and $Coord_{Fatemi}$ denotes the location of Fatemi station. Figure 3 shows the modeling data after removing the trend. We can see in Fig. 3 that the trend was almost removed. However, this model is not the best model to model trend of the data. Notice, we have used the physical modeling data to estimate the trend model. Then, the CSN model is applied on the real data (the six stations) to predict the missing values. After a summary exploratory data analysis, we define the spatial regression model as $y_{ij} - g(y_{ij}) = \beta_0 + z_{ij}$ where z_{ij}s are modeled as in Eq. 3 and $\boldsymbol{\varepsilon} \sim CSN_{256,2}(\mathbf{0}, \sigma^2 I_{256}, D, \mathbf{0}, I_2)$. Here, we consider two skewness parameters similar to the simulation example as $D = (\alpha_1 \mathbf{1}'_{256}, \alpha_2 \mathbf{1}'_{256})$. So, the model parameters are $\boldsymbol{\eta} = (\beta_0, \sigma^2, \boldsymbol{\alpha}, \boldsymbol{\theta})$.

For Bayesian analysis, the common priors are adopted as $(\beta_0|\beta_{01}) \sim N(\beta_{01}, 3)$, $\sigma^2|\lambda_0, \sigma_0) \sim IG(\lambda_0, \sigma_0)$, $(\boldsymbol{\alpha}|\boldsymbol{\alpha}_0, \tau_0^2) \sim N_2(\boldsymbol{\alpha}_0, \tau_0^2 I_2)$ and $(\theta_i|\theta_{0i}) \sim N(\theta_{0i}, 1)I(|\theta_{0i}| < 1)$ i=1,2, where β_{01}, λ_0, σ_0, α_0, τ_0^2 and θ_{0i} are considered as hyper parameters for using hierarchical Bayes to decrease the prior sensitivity. We use the hyper priors as follows: $\beta_{01} \sim U(-5, 5), \lambda_0 \sim U(0, 10), \sigma_0 \sim U(0, 5), \alpha_0 \sim U(-5, 5), \tau_0^2 \sim U(0, 10)$ and $\theta_{0i} \sim U(-1, 1)$. Using MCMC algorithms, Bayesian estimation of parameters is summarized in Table 5.



(a) (b)

Fig. 4 **a** Image of CO predictions with nine *gray* levels on Tehran map. **b** The prediction variance of CO predictions

Table 6 The CVMSE values of the CSN and Gaussian models

Model	CSN	Gaussian
CVMSE	1.2914	3.3183

Afterwards, a Bayesian prediction map is presented as an image with nine gray levels of CO prediction values on Tehran map in Fig. 4a. Obviously, we can see that city center and east side of Tehran is more polluted than the other parts of town. Figure 4b illustrates the variance of the CO prediction on the map. To check the accuracy of the Bayesian prediction for the CSN model, the cross validated mean-square error (CVMSE) is obtained as $\frac{1}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n}(y_{ij}-\hat{y}_{-ij})^2$ where $\hat{y}_{-ij}$ denotes prediction of y_{ij} without using the datum at grid point (i, j). Also the CVMSE value is obtained for a Gaussian model where the error terms are realizations from a Gaussian random field. The CVMSE values for CSN and Gaussian models in Table 6 show that the CSN model is more accurate than the Gaussian model.

6 Discussion and results

We modeled the skewness of the data by using CSN distribution for a spatial regression model with spatially autocorrelated errors and missing observations. In a Bayesian framework, we showed that the full conditionals of some model parameters and missing observations have closed form. Bayesian estimation of the model parameters were obtained in a simulation study, for the CSN model with two skewness parameters and the Gaussian model. We showed accuracy of our model by MSE criterion in this simulation.

We applied our model on air pollution data in Tehran city, then modeled the skewness of the data by the CSN model on a regular grid 16×16. We also used the physical modeling data to find the trend model. Then, Bayesian prediction of the CO data were illustrated on Tehran map at 12 PM. Finally, the MSECV was obtained for CSN and Gaussian models. It showed that the CSN model is better than Gaussian model.

The CSN model is analytically tractable, this provides the usage of time dimension for the CO data, while the time was considered to be fixed in this example. Therefore, we can use a space-time model in our method to forecast CO values in future time.

