# Approximate Bayesian Inference in Spatial GLMM with Skew Normal Latent Variables

## Running headline: Skew Normal Spatial GLMM

**Fatemeh Hosseini**[1], **Jo Eidsvik**[2], **and Mohsen Mohammadzadeh**[1]

*1) Department of Statistics, Tarbiat Modares University, Tehran, Iran.*

*2) Department of Mathematical Sciences, NTNU, Norway.*

*Address: Department of Statistics, Tarbiat Modares University, 14115-175 Tehran, Iran.*

*email: f.hosseini@modares.ac.ir*

# Abstract

Spatial generalized linear mixed models are common in applied statistics. Most users are satisfied using a Gaussian distribution for the spatial latent variables in this model, but it is unclear whether the Gaussian assumption holds. Wrong Gaussian assumptions cause bias in parameter estimates and affect the accuracy of spatial predictions. Thus, there is a need for more flexible priors for the latent variables, and to perform efficient inference and spatial prediction in the resulting models. In this paper we use skew normal distribution for the spatial latent variables. We propose new approximate Bayesian methods for inference and spatial prediction in this model. A key ingredient in our approximations is using the closed skew normal distribution to approximate the full conditional for latent variables. Our approximate inference and spatial prediction methods are fast and deterministic, using no sampling based strategies.



KEYWORDS: *approximate Bayesian inference*, *closed skew normal distribution*, *geostatistics*, *latent variables*, *mcmc*, *skew normal distribution*, *spatial generalized linear mixed model*

## 1. Introduction

Spatial generalized linear mixed models (SGLMM) are commonly used for count or proportion data acquired over a continuous spatial domain, see e.g. Diggle *et al.* (1998), Zhang (2002), Zhao *et al.* (2006) and Zhu *et al.* (2007). A fully Bayesian SGLMM consists of i) a likelihood in the exponential family, with conditional independence given a spatial latent variable, ii) a prior model for the latent random field and iii) priors for the model parameters, such as regression parameters, scale or spatial correlation length. The by far most common assumption for the latent random field in step ii) is to use a Gaussian (normal) random field. Inference of model parameters and spatial prediction has been studied intensely for this normal assumption, see e.g. Breslow and Clayton (1993), Christensen and Waagepetersen (2002), Christensen (2004), Ainsworth and Dean (2006), Rue and Martino (2007), Rue *et al.* (2009) and Eidsvik *et al.* (2009). The most common tool for inference and prediction is Markov chain Monte Carlo (MCMC) sampling, which is very general and allows one to assess any joint distribution of any quantities. MCMC sampling can be time consuming if the Markov chain mixes slowly, and an alternative approach is to focus on posterior marginals and use approximate inference based on deterministic and numerical methods. Rue and Martino (2007), Rue *et al.* (2009) and Eidsvik *et al.* (2009) argue that this type of approximate Bayesian inference takes seconds of computation time, while MCMC sampling can take days.

Assuming a normal formulation for the latent random field in step ii) above is convenient from a computational point of view, but is rather restrictive. Erroneous normal assumptions have influence on the estimation of model parameters and the accuracy of spatial prediction. This motivates the search for SGLMM with more flexible distributions for the spatial latent variables. Komarek and Lesaffre (2008) proposed to replace the normal prior model in GLMM with a penalized Gaussian mixture distribution. Kim and Mallick (2004) proposed a skew normal (SN) random field for the spatial variables as an extension to the usual Gaussian Bayesian kriging. The SN distribution (Azzalini (1985), Azzalini and Dalla Valle (1996)) is of a greater class than the normal. It is more flexible since it includes the normal with an extra parameter to regulate skewness. A generalization of the SN distribution, called the closed skew normal (CSN) distribution, allows even more flexibility than the SN. Moreover, the CSN distribution remains closed under linear condi-

tioning and marginalization. Dominguez-Molina *et al.* (2003) and Gonzalez-Farias *et al.* (2004) defined the CSN as a generalization of the SN and of the density defined by Gupta *et al.* (2003).

In this paper we propose a SN distribution for the spatial latent variable. For the resulting SGLMM we extend the approximate Bayesian inference methods of Eidsvik *et al.* (2009). The main contribution of the current work is to use the CSN as an approximation for the full conditional for latent variable, and to show that this allows us to perform fast approximate inference and prediction, without MCMC sampling. The inference step is done in a manner similar to the Laplace approximation (Tierney and Kadane (1986)), but based on the CSN approximation instead of the usual normal approximation. Our presentation is written for the typical geostatistical application with discrete spatial data acquired on a non-lattice continuous domain. The methods generalize to other settings such as lattice data or observations obtained on a map of regions and their neighbors.

This paper is organized as follows: In Section 2 the SGLMM with SN latent variables is defined. In this Section we also introduce the CSN which is important for our methodology. The proposed method of approximate Bayesian inference and prediction is described in Section 3. Section 4 shows results on two datasets. Closing remarks are given in Section 5. The full derivations of the proposed methodology are provided in an Appendix.

## 2. Spatial GLMM with Skew Normal Latent Variables

In this section we first define the SN and CSN distributions, and discuss the properties of the CSN distribution that we need for our methodology. Then we present the hierarchical Bayesian representation for SGLMM with SN as prior model for the spatial latent variables.

### *2.1. Skew Normal and Closed Skew Normal Distribution*

Let $\boldsymbol{x}$ be a $n$-dimensional random vector that follows a multivariate SN distribution. Then its density function is given by

$$f(\boldsymbol{x}|\boldsymbol{\lambda},\boldsymbol{\mu},\Sigma) = 2\phi_n(\boldsymbol{x};\boldsymbol{\mu},\Sigma)\Phi(\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{\mu})), \tag{1}$$

where $\phi_n(\cdot;\boldsymbol{\mu},\Sigma)$ is the $n$-dimensional normal density with mean vector $\boldsymbol{\mu} \in R^n$, positive definite $n\times n$ covariance matrix $\Sigma$, standard normal cumulative distribution function $\Phi(\cdot)$, and where $\boldsymbol{\lambda} \in R^n$ is a $n$-dimensional skewness parameter. In short we denote the SN distribution by $SN(\boldsymbol{\mu},\boldsymbol{\Sigma},\boldsymbol{\lambda})$. The

skewness parameter $\boldsymbol{\lambda}$ stretch the normal density in one direction, in this case defined by $\boldsymbol{\lambda}'\Sigma^{-\frac{1}{2}}$, (Arellano-Valle *et al.* (2005)). If skewness parameter $\lambda_i = 0$, $i = 1, \ldots, n$, the density of $\boldsymbol{x}$ in equation (1) reduces to the multivariate normal distribution.

The CSN distribution is a larger class which includes the SN distribution as a special case. The CSN distribution has some very desirable properties, similar to those of the normal distribution. For instance, the CSN distribution is closed under marginalization, conditioning, and linear transformations (full column or row rank), see Dominguez-Molina *et al.* (2003). A $n$-dimensional random vector $\boldsymbol{x}$ follows a multivariate CSN distribution, with parameters $\boldsymbol{\mu}$, $\Sigma$, $D$, $\boldsymbol{\nu}$, and $\Delta$, if its density function is given by

$$f_{n,q}(\boldsymbol{x}|\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = \Phi_q^{-1}(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D^{'})\, \phi_n(\boldsymbol{x}; \boldsymbol{\mu}, \Sigma)\, \Phi_q(D(\boldsymbol{x} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta). \tag{2}$$

In short we denote this distribution by $CSN_{n,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$, where length $n$ vector $\boldsymbol{\mu}$ is a location parameter, positive definite $n \times n$ matrix $\boldsymbol{\Sigma}$ is a scale matrix, the elements of $q \times n$ matrix $D$ are skewness parameters and $\Phi_q(\cdot; \boldsymbol{\nu}, \Delta)$ is the $q$-dimensional normal cumulative distribution function with mean $\boldsymbol{\nu}$ and covariance matrix $\Delta$. If we set $\boldsymbol{\nu} = 0$, $\Delta = I$ and $q = 1$, the CSN density reduces to that of the SN distribution. When $D$ is a zero matrix, the density in equation (2) reduces to the normal distribution. Figure 1 shows contour plots of CSN densities ($n = 2$) for different parameter values.

—

Figure 1 about here.

—

We use some key properties of the CSN distribution in this paper, and now list these briefly. Let $\boldsymbol{x} \sim CSN_{n,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$, then Dominguez-Molina *et al.* (2003) show the following:

- Let $A$ be a $r \times n (r \leq n)$ constant matrix, such that $AA'$ is nonsingular matrix and $\boldsymbol{b} \in R^r$. Then

$$A\boldsymbol{x} + \boldsymbol{b} \sim CSN_{r,q}(A\boldsymbol{\mu} + \boldsymbol{b}, A\Sigma A', D\Sigma A'(A\Sigma A')^{-1}, \boldsymbol{\nu}, \Delta + D\Sigma D' - D\Sigma A'(A\Sigma A')^{-1}A\Sigma D'). \tag{3}$$

- Partition $\boldsymbol{x}$ into $\boldsymbol{x} = (\boldsymbol{x}_1', \boldsymbol{x}_2')'$, where the dimension of $\boldsymbol{x}_1$ is k and the dimension of $\boldsymbol{x}_2$ is $n-k$. Then define $A = [I_k\ \mathbf{0}]$, with $\mathbf{0}$ a $k \times (n-k)$ zero matrix, such that $\boldsymbol{x}_1 = A\boldsymbol{x}$. Marginally,

$$\boldsymbol{x}_1 \sim CSN_{k,q}(\boldsymbol{\mu}_1, \Sigma_{11}, D^*, \boldsymbol{\nu}, \Delta^*), \tag{4}$$

where $D^* = D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}$, $\Delta^* = \Delta + D_2\Sigma_{22\cdot1}D_2'$, $\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$, with $\boldsymbol{\mu}_1$, $\Sigma_{11}$, $\Sigma_{22}$, $\Sigma_{12}$, $\Sigma_{21}$, $D_1$ and $D_2$ obtained by the partitioning $\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}$, $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$ and $D = (D_1,\ D_2)$.

