A Non-Gaussian Spatial Generalized Linear Latent Variable Model

Irina IRINCHEEVA, Eva CANTONI, and Marc G. GENTON

We consider a spatial generalized linear latent variable model with and without normality distributional assumption on the latent variables. When the latent variables are assumed to be multivariate normal, we apply a Laplace approximation. To relax the assumption of marginal normality in favor of a mixture of normals, we construct a multivariate density with Gaussian spatial dependence and given multivariate margins. We use the pairwise likelihood to estimate the corresponding spatial generalized linear latent variable model. The properties of the resulting estimators are explored by simulations. In the analysis of an air pollution data set the proposed methodology uncovers weather conditions to be a more important source of variability than air pollution in explaining all the causes of non-accidental mortality excluding accidents.

Key Words: Copula; Factor analysis; Latent variable; Mixture of Gaussians; Multivariate random field; Non-normal; Spatial data.

1. INTRODUCTION

Let $\mathbf{X}_s = (X_{s1}, \ldots, X_{sp})^T,\ s = 1, \ldots, S$, be a realization of a p-variate random field, i.e. the components $X_{s1}, \ldots, X_{sp}$ are correlated at each location s and $\mathbf{X}_s$ is correlated with $\mathbf{X}_{s'}$ for any $s' = 1, \ldots, S,\ s \neq s'$. Such multivariate spatial data frequently occur in social sciences, public health, environmental research and many other fields. It is appealing to reduce the dimensionality of $\mathbf{X}_s$ by assuming the presence of an underlying unobservable q-variate random vector $\mathbf{Z}_s$, with $q < p$, explaining all the systematic variability of $\mathbf{X}_s$ at location s. Such $\mathbf{Z}_s$ are called latent variables and can be constructed via various spatial linear latent variable models.

Irina Irincheeva (✉) is Postdoctoral Associate, Department of Statistical Science, Duke University, Durham, USA (E-mail: *Irina.Irincheeva@stat.duke.edu*). Eva Cantoni is Associate Professor, Research Center for Statistics and Department of Economics, University of Geneva, Geneva, Switzerland (E-mail: *Eva.Cantoni@unige.ch*). Marc G. Genton is Professor, Department of Statistics, Texas A&M University, College Station, USA (E-mail: *genton@stat.tamu.edu*).

The research was carried out while the first author was a Ph.D. Student at the University of Geneva.

© 2012 International Biometric Society

Journal of Agricultural, Biological, and Environmental Statistics, Volume 17, Number 3, Pages 332–353
DOI: 10.1007/s13253-012-0099-5

The most popular version of a spatial linear latent variable model is spatial factor analysis (Christensen and Amemiya 2002; Reich and Bandyopadhyay 2010) in which one assumes that the manifest vector $\mathbf{X}_s$ contains only continuous variables. The Spatial Generalized Linear Latent Variable model (SGLLVM; Wang and Wall 2003; Hogan and Tchernis 2004; Minozzo and Fruttini 2004) is an extension of factor analysis to the case when the components of the vector $\mathbf{X}_s$ conditionally on $\mathbf{Z}_s$ follow any distribution from the linear exponential family (for example, Bernoulli, Poisson and normal).

All of the currently known spatial linear latent variable models assume the latent variables to have a multivariate normal distribution, though it is known that the estimation and the inference in simple Generalized Linear Latent Variable Model (GLLVM) are sensitive to the incorrect specification of the latent variable distribution (Ma and Genton 2010). In particular, in a factor analysis model, the number of selected latent variables tends to be larger when the distribution of latent variables is not specified correctly (Montanari and Viroli 2010b). In this article, our focus is on relaxing the normality assumption in the marginal distribution of the latent variables at location s.

As a motivating example we consider a sub-sample of the well-known data base of the National Morbidity, Mortality and Air Pollution Study (NMMAPS, available from *www.ihapss.jhsph.edu/data/NMMAPS/R/*). The considered sample describes 93 U.S. cities on May 15, 2000 with the following nine variables: death.count, representing death counts for all causes of mortality except accidents; log.pop, being the log of the city population; tmpd—mean temperature; dptp—dew point temperature; mxrh—maximum relative humidity; mnrh—minimum relative humidity; o3tmean—trimmed mean for ozone hourly concentration; rmtmpd—adjusted three-day lag temperature; rmdptp—adjusted three-day lag dew point temperature. The exploratory graph for univariate and bivariate marginal distributions of all the nine variables is displayed in Figure 1.