Usually, it is difficult to check if a small data set comes from a normal distribution or from a CSN distribution. Since the CSN distribution is more general than a normal distribution, the CSN model is proposed to model data related to the Gaussian model.

Acknowledgements Partial support from Ordered and Spatial Data Center of Excellence of Ferdowsi University of Mashhad is acknowledged. The authors thank two referees and editor for many helpful and valuable comments and suggestions which improved the presentation of this paper.

Appendix A

For example, in a 3×2 regular grid the matrices in Eq. 5 are given by

$$B = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ \theta_2 & 0 & 0 & 0 & 0 & 0 \\ \theta_1 & 0 & 0 & 0 & 0 & 0 \\ -\theta_1\theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 \\ 0 & 0 & \theta_1 & 0 & 0 & 0 \\ 0 & 0 & -\theta_1\theta_2 & \theta_1 & \theta_2 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} -\theta_1\theta_2 & \theta_1 & \theta_2 & 0 & 0 & 0 \\ 0 & -\theta_1\theta_2 & \theta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\theta_1\theta_2 & \theta_2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -\theta_1\theta_2 & \theta_2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Appendix B: Proof of proposition 1

Part (i): $\boldsymbol{\varepsilon}$ is a additive model as $\boldsymbol{\varepsilon} = \boldsymbol{\mu} + K\boldsymbol{v} + G\boldsymbol{u}$, then according to Dominguez-Molina et al. (2007)

$$\boldsymbol{\varepsilon} \sim CSN_{mn,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta).$$

Since $\boldsymbol{z}$ in Eq. 5 is a linear combination of $\boldsymbol{\varepsilon}$, due to the linear combination property of CSN distribution (Dominguez-Molina et al. 2003), $\boldsymbol{z} \sim CSN_{mn,q}(WA\boldsymbol{z}_0 + W\boldsymbol{\mu}, W\Sigma W', DW^{-1}, \boldsymbol{\nu}, \Delta)$. Similarly, $\boldsymbol{Y} \sim CSN_{mn,q}(\boldsymbol{\mu}_y, \Sigma_y, DW^{-1}, \boldsymbol{\nu}, \Delta)$, where $\boldsymbol{\mu}_y = X\boldsymbol{\beta} + WA\boldsymbol{z}_0 + W\boldsymbol{\mu}$ and $\Sigma_y = W\Sigma W'$.

Part (ii): Let $\boldsymbol{Y}_{mis}$ denotes the vector of missing observations and $\boldsymbol{Y}_{obs}$ be the vector of observed values. Let $\boldsymbol{Y}^* = (\boldsymbol{Y}'_{obs}, \boldsymbol{Y}'_{mis})' = Q\boldsymbol{Y}$, where Q is an orthogonal matrix, that reorders the elements of $\boldsymbol{Y}$. Thus, $\boldsymbol{Y}^* \sim CSN(\boldsymbol{\mu}_Q, \Sigma_Q, D_Q, \boldsymbol{\nu}, \Delta)$, where $\boldsymbol{\mu}_Q = Q\boldsymbol{\mu}_y, \Sigma_Q = Q\Sigma_y Q', D_Q = DW^{-1}Q'$. Since CSN distribution is closed under conditioning, thus

$$\boldsymbol{Y}_{mis}|\boldsymbol{y}_{obs} \sim CSN_{N_{mis},q}(\boldsymbol{\mu}_{mis} + \Sigma_{mo}\Sigma_{oo}^{-1}(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs}), \Sigma_{mm.o}, D_{mis}, \boldsymbol{\nu}_m, \Delta),$$

where N_{mis} is dimension of missing values, $\boldsymbol{\nu}_m = \boldsymbol{\nu} - D^*(\boldsymbol{y}_{obs} - \boldsymbol{\mu}_{obs})$, $D^* = D_{obs} + D_{mis}\Sigma_{mo}\Sigma_{oo}^{-1}$, $\Sigma_{mm.o} = \Sigma_{mm} - \Sigma_{mo}\Sigma_{oo}^{-1}\Sigma_{om}$ and $\boldsymbol{\mu}_{mis}, \boldsymbol{\mu}_{obs}, \Sigma_{oo}, \Sigma_{mo}$,