- With the previous partitioning, the conditional distribution of $\boldsymbol{x}_2$ given $\boldsymbol{x}_1$ is

$$(\boldsymbol{x}_2|\boldsymbol{x}_1) \sim CSN_{n-k,q}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22\cdot1}, D_2, \boldsymbol{\nu} - D^*(\boldsymbol{x}_1 - \boldsymbol{\mu}_1), \Delta). \tag{5}$$

- The first moment of the CSN distribution is

$$\mathrm{E}(\boldsymbol{X}) = \boldsymbol{\mu} + \Sigma D'\psi, \quad \psi = \psi(0, \Sigma, D, \nu, \Delta) \tag{6}$$

where $\psi(0, \Sigma, D, \nu, \Delta) = \frac{\Phi_q^*(r,\nu,\Delta+D\Sigma D')}{\Phi_q(0,\nu,\Delta+D\Sigma D')}|_{r=0}$, and for any positive definite matrix $\Omega$, $\Phi_q^*(r;\nu,\Omega) = [\nabla_r\Phi_q(r;\nu,\Omega)]'$ with gradient operator $\nabla_r = (\frac{\partial}{\partial r_1}, \cdots, \frac{\partial}{\partial r_q})'$. For example, when $q = 1$, $\Phi_1^*(r;\nu,\Omega)|_{r=0} = \Omega^{-\frac{1}{2}}\phi(\nu\Omega^{-\frac{1}{2}})$.

- A sample from the CSN distribution can be obtained by a rejection sampling strategy as follows: i) Draw $\boldsymbol{w}_0 \sim N(\boldsymbol{\nu}, \Delta)$, ii) Draw $\boldsymbol{w}_1 \sim N(\boldsymbol{\mu}, \Sigma)$, iii) Set $\boldsymbol{w}_2 = \boldsymbol{w}_0 - D\boldsymbol{w}_1$. iv) Accept $\boldsymbol{w}_1$ if $\boldsymbol{w}_2 < \mathbf{0}$ for all components.

### *2.2. Hierarchical Bayesian model*

The SGLMM is defined at three levels: i) likelihood model for data $\boldsymbol{y}$ assumed to be of an exponential family distribution, ii) a SN prior for the spatial latent variable $\boldsymbol{x}$, iii) priors for the model parameters $\boldsymbol{\eta}$ in the mean, spatial interaction and skewness parameters of this SN density.

We first discuss the SN prior model. Let $\boldsymbol{x} = (x_1, \cdots, x_n)'$ be spatial latent variables at $n$ sites $\{s_1, \cdots, s_n\}$ with density $\pi(\boldsymbol{x}|\boldsymbol{\eta}) = SN(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda})$. Here, the location parameter $H\boldsymbol{\beta}$ consists of $n \times (p+1)$ matrix $H$ of covariates, and $p+1$ regression parameters $\boldsymbol{\beta} = (\beta_0, \cdots, \beta_p)'$. The spatial interaction matrix $\Sigma_\theta$ is a positive definite $n \times n$ matrix, with two dimensional parameter $\boldsymbol{\theta} = (\sigma, \varphi)$ indicative of the scale and spatial correlation length, respectively. In the examples below we use an isotropic exponential 'correlation' structure for the entries in this matrix. This entails that $\Sigma_{\boldsymbol{\theta}}(i, j) = \sigma^2 \exp(-|s_i - s_j|/\varphi)$, where $|s_i - s_j|$ is the euclidean distance between sites $s_i$ and $s_j$. The

skewness parameter $\boldsymbol{\lambda} = \lambda_0 \mathbf{1}$ is a length $n$ vector. We define $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\theta}, \lambda_0)$. From equation (1) we have

$$\pi(\boldsymbol{x}|\boldsymbol{\eta}) = \frac{2}{(2\pi)^{n/2}|\Sigma_\theta|^{1/2}} \exp\left(-\frac{1}{2}(\boldsymbol{x} - H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\boldsymbol{x} - H\boldsymbol{\beta})\right) \cdot \Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\boldsymbol{x} - H\boldsymbol{\beta})).$$

We consider the situation where sites $\{s_1, \cdots, s_k\}$ are observation sites, while one of our goals is to predict the spatial latent variables at the unobserved sites $\{s_{k+1}, \cdots, s_n\}$. The latent variables at the $k$ observed sites are denoted $\boldsymbol{x}^{obs} = A\boldsymbol{x}$, $A = [I_{k\times k} \mid 0_{k\times n-k}]$. Thus, $A$ decomposes vector $\boldsymbol{x}$ into $\boldsymbol{x} = (\boldsymbol{x}^{obs'}, \boldsymbol{x}^{pred'})'$, where $\boldsymbol{x}^{pred}$ are latent variables at $(n-k)$ chosen prediction sites.

The likelihood part of the model links data to the latent variables at the $k$ observation sites. Let $\boldsymbol{y}' = (y_1, \cdots, y_k)$ represent the discrete spatial response variables at the observation sites $\{s_1, \cdots, s_k\}$. We assume that the measurements are conditionally independent with likelihood $\pi(\boldsymbol{y}|\boldsymbol{x}^{obs})$ of an exponential family (McCullagh and Nelder (1989)). For each observation site $i = 1, \ldots, k$, this can be written $\pi(y_i|x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}$, where $b(x_i)$ is the cumulant function, such that $\frac{\partial b(x_i)}{\partial x_i} = E(y_i|x_i)$, $\frac{\partial^2 b(x_i)}{\partial x_i^2} = var(y_i|x_i)$. For example, with binomial likelihood $b(x_i) = u_i \log(1+\exp(x_i))$, where $u_i$ is the number of trials. The mean $E(y_i|x_i)$ and $x_i$ are in general related by $E(y_i|x_i) = g^{-1}(x_i)$, where $g(\cdot)$ is the known link function.

The last part of the hierarchical model is the prior for model parameters $\boldsymbol{\eta}$. We use proper priors for all these parameters, and denote this prior by $\pi(\boldsymbol{\eta})$. For the regression parameters we set $\boldsymbol{\beta} \sim N(\boldsymbol{a}, B)$, with large variance terms in $B$. For the scale parameter $\sigma$ we use an inverse gamma prior, while the spatial 'correlation' parameter $\phi$ is assigned a gamma prior. Finally, the skewness parameter $\lambda_0$ is assigned a normal prior. The specific parameters are set in the examples below. Using the normal prior for the regression parameters $\boldsymbol{\beta}$, one can in fact marginalize over these parameters in the SN prior for latent variables. Since the SN distribution is a special case of CSN, we have $(\boldsymbol{x}|\boldsymbol{\beta}, \boldsymbol{\theta}, \lambda_0) \sim CSN_{n,1}(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}, 0, 1)$, and when $\boldsymbol{\beta} \sim N(\boldsymbol{a}, B)$ we get

$$(\boldsymbol{x}|\boldsymbol{\theta}, \lambda_0) \sim CSN_{n,1}(\boldsymbol{\mu}_x, \Sigma_x, D_x, \boldsymbol{\nu}_x, \Delta_x),$$

where $\boldsymbol{\mu}_x = H\boldsymbol{a}$, $\Sigma_x = (\Sigma_\theta + HBH')$, $D_x = \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}}\Sigma_x^{-1}$, $\boldsymbol{\nu}_x = 0$, $\Delta_x = 1 + \boldsymbol{\lambda}'\boldsymbol{\lambda} - \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}}\Sigma_x^{-1}\Sigma_\theta^{\frac{1}{2}}\boldsymbol{\lambda}$. See Appendix. In our examples below the regression parameter $\beta_0$, corresponding to the constant term, is integrated out using this result. The other parameters $\beta_1, \ldots, \beta_p$ could be of explanatory interest, and are kept as part of the parameter vector $\boldsymbol{\eta}$.

To summarize, the model has the following components:

$$\begin{aligned}\pi(\boldsymbol{y},\boldsymbol{x},\boldsymbol{\eta}) &= \pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta}) \\ &= \frac{2}{(2\pi)^{n/2}|\Sigma_\theta|^{1/2}}\exp\left(\sum_{i=1}^{k}[y_i x_i - b(x_i) + c(y_i)] - \frac{1}{2}(\boldsymbol{x}-H\boldsymbol{\beta})'\Sigma_\theta{}^{-1}(\boldsymbol{x}-H\boldsymbol{\beta})\right) \\ &\quad \times\Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\boldsymbol{x}-H\boldsymbol{\beta}))\pi(\boldsymbol{\eta}), \end{aligned} \tag{7}$$

Our main interest is in the posterior distribution for model parameters $\boldsymbol{\eta}$, and in predicting latent variables $x_j$ at unobserved sites, or perhaps prospective data $y_j$ at unobserved sites. The posteriors are not directly available, and in the next Section we present a methodology for approximating the posterior densities.