Clearly, none of the continuous variables in Figure 1 can be assumed to be normal, but the normality assumption of the latent variables, together with the usual identity link function, would imply the normality of the manifest variables. Although Figure 1 does not take into account the spatial correlations, one sees clearly the need to model the overdispersion of count manifest variables, as well as the asymmetry and bimodality of continuous manifest variables. Various ad hoc approaches, such as data transformations, or separating data in two sub-samples, allow one to bypass this necessity, but can induce considerable bias in the inference. Therefore, we assume the latent variables to be a bimodal mixture of normals, marginally at location s.

This seemingly simple assumption raises the challenge of constructing a multivariate distribution with given non-overlapping multivariate margins, i.e. we need to construct the probability density function of the qS-variate random vector $(\mathbf{Z}_1^T, \ldots, \mathbf{Z}_S^T)^T$, knowing the probability density functions of q-variate vectors $\mathbf{Z}_s$, $s = 1, \ldots, S$. This problem is an active field of statistics involving works on copulas (Nelsen 2006; Bardossy 2006; Kazianka and Pilz 2011), Markov random fields (Rue and Held 2005), non-parametric approaches (Brown, Diggle, and Henderson 2003; Gelfand, Kottas, and MacEachern 2005), among many others. In this article we use a "linkage" tool, proposed by Li, Scarsini, and Shaked (1996), which can be viewed as a generalization of copulas to multivariate marginal distributions. To the best of our knowledge, it is the first work with linkages in spatial statistics.

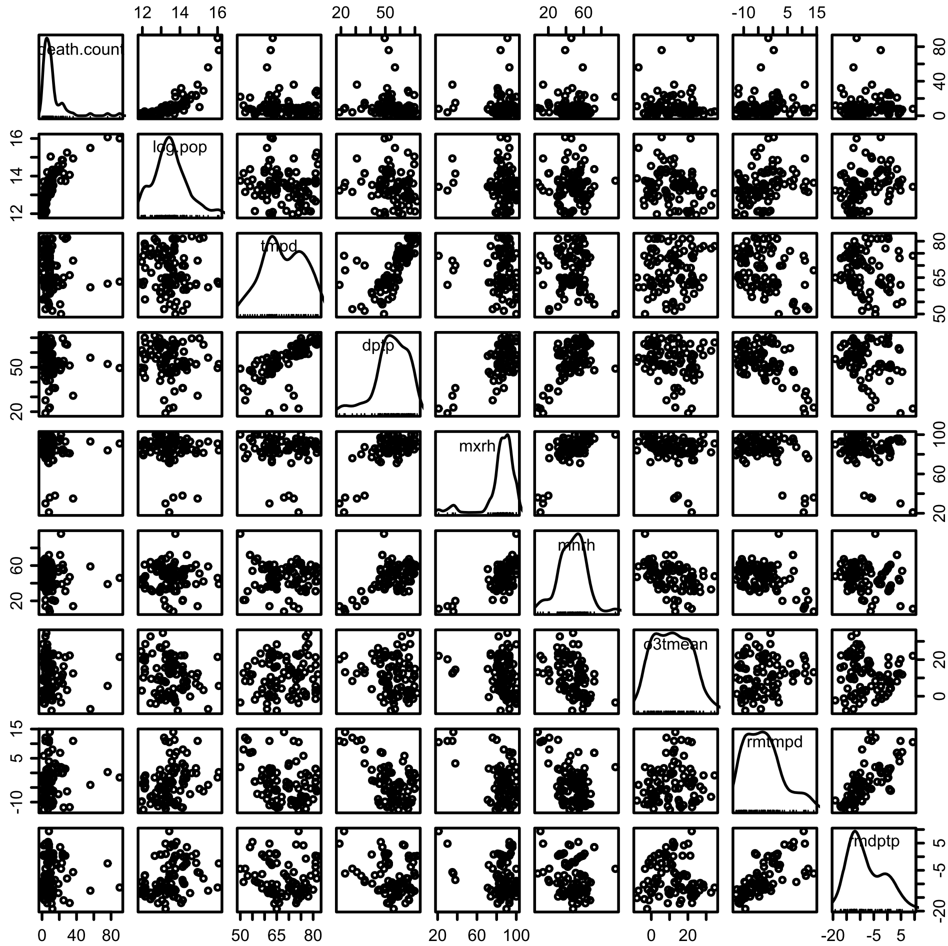

Figure 1. Exploratory plot of NMMAPS data set.