Σ_{om}, Σ_{mm}, D_{mis} and D_{obs} come from the partitions

$$Q\boldsymbol{\mu}_y = \begin{pmatrix} \boldsymbol{\mu}_{obs} \\ \boldsymbol{\mu}_{mis} \end{pmatrix}, \quad Q\Sigma_y Q' = \begin{pmatrix} \Sigma_{oo} & \Sigma_{om} \\ \Sigma_{mo} & \Sigma_{mm} \end{pmatrix}, \quad DW^{-1}Q' = \begin{pmatrix} D_{obs} & D_{mis} \end{pmatrix}.$$

Appendix C

Here, we show that the full conditional of $\boldsymbol{\beta}$ in the Bayesian CSN model of Eq. 8 is a CSN distribution as $CSN_{r,q}(\boldsymbol{\mu}_\beta, \Sigma_\beta, -DW^{-1}X, \boldsymbol{\nu}_\beta, I_q)$.

Proof

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\alpha}) &\propto f(\boldsymbol{y}|\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\alpha})\pi(\boldsymbol{\beta}) \\
&\propto \phi_{mn}(\boldsymbol{y}; \boldsymbol{\mu}_y, \sigma^2 WW')\Phi_q(DW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y); 0, I_q)\phi_r(\boldsymbol{\beta}; \boldsymbol{\beta}_0, \Sigma_0) \\
&\propto \phi_r(\boldsymbol{\beta}; \boldsymbol{\mu}_\beta, \Sigma_\beta)\Phi_q(DW^{-1}(\boldsymbol{y} - X\boldsymbol{\beta} - WA\boldsymbol{z}_0); \mathbf{0}, I_q),
\end{aligned}$$

where $\Sigma_\beta = (X'(\sigma^2 WW')^{-1}X + \Sigma_0^{-1})^{-1}$, $\boldsymbol{\mu}_\beta = \Sigma_\beta[X'(\sigma^2 WW')^{-1}(\boldsymbol{y} - WA\boldsymbol{z}_0) + \Sigma_0^{-1}\boldsymbol{\beta}_0]$. Thus

$$\begin{aligned}
\pi(\boldsymbol{\beta}|\boldsymbol{y}, \sigma^2, \boldsymbol{\theta}, \boldsymbol{\alpha}) &\propto \phi_r(\boldsymbol{\beta}; \boldsymbol{\mu}_\beta, \Sigma_\beta)\Phi_q(-DW^{-1}X\boldsymbol{\beta} + DW^{-1}(\boldsymbol{y} - WA\boldsymbol{z}_0); \mathbf{0}, I_q), \\
&= \phi_r(\boldsymbol{\beta}; \boldsymbol{\mu}_\beta, \Sigma_\beta)\Phi_q(-DW^{-1}X(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta) \\
&\quad + DW^{-1}(\boldsymbol{y} - WA\boldsymbol{z}_0 - X\boldsymbol{\mu}_\beta); \mathbf{0}, I_q), \\
&= \phi_r(\boldsymbol{\beta}; \boldsymbol{\mu}_\beta, \Sigma_\beta)\Phi_q(-DW^{-1}X(\boldsymbol{\beta} - \boldsymbol{\mu}_\beta); \\
&\quad DW^{-1}(WA\boldsymbol{z}_0 + X\boldsymbol{\mu}_\beta - \boldsymbol{y}); \mathbf{0}, I_q).
\end{aligned}$$

Let $\boldsymbol{\nu}_\beta = DW^{-1}(WA\boldsymbol{z}_0 + X\boldsymbol{\mu}_\beta - \boldsymbol{y})$, therefore the full conditional of $\boldsymbol{\beta}$ is a CSN distribution as $CSN_{r,q}(\boldsymbol{\mu}_\beta, \Sigma_\beta, -DW^{-1}X, \boldsymbol{\nu}_\beta, I_q)$. □