## 3. Approximation Bayesian Inference and prediction

The main methodological idea is to use a CSN approximation for the conditional density $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$, and next use this to obtain approximate inference for parameters and spatial prediction. Recall from Section 2 that the CSN distribution is closed under linear transformation and conditioning. We use these properties to construct an approximation for the full conditional $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$. This is done by linearization of the full conditional in an iterative manner. From equation (7), we can write

$$\begin{aligned}\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) &\propto \exp\{-\frac{1}{2}\boldsymbol{x}'\Sigma_\theta^{-1}\boldsymbol{x} + \boldsymbol{x}'\Sigma_\theta^{-1}H\boldsymbol{\beta} + \sum_{i=1}^{k}[y_i x_i - b(x_i)]\}\cdot\Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\boldsymbol{x}-H\boldsymbol{\beta})) \\ &\approx \exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_{x|y,\eta})'\Sigma_{x|y,\eta}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}_{x|y,\eta})\}\cdot\Phi(\boldsymbol{\lambda}'^{*}\Sigma_{x|y,\eta}^{-\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{\mu}_{x|y,\eta}), \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(H\boldsymbol{\beta}-\boldsymbol{\mu}_{x|y,\eta}), 1),\end{aligned}$$

where $\boldsymbol{\lambda}'^{*} = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}\Sigma_{x|y,\eta}^{\frac{1}{2}}$. Then $(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) \approx CSN_{n,1}(\boldsymbol{\mu}_{x|y,\eta}, \Sigma_{x|y,\eta}, D_{x|y,\eta}, \boldsymbol{\nu}_{x|y,\eta}, 1)$, where $\boldsymbol{\mu}_{x|y,\eta} = H\boldsymbol{\beta} + \Sigma_\theta A'R^{-1}(z(\boldsymbol{y},\boldsymbol{x}^{obs}) - AH\boldsymbol{\beta})$, and $z_i(y_i, x_i) = [y_i - b'(x_i) + x_i b''(x_i)]/b''(x_i)$, $i = 1,\ldots,k$, is a linearization of the exponential family likelihood part of equation (7) at a fixed value of $\boldsymbol{x}$. Moreover, $R = A\Sigma_\theta A' + P$ and $P = P(\boldsymbol{x})$ is a diagonal matrix with entries element $P(i,i) = 1/b''(x_i)$, $i = 1,\cdots,k$. Finally, $\Sigma_{x|y,\eta} = \Sigma_\theta - \Sigma_\theta A'R^{-1}A\Sigma_\theta$, $D_{x|y,\eta} = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}$, and $\boldsymbol{\nu}_{x|y,\eta} = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(H\boldsymbol{\beta} - \boldsymbol{\mu}_{x|y,\eta})$. See Appendix for further explanation. Note that $\boldsymbol{\mu}_{x|y,\eta}$, $\Sigma_{x|y,\eta}$ and $\boldsymbol{\nu}_{x|y,\eta}$ are all functions of data $\boldsymbol{y}$ or the fixed linearization point through $z(\boldsymbol{y},\boldsymbol{x}^{obs})$ and $P$. We remark that if the data $\boldsymbol{y}$ are continuous with a linear and normal distributed likelihood function, the CSN expression for the full conditional is exact.

We now present the iterative algorithm for fitting a CSN approximation in more detail. At each step of the iteration $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ is updated using the method described above. The algorithm goes as follows:

1. Choose starting value $\boldsymbol{x}^{(0)}$, set $m = 0$.
2. Calculate $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) = CSN_{n,1}(\hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}), \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}), D_{x|y,\eta}, \hat{\boldsymbol{\nu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}), 1)$.
3. Let $\boldsymbol{x}^{(m+1)} = \hat{\mathrm{E}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) = \hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}) + \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)})D'_{x|y,\eta}\hat{\psi}$,

$$\hat{\psi} = \psi(0, \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}), D_{x|y,\eta}, \hat{\nu}_{x|y,\eta}(\boldsymbol{x}^{(m)}), 1)$$

4. Set $m = m + 1$. Go to 2.
5. Convergence is obtained after a few iterations and an approximation of $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ is then

$$\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) = CSN_{n,1}(\hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m+1)}), \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m+1)}), D_{x|y,\eta}, \hat{\boldsymbol{\nu}}_{x|y,\eta}(\boldsymbol{x}^{(m+1)}), 1).$$

We illustrate the approximation with a simple graphical example. Let the prior for $\boldsymbol{x}$ be SN with

—

Figure 2 about here.

—

fixed location parameter $\boldsymbol{\mu} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, scale $\Sigma = I_2$, and skewness $\boldsymbol{\lambda} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}$. Let further $\boldsymbol{y}' = (2\ \ 5)$ be two binomial observations with trials $u_i = 10,\ i = 1, 2$. Figure 2 Left) shows a contour plot of the exact $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ when we use $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) \propto \pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{y}|\boldsymbol{x})$, while Figure 2 Middle) shows a contour plot of the approximation $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$. In both these displays the densities are evaluated on a dense grid of $\boldsymbol{x}$ values. Figure 2 Right) shows the difference between the exact and the approximate solution, i.e. $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) - \hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$. There is not much difference between the two, and it seems like the CSN approximation is very good in this case. The slight differences in Figure 2 Right) are mainly along the skewness direction, away from the mean, with underestimation in the positive direction and overestimation in the negative direction. In the examples below we document the quality of the approximation on spatial data.

### 3.1. Approximate Parametric Inference

We next show methods for using the approximate full conditional $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ to construct an approximation for $\pi(\boldsymbol{\eta}|\boldsymbol{y})$. The posterior marginal for model parameters $\boldsymbol{\eta}$ is

$$\pi(\boldsymbol{\eta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})}{\pi(\boldsymbol{y})\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})} \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})}{\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})},$$

which is valid for any value of $\boldsymbol{x}$. For example $\boldsymbol{x} = \hat{\mathrm{E}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$, the mean of the approximate CSN posterior $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$, for fixed $\boldsymbol{\eta}$. By using the fitted CSN approximation to the full conditional $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$, we can write an approximate posterior marginal density for the model parameters as

$$\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})}{\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})}\Big|_{\boldsymbol{x}=\hat{\mathrm{E}}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})}. \tag{8}$$

This expression is similar to the Laplace approximation of Tierney and Kadane (1986). They employed the Laplace approximation to approximate the posterior marginal density of individual parameters in a multiparameter setting and showed that the error in the approximation $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$ is of a functional form of order $O(n^{-\frac{3}{2}})$. Eidsvik *et al.* (2009) used this Laplace approximation to the posterior marginals for model parameters in SGLMM with normal spatial latent variables. Rue and Martino (2007) and Rue *et al.* (2009) used the Laplace approximation with a Gaussian Markov random field for spatial latent variables. Our formulation is different since we use a CSN approximation, and not the normal approximation which is standard in Laplace's approximation. Still, the approximation bears some resemblence to that of the Laplace approximation used in Rue and Martino (2007), Eidsvik *et al.* (2009) and Rue *et al.* (2009), since the CSN is of a conjugate form and used both as prior for spatial latent variables, and as an approximation to the full conditional for these latent variables. Another difference is that we compute the approximation at the mean and not the mode, see also Hsiao *et al.* (2003). One would intuitively presume that the approximation $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$ is good if the inserted approximation for the full conditional, in the denominator of equation (8), is sufficiently accurate. Similar ideas, albeit sampling based, have been used to approximate the marginal likelihood from MCMC realizations, see e.g. Chib (1995). These sampling based approximations are exact in the limit, as the number of MCMC realizations go to infinity.

In practice we must evaluate $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$ numerically on a grid of $\boldsymbol{\eta}$ values, or fit a parametric form to the density. Here, we adopt the numerical scheme of Rue *et al.* (2009). This entails the following

steps: i) Optimize $\log \hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$, and denote the argument at the mode by $\boldsymbol{\eta}^*$. ii) Compute $\Sigma = H^{-1}$ at $\boldsymbol{\eta}^*$, where $H$ is the Hessian matrix. iii) Find the eigen-decomposition of $\Sigma = V\Lambda V'$, where $V$ is eigenvectors and $\Lambda$ is a diagonal eigenvalue matrix. iv) Use a stepping out procedure away from $\boldsymbol{\eta}^*$, oriented according to the grid spanned by the eigenvectors. The stepping out procedure requires tuning parameters which are the stepping length and a truncation factor indicating when the density becomes negligible relative to the density at the mode. In our examples below we always have two parameters in the spatial interaction matrix $\Sigma_{\boldsymbol{\theta}}$, one parameter for skewness $\lambda_0$, and sometimes one regression coefficient $\beta_1$. Then the dimension of $\boldsymbol{\eta}$ is $3-4$, and the above numerical procedure is fast. For visualization of the approximate posterior marginals for each model parameter, one can apply interpolation on top of the numerical scheme.

### *3.2. Approximate Bayesian Spatial Prediction*

One of the goals of SGLMM is to predict the spatial latent variables, $\boldsymbol{x}^{pred}$, at unobserved sites. The approximate CSN distribution for the full conditional again plays a key role. Recall from Section 2 that when the joint $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ is CSN, the marginal distribution of one variable, denoted $\hat{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta})$, is also within the CSN class. Spatial prediction is done weighting these CSN densities with the posterior marginals for parameters $\boldsymbol{\eta}$. The approximate predictive marginals are then

$$\hat{\pi}(x_j|\boldsymbol{y}) = \sum_{\ell} \hat{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta}_\ell)\cdot\hat{\pi}(\boldsymbol{\eta}_\ell|\boldsymbol{y}), \quad j = 1,\cdots,n,$$

where the sum is over the numerical evaluation points of $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$. This expression for the posterior marginal is a mixture of CSN distributions. We are particularly interested in $j = k+1,\ldots,n$, which index the latent variables at unobserved sites $\boldsymbol{x}^{pred}$.

Following Eidsvik *et al.* (2009) and Rue *et al.* (2009) we can refine the predictive marginal for $x_j$ using what Rue *et al.* (2009) refer to as the integrated nested Laplace approximation. In our case we extend their approach by fitting CSN distributions to obtain the improved marginals, rather than normal densities. Nevertheless, the form of the nested approximations remains the same. The improved approximation for predictive marginals is denoted $\tilde{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta})$. If the improved version is indistinguishable from the direct approximation $\hat{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta})$, this can be used to verify the direct one. The improved version is

$$\tilde{\pi}(x_j|\boldsymbol{y},\boldsymbol{\eta}) \propto \frac{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})}{\tilde{\pi}(\boldsymbol{x}_{-j}|x_j,\boldsymbol{y},\boldsymbol{\eta})}\Big|_{\boldsymbol{x}=\hat{\mathrm{E}}(\boldsymbol{x}_{-j}|x_j,\boldsymbol{y},\boldsymbol{\eta})}, \tag{9}$$

where $\tilde{\pi}(\boldsymbol{x}_{-j}|x_j, \boldsymbol{y}, \boldsymbol{\eta})$ is a CSN distribution, approximated and evaluated for fixed $x_j$. The denominator is a CSN distribution because we linearize the likelihood and because of the conditioning property of the CSN distribution remarked in Section 2. See also Appendix. The improved posterior marginal for latent variables requires that we fix $x_j$ at different values (on a grid), and for each of these $x_j$'s we approximate a new CSN density in the denominator. This scheme constructs a numerical approximation to $\tilde{\pi}(x_j|\boldsymbol{y}, \boldsymbol{\eta})$, which is more reliable than the direct one, but more time consuming. To obtain the marginal $\tilde{\pi}(x_j|\boldsymbol{y})$ one must carry out this procedure for all $\boldsymbol{\eta}_\ell$ parameters, and weight the resulting densities according to $\hat{\pi}(\boldsymbol{\eta}_\ell|\boldsymbol{y})$.

One might also predict data values $y_j$ at unobserved sites. For a prediction site, $j = k+1, \ldots, n$, we have

$$\pi(y_j|\boldsymbol{y}) = \int_{x_j} \pi(y_j|x_j)\pi(x_j|\boldsymbol{y})dx_j. \tag{10}$$

We get a direct or improved approximation by inserting either $\hat{\pi}(x_j|\boldsymbol{y})$ or $\tilde{\pi}(x_j|\boldsymbol{y})$ in the integrand, respectively. In practice we further approximate the integral with a sum over a grid of $x_j$ values. Equation (10) must be computed for all $y_j$ values in the discrete sample space of the data. Eidsvik *et al.* (2009) used this approximation for detecting outliers. In an example below we use the relation for comparing the two assumptions of normal or skew normal models for spatial latent variables.

### *3.3. Approximations as proposal distributions in MCMC algorithms*

The approximations $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$ and $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$ described above can be used as proposal distributions in a MCMC algorithm. Note that the MCMC algorithm is then run after the approximations have been constructed. The computational time of the MCMC algorithm is large compared with the direct numerical approach that we outline above. Nevertheless, there are advantages of MCMC sampling. For instance, the results are exact in the limit, as the number of samples go to infinity. Further, MCMC sampling allows one to assess any joint probability for any non-linear function of variables, by simply computing those non-linear functions for each MCMC sample. For this paper we use MCMC sampling to check the quality of our proposed approximations for posterior marginals for parameters and predictions of the latent variables.

We use an independent proposal Metropolis–Hastings (MH) algorithm as follows: (i) propose $\boldsymbol{\eta}^*$ from the numerical approximation $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$; (ii) propose $\boldsymbol{x}^*$ from the CSN approximation

$\hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta}^*)$; (iii) accept them jointly, otherwise keep the current values. The acceptence rate equals:

$$\min\left(1, \frac{\pi(\boldsymbol{y}|\boldsymbol{x}^*)\pi(\boldsymbol{x}^*|\boldsymbol{\eta}^*)\pi(\boldsymbol{\eta}^*)}{\pi(\boldsymbol{y}|\boldsymbol{x})\pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{\eta})} \frac{\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})\hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})}{\hat{\pi}(\boldsymbol{\eta}^*|\boldsymbol{y})\hat{\pi}(\boldsymbol{x}^*|\boldsymbol{y}, \boldsymbol{\eta}^*)}\right),$$

where $\boldsymbol{x}$ and $\boldsymbol{\eta}$ denote the current values of the Markov chain. Step ii) of the MH scheme above is done using properties of the CSN distribution described in Section 2. The presented independent proposal algorithm can also be used in combination with another MH algorithm, for instance random walk, to avoid getting stuck for very many iterations. We did not experience this problem to a large extent in our examples, and used only the independent proposal. One can similarly run a MH algorithm for fixed parameters $\boldsymbol{\eta}$, drawing the proposals from $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$.

## 4. Examples

In this Section we study two examples with (non-lattice) discrete data at observation sites over a continuous domain. The first is on precipitation data, using a binomial likelihood with $u_i = 14$ trials. The second is on trout 'population status' data in relation to acid pollution. This is modeled with a Bernoulli likelihood. For both examples we compare the accuracy of the proposed approximate Bayesian methods. We also campare results obtained with a normal or SN model for the latent variables.

—

Figure 3 about here.

—

### *4.1. Precipitation in Middle Norway*

Eidsvik *et al.* (2009) studied rainfall data from Fall 2006 in the middle part of Norway. They used a SGLMM for the dataset, with a normal prior model for the spatial latent variables. We regard the same data, but with a SN prior for the latent variables. The data are number of days with rain out of the number of days in operation ($u_i = 14$, two weeks), observed for $i = 1, \cdots, 92$ observation sites. Figure 3 shows the precipitation data on the map of middle Norway. The counts are assumed to be conditionally independent Binomial variables, hence $\pi(y_i|x_i) = \exp\{y_i x_i - u_i \log(1 + \exp(x_i))\}$. Following Eidsvik *et al.* (2009) we use a scale matrix based on an isotropic exponential covariance function for the spatial latent variable. The scale parameter $\sigma$ is assigned an inverse gamma prior

model with mean 0.5 and variance 1, and the correlation range parameter $\phi$ is assigned a gamma prior with mean 40 km and variance 200 km$^2$. The skewness parameter $\lambda_0$ is assigned a normal prior with mean 3 and variance 5. There are no covariates in the model, and the constant term $\beta_0$, which has normal prior with mean 0.8 and variance 1, is integrated out analytically (Appendix).

The numerical scheme for assessing $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$ ended up with around 2000 evaluation points in the three dimensional space spanned by $\sigma$, $\phi$ and $\lambda_0$. To construct approximate posterior marginals $\hat{\pi}(\sigma|\boldsymbol{y})$, $\hat{\pi}(\phi|\boldsymbol{y})$ and $\hat{\pi}(\lambda_0|\boldsymbol{y})$ we created an interpolated surface on a three dimensional grid oriented along each variable, and then summed out the appropriate directions. Figure 4 Left) shows the resulting posterior marginals (solid line marked SN). For the scale and correlation parameter there is much information in the data, since the posterior distributions are shifted from the prior mean and much narrower than the priors. For the skewness parameter $\lambda_0$ (Figure 4, Left, below)) the posterior has a mode about 3 and variance about 3, which is only a little narrower than the prior. Thus, the data do not carry much information about this parameter. We tested with other priors for $\lambda_0$, and the posterior usually has the same mode as the prior, but the density is a little narrower. The effects of the different priors we tested for $\lambda_0$ were not noticeable on the predictions for latent variable. The dashed-dotted lines falling close to the solid lines in Figure 4 Left) are computed from a very long MCMC run using the MH scheme described in Section 3.3. We note that the results of approximate Bayesian inference and MCMC sampling are almost identical for these posterior marginals. The fluctuations in the lines are caused by Monte Carlo error, or interpolation of the evaluation points. These fluctuations are larger than the difference between the results of approximate Bayesian inference and MCMC sampling. The acceptance rate of the MCMC sampler was about 80 percent.

We also applied the same scheme for a normal prior model, which entails treating $\lambda_0$ fixed at 0. Figure 4 Left) shows the posterior marginals for $\sigma$ and $\phi$ in this case (solid line marked Normal). Compared with the SN prior model, we notice that the standard deviation is larger for normal prior, while the correlation length is smaller for normal. Thus, when we account for the skewness using SN, we obtain more spatial smoothness in the model. The results of normal prior are also compared with MCMC simulations (dashed-dotted). Approximate Bayesian inference and MCMC simulations give almost identical results. The acceptance rate of the MCMC sampler was

about 80 percent. One interpretation of this smaller acceptance rate for normal, compared with SN, is that there is more smoothness (higher $\phi$) in the SN model, and this implies that the effective dimension of the spatial variable is smaller for SN prior model.

In Figure 4 Right) we show predictive distributions at an unobserved site at longitude 10.3 and latitude 63.3. In this case the prediction is done fixing the $\boldsymbol{\eta}$ parameters at the mode of $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$. The parameter values at the mode are $\hat{\boldsymbol{\eta}}^{SN} = (0.26, 51.2, 3.2)$ for SN and $\hat{\boldsymbol{\eta}}^{N} = (0.35, 35)$ for normal. The prediction is done for SN prior model (Figure 4 Right, top)) and normal prior model (Figure 4 Right, below)). In both situations we compute three approximation: direct $\hat{\pi}(x_j|\boldsymbol{y}, \hat{\boldsymbol{\eta}})$, improved $\tilde{\pi}(x_j|\boldsymbol{y}, \hat{\boldsymbol{\eta}})$, and MCMC sampling $\hat{\pi}^{MCMC}(x_j|\boldsymbol{y}, \hat{\boldsymbol{\eta}})$. For both models the differences between the three approximations are very small. Direct inference is slightly biased to the left for both models, compared with improved and MCMC results, but the effect is hardly noticeable. The improved approximation and MCMC results are indistinguishable. The predictions are centered around 1.2 for both models, but the SN case is narrower than the normal result. The approximations for the marginal prediction $\pi(x_j|\boldsymbol{y})$ are even more similar than those displayed in Figure 4 Right). The same holds for other predicition sites we looked at. For this dataset the results of normal or SN prior model are near identical. For both normal and SN model the approximate Bayesian prediciton results are near identical to that of MCMC sampling. The computation time of the direct approximation is seconds, the improved approximation takes minutes, while the MCMC algorithm runs for several hours.