Appendix D: Proof of proposition 2

Set $D = \alpha J$ in the full conditional of α, thus $k = 1$ and

$$\begin{aligned}
\pi(\alpha|\boldsymbol{y}, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\theta}) &\propto \phi(\alpha; \alpha_0, \tau_0^2)\Phi_q(\alpha JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y); \mathbf{0}, I_q) \\
&\quad \times [\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2\alpha^2 JJ')]^{-1}, \\
&= \phi(\alpha; \alpha_0, \tau_0^2)\Phi_q(JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)(\alpha - \alpha_0) \\
&\quad + JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)\alpha_0; 0, I_q) \\
&\quad \times [\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2\alpha^2 JJ')]^{-1}, \\
&= \phi(\alpha; \alpha_0, \tau_0^2)\Phi_q(JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)(\alpha - \alpha_0); \\
&\quad -JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)\alpha_0, I_q) \\
&\quad \times [\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2\alpha^2 JJ')]^{-1},
\end{aligned}$$

set $D_\alpha = JW^{-1}(\boldsymbol{y} - \boldsymbol{\mu}_y)$ and $\boldsymbol{v}_\alpha = -D_\alpha \alpha_0$, thus

$$\pi(\alpha|\boldsymbol{y}, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\theta}) \propto \phi(\alpha; \alpha_0, \tau_0^2)\Phi_q(D_\alpha(\alpha - \alpha_0); \boldsymbol{v}_\alpha, I_q) \times [\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2\alpha^2 JJ')]^{-1}.$$

According to CSN density, full conditional of α is proportional to

$$CSN_{1,q}(\alpha_0, \tau_0^2, D_\alpha, \boldsymbol{v}_\alpha, I_q)[\Phi_q(\mathbf{0}; \mathbf{0}, I_q + \sigma^2\alpha^2 JJ')]^{-1}.$$

References

Allard D, Naveau P (2007) A new spatial skew-normal random field model. Commun Stat-Theory Method 36:1821–1834

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

Anselin L (1990) Spatial dependence and spatial structural instability in applied regression analysis. J Reg Sci 30:185–207

Arellano-Valle RB, Bolfarine H, Lachos VH (2005) Skew-normal mixed linear models. J Data Sci 3: 415–438

Azzalini A (1985) A class of distributions which includes the normal ones. Scand J Stat 12:171–178

Azzalini A, Dalla Valle A (1996) The multivariate skew-normal distribution. Biometrika 83:715–726

Basu S, Reinsel GC (1994) Regression models with spatially correlated errors. J Am Stat Assoc 89:88–99

Bolfarinea H, Lachos VH (2007) Skew-probit measurement error models. Stat Method 4:1–12

Dominguez-Molina J, Gonzalez-Farias G, Gupta A (2003) The multivariate closed skew normal distribution. Technical report 03-12, Department of Mathematics and Statistics, Bowling Green State University

Dominguez-Molina J, Gonzalez-Farias G, Ramos-Quiroga R, Gupta A (2007) A matrix variate closed skew-normal distribution with applications to stochastic frontier analysis. Commun Stat-Theory Method 36:1691–1703

Genz A (2004) Numerical computation of rectangular bivariate and trivariate normal and t probabilities. Stat Comput 14:251–260

Gschlößl S, Czado C (2008) Modelling count data with overdispersion and spatial effects. Stat Pap 49: 531–552

Karimi O, Mohammadzadeh M (2010) Bayesian spatial prediction for discrete closed skew Gaussian random field. Math Geosci (in press)

Karimi O, Omre H, Mohammadzadeh M (2010) Bayesian closed-skew Gaussian inversion of seismic AVO data for elastic material properties. Geophysics **75**:185–198. doi:10.1190/1.3299291

Kim HM, Mallick BK (2004) A Bayesian prediction using the skew Gaussian distribution. J Stat Plan Inference 120:85–101

Liseo B, Loperfido N (2006) A note on reference priors for the scalar skew-normal distribution. J Stat Plan Inference 136:373–389

Oh M, Shina DW, Kim HJ (2002) Bayesian analysis of regression models with spatially correlated errors and missing observations. Comput Stat Data Anal 39:387–400

Sahu SK, Dey DK, Branco MD (2003) A new class of multivariate skew distributions with applications to Bayesian regression models. Can J Stat 31:129–150

Shin DW, Sarkar S (1994) Parameter estimation in regression models with autocorrelated errors using irregular data. Commun Stat Theory Method 23:3567–3580