—

Figure 4 about here.

—

—

Figure 5 about here.

—

### *4.2. Acidification in Norwegian lakes*

Lake acidification is defined as the process in which annual mean lake acidity changes from circumneutral PH to values less than 5.6. This bounding value is based on the fact that pure water (PH 7) can acidify naturally to PH 5.6 in equilibrium with atmospheric carbon dioxide (Dobson *et al.*

(1990)). The acidification process occurs over time, and the lake water stabilizes at a certain acidity, maintaining a small number of species of plants and animals, but usually lacking many fish. For example, the trout populations are very sensitive to acidification.

Varin *et al.* (2005) studied trout and acidification data made in 542 Norwegian lakes. They modeled these data with a SGLMM, using a normality assumption for the latent variables. The data were collected during 1986 from interviews with local fishermen. The response is 'population status' of trout for each lake $i = 1, \ldots, 542$, coded as unaffected ($y_i = 0$) or decreased/extinct ($y_i = 1$). Data are shown in Figure 5. In this example we split the data in two parts. A training set of size 400 is picked among the 542 sites, while the remaining 142 sites are used for cross-validation. The 400 sites in the training set are picked at random, and a map of only these sites shows a similar pattern of sites and of 0 / 1 responses (filled, open circles) as in Figure 5. The measured Acid Neutralizing Capacity (ANC) is used as a covariate. ANC is a measure for the overall buffering capacity against acidification for a solution. The data are assumed to be conditionally independent Bernoulli variable, i.e. $\pi(y_i|x_i) = \exp\{y_i x_i - \log(1 + \exp(x_i))\}$.

We reanalyze these data using a SN prior model for latent variables, i.e. $\pi(\boldsymbol{x}|\boldsymbol{\eta}) = SN(\beta_0 + \beta_1 ANC, \Sigma_\theta, \boldsymbol{\lambda})$, where $\beta_0$, $\beta_1$, are regression parameters, $\boldsymbol{\lambda} = \lambda_0 \mathbf{1}$ skewness parameter, and $\Sigma_\theta$ the scale matrix with a spatial correlation structure. We again use an isotropic exponential covariance model with parameters $(\phi, \sigma)$ for 'correlation range' and scale, respectively. The prior for $\beta_0$ is N(0.5, 1), and this parameter is integrated out analytically. For the remaining four parameters we use the following priors: $\beta_1$ is normal with mean $-2$ and variance 5, $\lambda_0$ is normal with mean 3 and variance 5, $\sigma$ is inverse gamma with mean 2 and variance 1, and $\phi$ is gamma with mean 250 and variance 2500.

Figure 6 Left) shows posterior marginals for $\boldsymbol{\eta} = (\sigma, \phi, \beta_1, \lambda_0)$. The solid lines are made using direct approximate Bayesian inference, while the dashed-dotted lines are based on MCMC samples. Differences are very small, possibly due to Monte Carlo error or interpolation routines used for visualizing marginals. The acceptance rate for the MCMC sampler was about 90 percent in this case. Figure 6 Right) shows similar posterior marginals when we use a normal prior for the spatial latent variable, where the skewness parameter $\lambda_0 = 0$. Results of direct inference and MCMC sampling are very similar. When comparing the plots for SN and the normal prior models, we note

that the scale $\sigma$ is about the same, while the correlation range parameter $\varphi$ is somewhat larger for SN. The range seem better determined in the normal case with a narrower posterior. The regression parameter $\beta_1$ for the ANC covariate has its posterior mass between $-6$ and $-3.5$ for both models, and this indicates that there is a clear connection between this covariate and the 'population status' of trout. Note that the ANC covariate is normalized in our case, when renormalized our estimates are comparable with those in Varin *et al.* (2005). For the scale and the correlation range we obtain the same as they do. For the skewness parameter in the SN model, the posterior is only a little narrower than that from the prior specification, indicating that the data do not carry much information about this parameter.

In Figure 7 Left) we show predictive densities for an unobserved site at longitude 24.2 and latitude 69.9. These predictive densities are computed for model parameter $\boldsymbol{\eta}$ fixed at the approximate posterior mode, for both normal and SN prior model, and for direct (solid), improved (+) and MCMC sampling (dashed-dotted) approximations. The SN prediction has a mode at about 5.4, while the normal case has a mode at about 5.6. In each subplot the direct, improved and MCMC sampling approximations are indistinguishable. The acceptance rates of the MCMC algorithms, treating $\boldsymbol{\eta}$ fixed, was about 90 percent for both normal and SN prior model. Figure 7 Right) shows predictive marginals $\hat{\pi}(x_j|\boldsymbol{y})$ (solid), $\tilde{\pi}(x_j|\boldsymbol{y})$ (+) and $\pi^{MCMC}(x_j|\boldsymbol{y})$ (dashed-dotted) at the same site. These densities are wider than the ones for fixed model parameters in the left display, a natural consequence when we account for the uncertainty in model parameters. In this marginal predictive case the SN results are again narrower and to the left compared with the predictions based on a normal prior. The three approximations are very similar in both left and right displays. An exception is the approximation based on MCMC sampling in the SN marginal prediction (Figure 7, right, top). This seems to be a consequence of poor mixing in the MCMC algorithm, which was not long enough in this case. The computation time of the direct approximation is seconds, the improved approximation takes minutes, while the MCMC algorithm was run for several hours.

Recall that 142 of the original observations were left out of the analysis, and will now be used for cross validation. From the data we know that 66 of these 142 sites represent unaffected trout ($y_j = 0$), the remaining 76 represent decreased/extinct trout ($y_j = 1$). By using the approximation in equation (10) we compute $\hat{\pi}(0|\boldsymbol{y})$ and $\hat{\pi}(1|\boldsymbol{y})$ at all 142 prediction sites, for normal and SN prior

model. We consider a classification rule as follows

$$\hat{y}_j = \begin{cases} 0 & \hat{\pi}(0|\boldsymbol{y}) > \hat{\pi}(1|\boldsymbol{y}) \\ 1 & \hat{\pi}(1|\boldsymbol{y}) > \hat{\pi}(0|\boldsymbol{y}) \end{cases}, \quad j = 401, \cdots, 542. \tag{11}$$

With this classification rule we obtain 60 sites predicted as unaffected for the SN model and 59 for normal model. For SN prior there are 10 sites predicted $\hat{y}_j = 1$, while observed $y_j = 0$. On the other hand there are 4 sited where SN prediction is $\hat{y}_j = 0$, while observed $y_j = 1$. Differences between predicted and observed values typically occur when $\hat{\pi}(1|\boldsymbol{y})$ is near 0.5. Both SN and normal model overestimate $\pi(y_j = 1|\boldsymbol{y})$ slightly on average, but SN is a little bit closer to the true proportions observed in the data. SN and normal model classifications disagree only at one prediction site. The predicted probabilities $\hat{\pi}(y_j|\boldsymbol{y})$ look almost the same for SN and normal model, but show clear differences at some sites. These sites are the ones where the skewness parameter kicks in for the CSN posteriors, given partly by the prior, partly by the data. The moderate differences between SN and normal predictions have been noted before; for instance in Kim and Mallick (2004) who displayed predictions for both models, and in Liu and Dey (2008) where they presented, amongst others, the Deviance Information Criteria for both models with only slight differences.

—

Figure 6 about here.

—

—

Figure 7 about here.

—

## 5. Conclusion

We have considered a SN prior model for the spatial latent variables in SGLMM. This model is more general than the usual one based on a normal prior, since it allows a parameter for skewness. The marginal contribution of our paper is to propose methods for approximate Bayesian inference and prediction for SGLMM with this SN prior. Our methods for inference and spatial prediction extend the approach of Eidsvik *et al.* (2009) to SN prior models. To achieve the approximations we use the CSN distribution as an important building block. The CSN distribution is a generalization of the SN, which is closed under linear conditioning and marginalization. Based on our

CSN approximation we propose direct and improved approximate prediction densities, which are compared with long MCMC runs. The direct approximation takes seconds of computer time, the improved takes minutes, while the MCMC algorithm takes several hours. The three approximations are indistinguishable for marginal inference for model parameters and spatial predictions for the datasets we studied.

We also compared results obtained with the SN prior with that of a normal prior. The results of parameter estimates show that the 'correlation range' is larger using the SN prior model, while the scale is reduced for SN. The spatial data carry much information about the regression parameters and the scale parameter, some about the 'correlation range', only limited information about the skewness parameter. For spatial prediction we notice very moderate differences between the SN and normal priors. In our examples the predictive densities with SN prior were a little narrower than using normal. For some sites, possibly directions of the multivariate spatial dimensions where the skewness parameter kicks in, the predictive densities were a little shifted compared with normal prior, but not very much.

We only considered the SN prior model. Our methods could be extended to a CSN prior model ($q > 1$), which entails more directions for the skewness in a multivariate setting. This certainly allows more flexible models, but one caveat with this approach is that the skewness parameters seem hard to estimate. There is a trade off between flexibility and identification. Nevertheless, prior information from external sources, or added structure in the model, could provide more realism in specific examples.

## 6. Acknowledgment

The authors thank the Norwegian Institute of Water Research for the trout population data. We also thank Prof. Dominguez-Molina for his help on the closed skew normal distributions. FH thanks the Department of Mathematical Sciences for support September 2008 - February 2009.

## References


Ainsworth, L.M. & Dean, C.B. (2006). Approximate Inference for Disease Mapping, *Computational statistics and data analysis,* **50**, 2552-2570.

Arellano-Valle, R. B., Bolfarine, H. & Lachosv V. H., (2005). Skew-normal Linear mixed models, *Journal of data science,* **3**, 415-438.

Azzalini, A. (1985). A class of distributions which includes the normal one, *Scandinavian journal of statistics,* **12**, 171-178.

Azzalini, A. & Dalla Valle, A. (1996). The Multivariate skew-normal distribution, *Biometrika,* **83**, 715-726.

Breslow, N. E. & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the american statistical association*, **88**, 9-25.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the american statistical association,* **90** 1313-1321.

Christensen O. F. (2004). Monte Carlo maximum likelihood in model-based geostatistics, *Journal of computational and graphical statistics,* **13**, 702-718.

Christensen O. F. & Waagepetersen R. P. (2002). Bayesian prediction of spatial count data using generalized linear mixed models, *Biometrics,* **58**, 280-286.

Diggle, P., Tawn, J. A. & Moyeed, R. A. (1998). Model-based geostatistic, *Journal of the royal statistical society, Series C. Applied statistics,* **47**, 299-350.

Dobson, J.E., Rush, R. M. & Peplies, R. W. (1990). Forest blowdown and lake acidification, *Annals of the association of american geographers,* **80**, 343-361.

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.

Eidsvik, J., Martino, S. & Rue, H. (2009). Approximate bayesian inference in spatial generalized linear mixed models, *Scandinavian journal of statistics*, **36**, 1-22.

Gonzalez-Farias, G. & Dominguez-Molina, J. A. and Gupta, A. K. (2004). The closed skewnormal distribution. In: *Genton M. G., ed. Skew-elliptical distributions and their applications: A journey beyond normality. Boca Raton, FL: Chapman and Hall CRC,* 2542.

Gupta, A.K., Gonzales-Farias, G. & Dominguez-Molina, J.A. (2003). A multivariate skew normal distribution. *Journal of multivariate analysis,* **89**, 181-190.

Hsiao, C. K., Huang, S. Y. & Chang, C.W., (2004), Bayesian marginal inference via candidates formula, *Statistics and computing*, **14**, 59-66.

Kim, H. M. & Mallick, B. K. (2004). A Bayesian prediction using the skew Gaussian distribution. *Journal of statististical planning and inference,* **120**, 85-101.

Komarek, A. & Lesaffre, E. (2008). Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution, *Computational statistics and data analysis,* **52**, 3441-3458.

Liu J. & Dey D. K. (2008). Skew random effects in multilevel binomial models: an alternative to nonparametric approach, *Statistical modelling,* **8**, 221-241.

McCullagh, P. & Nelder, J. A. (1989). *Generalized linear models,* Chapman and Hall, London.

Rue, H. & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of statistical planning and inference,* **137**, 3177-3192.

Rue, H., Martino, S. & Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations, *Journal of the royal statistical society, Series B.* To Appear.

Tierney, L. & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association,* **81**, 82-86.

Varin, C., Høst, G. & Skare, Ø. (2005). Pairwise likelihood inference in spatial generalized linear mixed models, *Computational statistics and data analysis,* **49**, 1173-1191.

Zhang, H. (2002). On estimation and prediction for spatial generalized linear mixed models, *Biometrics,* **58**, 129-136.

Zhao, Y., Staudenmayer, J., Coull, B. A. & Wand, M. P. (2006). General design Bayesian generalized linear mixed models with applications to spatial statistics, *Statistical science,* **21**, 35-51.

Zhu, H., Gu, M. & Peterson, B. (2007). Maximum likelihood from spatial random effects models via the stochastic approximation expectation maximization algorithm, *Statistics and computing*, **17**, 163-177.

Address of Correspondence:

*Fatemeh Hosseini*

*Department of Statistics, Tarbiat Modares University*

*14115-175 Tehran*

*Iran*

———————-

*EMAIL: f.hosseini@modares.ac.ir*

*PHONE: 98 21 82883483*

## 7. Appendix

### *A.1. Conjugate closed skew normal posterior distribution*

Let $W \sim N_n(\mu, \Sigma)$, $E \sim N_q(0, \Delta)$ and $X_0 = -\nu + D(W - \mu) + E$ where $D_{q\times n}$ is an arbitrary matrix, $\mu \in R^n$, $\nu \in R^q$ and $\Delta_{q\times q} > 0$. The joint distribution of $W$ and $X_0$ is

$$\begin{pmatrix} W \\ X_0 \end{pmatrix} \sim N_{n+q}\left(\begin{pmatrix} \mu \\ -\nu \end{pmatrix}, \begin{pmatrix} \Sigma & \Sigma D' \\ D\Sigma & \Delta + D\Sigma D' \end{pmatrix}\right).$$

The conditional density of $(W|X_0 > 0)$ is $CSN_{n,q}(\mu, \Sigma, D, \nu, \Delta)$ (Dominguez-Molina *et al.* (2003)). Assume $Y = AW + \varepsilon$, $Y \in R^m$, where $A$ is a $m \times n$ matrix and $\varepsilon \sim N(0, P)$. Let $Z = (Y', W')'$, then the joint distribution of $Z$ and $X_0$ is

$$\begin{pmatrix} Z \\ X_0 \end{pmatrix} \sim N_{m+n+q}\left(\begin{pmatrix} \mu_z \\ -\nu \end{pmatrix}, \begin{pmatrix} \Sigma_z & A'^*\Sigma D' \\ D\Sigma A^* & \Delta + D\Sigma D' \end{pmatrix}\right),$$

where $\mu_z = \begin{pmatrix} A\mu \\ \mu \end{pmatrix}$, $\Sigma_z = \begin{pmatrix} A\Sigma A' + P & A\Sigma \\ \Sigma A' & \Sigma \end{pmatrix}$, $A^* = (A' \;\; I)$ is an $n \times (m + n)$ matrix. Thus, jointly $(Z|X_0 > 0) \sim CSN_{m+n,q}(\mu_z, \Sigma_z, D_z, \nu, \Delta_z)$, where $D_z = D\Sigma A^*\Sigma_z^{-1}$ and $\Delta_z = \Delta + D\Sigma D' - D_z\Sigma_z D_z'$. On the other hand we know that $\Sigma_z^{-1} = \begin{pmatrix} P^{-1} & -P^{-1}A\Sigma \\ A'P^{-1} & \Sigma^{-1} + A'P^{-1}A \end{pmatrix}$, therefore $D_z = (0_{q\times n} \;\; D_{q\times n})$, $D_z\Sigma_z D' = D\Sigma D$ and $\Delta_z = \Delta$. By using the property of the CSN distribution in equation (5), we have

$$W|Y, X_0 > 0 \sim CSN_{n,q}(\mu + \Sigma A'(A\Sigma A' + P)^{-1}(Y - A\mu), \Sigma - \Sigma A'(A\Sigma A' + P)^{-1}A\Sigma, D, \nu - D^*(Y - A\mu), \Delta), \quad (12)$$

where $D^* = D\Sigma A'(A\Sigma A' + P)^{-1}$.

Now, assume $\boldsymbol{x}|\boldsymbol{\eta} \sim SN_n(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda}') = CSN_{n,1}(H\boldsymbol{\beta}, \Sigma_\theta, D, 0, 1)$ where $D = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}$. Further, likelihood is $\pi(y_i|x_i) = \exp\{y_i x_i - b(x_i)\}$, and $A$ is defined in Section 2.2. By linearizing the likelihood part of equation (7) and using equation (12) we get

$$\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta} \sim CSN_{n,1}(\boldsymbol{\mu}_{x|y,\eta}, \Sigma_{x|y,\eta}, D_{x|y,\eta}, \boldsymbol{\nu}_{x|y,\eta}, 1),$$

where $\boldsymbol{\mu}_{x|y,\eta} = H\boldsymbol{\beta} + \Sigma_\theta A'(A\Sigma A' + P)^{-1}(z(\boldsymbol{y}, \boldsymbol{x}^{obs}) - AH\boldsymbol{\beta})$, with $z(y_i, x_i)$ and $P$ defined in Section 3. Also, $\Sigma_{x|y,\eta} = \Sigma_\theta - \Sigma_\theta A'(A\Sigma A' + P)^{-1}A\Sigma_\theta$, $D_{x|y,\eta} = D = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}$, and $\boldsymbol{\nu}_{x|y,\eta} = -\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}\Sigma_\theta A'(A\Sigma A' +$

$P)^{-1}(z(\boldsymbol{y}, \boldsymbol{x}^{obs}) - AH\boldsymbol{\beta}) = \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(H\boldsymbol{\beta} - \boldsymbol{\mu}_{x|y,\eta})$.

*A.2. Direct and improved approximations for* $\pi(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$

Let $\boldsymbol{x}^* = \begin{pmatrix} x_j \\ \boldsymbol{x}_{-j} \end{pmatrix} = A_j\boldsymbol{x}$, $j = 1, \cdots, n$, where $\boldsymbol{x}_{-j} = (x_1, \cdots, x_{j-1}, x_{j+1}, \cdots, x_n)$ and $A_j$ is an identity matrix such that its $j$th row is exchanged with the first row. Thus, $A_jA_j' = I$ and $A_j$ is an orthogonal matrix. By using the property of the CSN distribution in equation (3) we get

$$\boldsymbol{x}^*|\boldsymbol{y}, \boldsymbol{\eta} \sim CSN_{n,1}(\mu_{x^*|y,\eta}, \Sigma_{x^*|y,\eta}, D_{x^*|y,\eta}, \nu_{x|y,\eta}, \Delta_{x^*|y,\eta}), \tag{13}$$

where, $\mu_{x^*|y,\eta} = A_j\hat{\mu}_{x|y,\eta} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}$, $\Sigma_{x^*|y,\eta} = A_j\hat{\Sigma}_{x|y,\eta} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$, $D_{x^*|y,\eta} = \hat{D}_{x|y,\eta}\hat{\Sigma}_{x|y,\eta}A_j'(A_j\hat{\Sigma}_{x|y,\eta}A_j')^{-1} = \hat{D}_{x|y,\eta}A_j' = [D_1 \;\; D_2]$, $\Delta_{x^*|y,\eta} = \hat{\Delta}_{x|y,\eta} + \hat{D}_{x|y,\eta}\hat{\Sigma}_{x|y,\eta}\hat{D}'_{x|y,\eta} - \hat{D}_{x|y,\eta}\hat{\Sigma}_{x|y,\eta}A_j'(A_j\hat{\Sigma}_{x|y,\eta}A_j')^{-1}A_j\hat{\Sigma}_{x|y,\eta}\hat{D}'_{x|y,\eta} = 1$.
Therefore, by the property of CSN in equation (4)

$$x_j|\boldsymbol{y}, \boldsymbol{\eta} \sim CSN(\mu_1, \Sigma_{11}, D^*, \nu_{x|y,\eta}, \Delta^*)$$

where, $D^* = D_1 + D_2\Sigma_{21}\Sigma_{11}^{-1}$, $\Delta^* = 1 + D_2\Sigma_{22\cdot1}D_2'$, $\Sigma_{22\cdot1} = \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}$.
Also, by using equation (5) and (13) we get

$$\boldsymbol{x}_{-j}|x_j, \boldsymbol{y}, \boldsymbol{\eta} \sim CSN(\mu_2 + \Sigma_{21}\Sigma_{11}^{-1}(x_j - \mu_1), \Sigma_{22\cdot1}, D_2, \nu - D^*(x_j - \mu_1), \Delta_{x^*|y,\eta})$$

*A.3. Marginalizing over a normal prior for the regression parameters*

**Theorem.** Let $(\boldsymbol{x}|\boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}) \sim CSN_{n,1}(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}, 0, 1)$ and $\boldsymbol{\beta} \sim N(\boldsymbol{a}, B)$, such that $\boldsymbol{\beta}$ is a $p \times 1$ vector, then,

$$(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\lambda}) \sim CSN_{n,1}(\boldsymbol{\mu}_x, \Sigma_x, D_x, \boldsymbol{\nu}_x, \Delta_x),$$

where, $\boldsymbol{\mu}_x = H\boldsymbol{a}$, $\Sigma_x = (\Sigma_\theta + HBH')$, $D_x = \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}}\Sigma_x^{-1}$, $\boldsymbol{\nu}_x = 0$, $\Delta_x = 1 + \boldsymbol{\lambda}'\boldsymbol{\lambda} - \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}}\Sigma_x^{-1}\Sigma_\theta^{\frac{1}{2}}\boldsymbol{\lambda}$.

*Proof.* We have $\boldsymbol{x}|\boldsymbol{\beta} \sim SN(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda})$, equivalent to $CSN_{n,1}(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}, 0, 1)$. We can write

$$\begin{pmatrix} W \\ X_0 \end{pmatrix} | \boldsymbol{\beta} \sim N_{n+1}\left(\begin{pmatrix} H\boldsymbol{\beta} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_\theta & \Sigma_\theta^{\frac{1}{2}}\boldsymbol{\lambda} \\ \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}} & 1 + \boldsymbol{\lambda}\boldsymbol{\lambda}' \end{pmatrix}\right),$$

such that $\boldsymbol{x}|\boldsymbol{\beta} \overset{d}{=} W|(X_0 > 0|\boldsymbol{\beta})$. Now, assume $\boldsymbol{\beta} \sim N(\boldsymbol{a}, B)$, we have

$$\mathrm{E}\begin{pmatrix} W \\ X_0 \end{pmatrix} = \mathrm{E}\left[\mathrm{E}\left(\begin{pmatrix} W \\ X_0 \end{pmatrix} | \boldsymbol{\beta}\right)\right] = \mathrm{E}\begin{pmatrix} H\boldsymbol{\beta} \\ 0 \end{pmatrix} = \begin{pmatrix} H\boldsymbol{a} \\ 0 \end{pmatrix}$$

and

$$\mathrm{Var}\begin{pmatrix} W \\ X_0 \end{pmatrix} = \mathrm{E}\left[\mathrm{Var}\left(\begin{pmatrix} W \\ X_0 \end{pmatrix} \mid \boldsymbol{\beta}\right)\right] + \mathrm{Var}\left[\mathrm{E}\left(\begin{pmatrix} W \\ X_0 \end{pmatrix} \mid \boldsymbol{\beta}\right)\right] = \begin{pmatrix} \Sigma_\theta + HBH' & \Sigma_\theta^{\frac{1}{2}}\boldsymbol{\lambda} \\ \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}} & 1 + \boldsymbol{\lambda}\boldsymbol{\lambda}' \end{pmatrix}.$$

Let $\boldsymbol{\mu}_x = H\boldsymbol{a}$, $\boldsymbol{\Sigma}_x = \Sigma_\theta + HBH'$, then

$$\begin{pmatrix} W \\ X_0 \end{pmatrix} \sim N_{n+1}\left(\begin{pmatrix} H\boldsymbol{a} \\ 0 \end{pmatrix}, \begin{pmatrix} \Sigma_x & \Sigma_x D_x' \\ D_x\Sigma_x & \Delta_x + D_x\Sigma_x D_x' \end{pmatrix}\right),$$

where $D_x = \boldsymbol{\lambda}'\Sigma_\theta^{\frac{1}{2}}\Sigma_x^{-1}$ and $\Delta_x = 1 + \boldsymbol{\lambda}'\boldsymbol{\lambda} - D_x\Sigma_x D_x'$ which completes the proof.

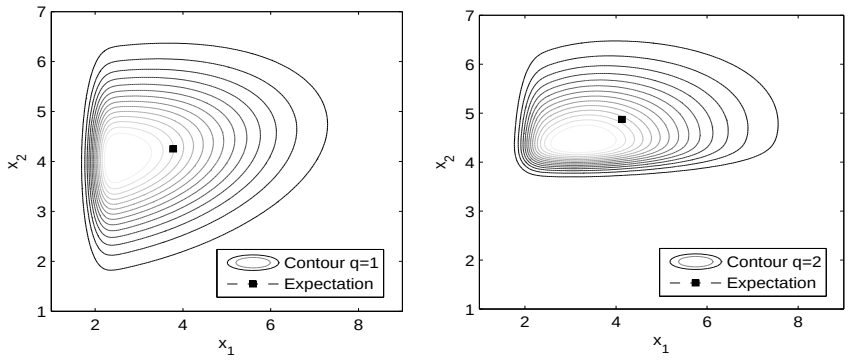


Figure 1: Left: contour plot of CSN density with parameters $\boldsymbol{\mu} = (2,4)'$, $\Sigma = [5\ \ 0.7; 0.7\ \ 1]$, $D = (5,0)$, $\nu = 0$, $\Delta = 1$. Right: contour plot of CSN density with parameters $\boldsymbol{\mu} = (2,4)'$, $\Sigma = [5\ \ 0.7; 0.7\ \ 1]$, $D = [5\ 0; 0\ 5]$, $\boldsymbol{\nu} = (0,0)'$, $\Delta = I_2$.

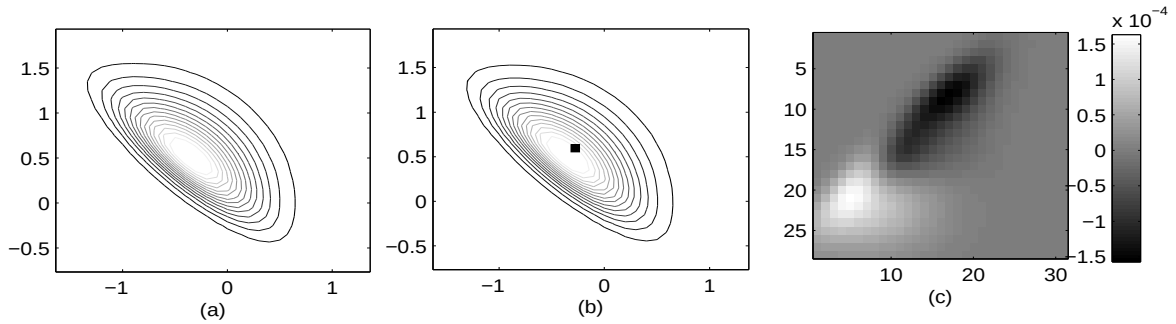


Figure 2: Left: Full conditional for $\boldsymbol{x}$ when $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta}) \propto \pi(\boldsymbol{x}|\boldsymbol{\eta})\pi(\boldsymbol{y}|\boldsymbol{x})$. Middle: Full conditional for $\boldsymbol{x}$ when $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ is a approximate of CSN. Right: Difference between $\pi(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$ and $\hat{\pi}(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\eta})$.

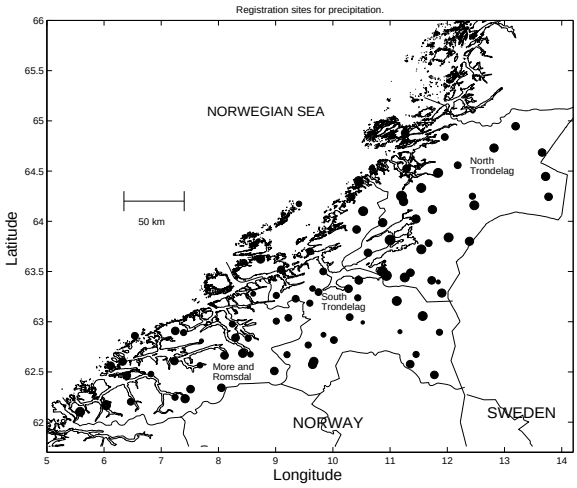


Figure 3: Precipitation data shown in a map of middle Norway. Filled circles are 92 registration sites of rainfall. Size of circles indicate the proportion of days with rain.

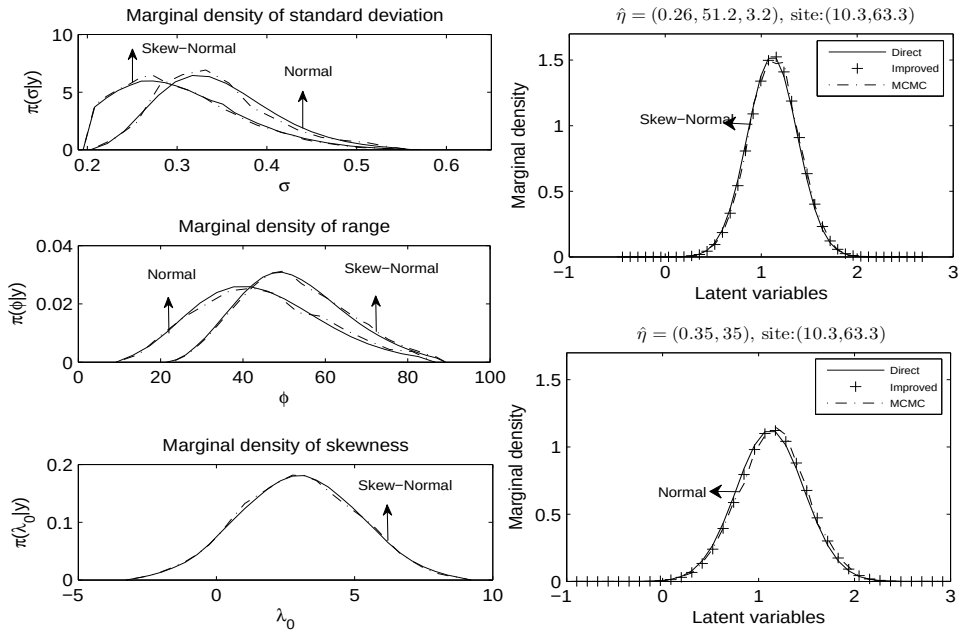


Figure 4: Left: Direct approximation of posterior marginals for $\sigma$, $\varphi$ and $\lambda_0$ (solid) and MCMC approximation of posterior marginals (dashed-dotted) when latent variables is SN and normal. Right: Conditional density $\hat{\pi}(x_j|\boldsymbol{y},\hat{\eta})$ ($\hat{\eta}$ is mode of $\hat{\pi}(\boldsymbol{\eta}|\boldsymbol{y})$) obtained by approximation Bayesian inference for model with SN (top) and normal (below) prior models for spatial latent variables. Solid is direct approximation, '+' is improved approximation, and dashed-dotted is approximation by MCMC sampling.

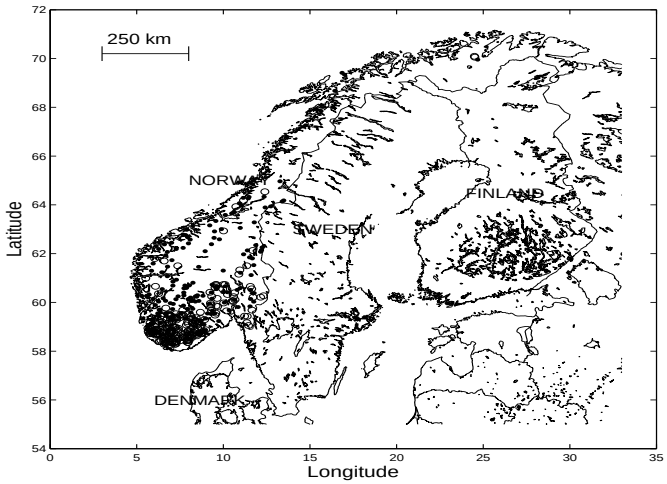


Figure 5: Trout data. Open circles denote lakes where trout are not affected by acidification, Filled circles are lakes where the trout population has decreased or is extinct. In total there are 542 observation sites.

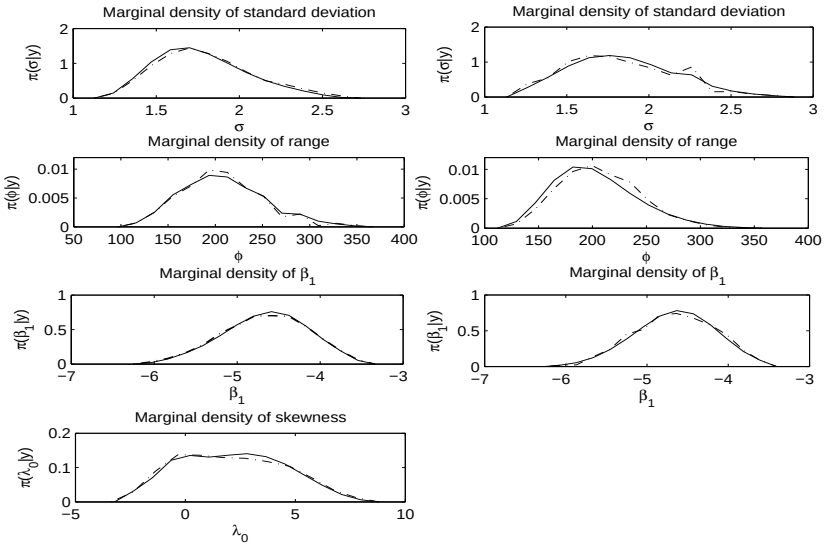


Figure 6: Left: Direct approximation of posterior marginals for $\sigma$, $\phi$, $\beta_1$ and $\lambda_0$ (solid) and MCMC sampling approximation of posterior marginals (dashed-dotted) when latent variable has SN prior. Right: Direct approximation of posterior marginals for $\sigma$, $\phi$ and $\beta_1$ (solid) and MCMC sampling approximation of posterior marginals (dashed-dotted) when latent variable has normal prior.

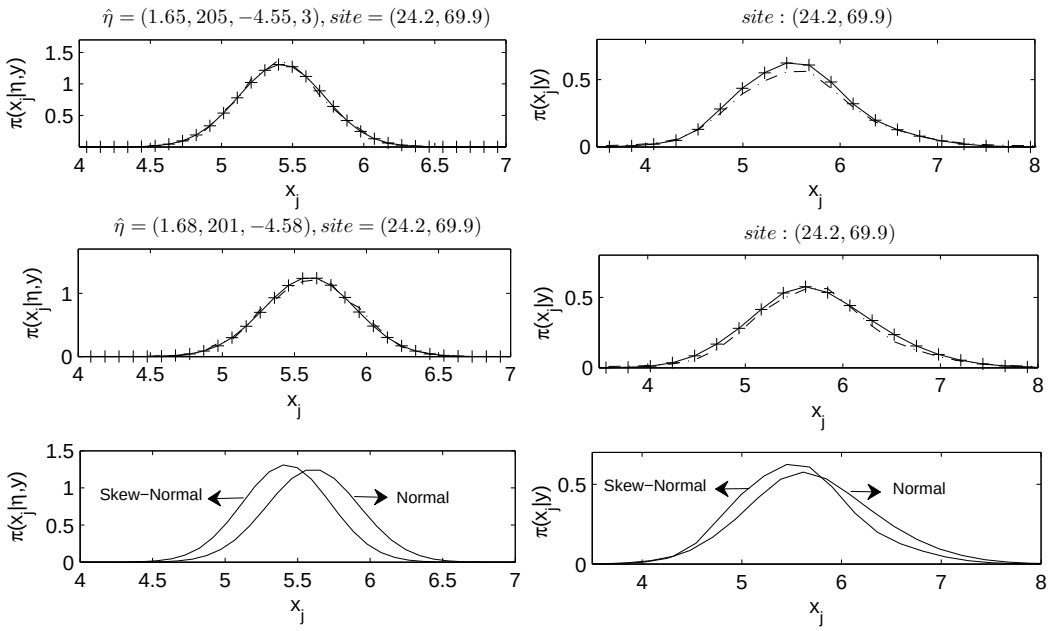


Figure 7: Left: Approximations to conditional density $\pi(x_j|\boldsymbol{y}, \hat{\eta})$ ($\hat{\eta}$ is mode of $\hat{\pi}(\eta|y)$) obtained with SN prior for latent variable (top) and for normal prior for latent variable (middle). Right: Approximate predictive marginal $\pi(x_j|y)$ obtained by SN prior (top) and normal prior (middle). Solid is direct approximation, + is improved approximation, and dashed-dotted approximation by MCMC sampling. Below plots show SN and normal results (top and middle) at the chosen prediction site in combined displays.