In Section 2 we review the spatial factor analysis model and SGLLVM, proposing a Laplace approximation for SGLLVM. In Section 3 we introduce linkages, discuss the distribution of the latent variables, propose a Monte Carlo EM pairwise likelihood method and approximate (with adaptive integration) the pairwise likelihood. Finally, Section 4 explores by simulations the distributional properties of the proposed estimators, and Section 5 contains supplementary motivations and an analysis of NMMAPS data.

2. SGLLVM WITH GAUSSIAN LATENT VARIABLES

2.1. Conditionally Gaussian Manifest Variables

When all the manifest variables can be assumed to be conditionally normal given multivariate normal latent variables, the SGLLVM model is simply spatial factor analysis. In this case, it is known that the final distribution of the manifest variables exists in closed form and is multivariate normal. We did not find in the literature the description of the frequentist maximum likelihood estimator so we describe it briefly here. At location s for manifest variable i, the factor analysis model can be written as follows:

$$\begin{aligned} X_{si} &= \mu_i + \boldsymbol{\gamma}_i^T \mathbf{Z}_s + \varepsilon_{si}, \\ \mathbf{Z}_s &\sim N_q(\mathbf{0}, \mathbf{I}_q) \quad \text{independent of } \varepsilon_{si} \sim N(0, \psi_i), \end{aligned} \tag{2.1}$$

where $N_q(\mathbf{0}, \mathbf{I}_q)$ denotes a q-variate normal distribution with location $\mathbf{0}$ and $q \times q$ identity matrix $\mathbf{I}_q$ as covariance. Notice that mutual independence of ε_{si} is crucial to any latent variable model and translates the assumption that the latent variables explain all the systematic variability of the manifest variables. Equation (2.1) can be rewritten in matrix form as

$$\operatorname{vec}\mathbf{X} = \boldsymbol{\mu} \otimes \boldsymbol{\iota} + (\boldsymbol{\Gamma} \otimes \mathbf{I}_S)\operatorname{vec}\mathbf{Z} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim N_{qS}(\mathbf{0}, \boldsymbol{\Psi} \otimes \mathbf{I}_S), \tag{2.2}$$

where $\mathbf{X} = (\mathbf{X}_1, \dots, \mathbf{X}_S)^T \in \mathbb{R}^{S\times p}$, $\mathbf{Z} = (\mathbf{Z}_1, \dots, \mathbf{Z}_S)^T \in \mathbb{R}^{S\times q}$, $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T$, $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_p)^T$ and $\boldsymbol{\Psi} = \operatorname{diag}(\psi_1, \dots, \psi_p)$ are parameters; $\boldsymbol{\iota}$ is the $S \times 1$-vector of ones and $\mathbf{I}_S$ is the $S \times S$ identity matrix. An alternative form of Equation (2.2) is

$$\operatorname{vec}\mathbf{X}^T = \boldsymbol{\iota} \otimes \boldsymbol{\mu} + (\mathbf{I}_S \otimes \boldsymbol{\Gamma})\operatorname{vec}\mathbf{Z}^T + \boldsymbol{\epsilon}, \qquad \boldsymbol{\epsilon} \sim N_{qS}(\mathbf{0}, \mathbf{I}_S \otimes \boldsymbol{\Psi}). \tag{2.3}$$

This form can be obtained by applying commutation matrices (Magnus and Neudecker 1988, p. 54) to $\operatorname{vec}\mathbf{X}$ and $\operatorname{vec}\mathbf{Z}$ in (2.2) and is basically another arrangement of the equations in (2.1).

Given (2.3) the joint distribution of $\operatorname{vec}\mathbf{X}$ is

$$\begin{aligned} f(\mathbf{x}) &= \frac{1}{|2\pi\boldsymbol{\Omega}|^{1/2}|2\pi\boldsymbol{\Sigma}|^{1/2}} \int_{\mathbb{R}^{qS}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\nu} - \mathbf{F}\mathbf{z})^T \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\nu} - \mathbf{F}\mathbf{z}) - \frac{1}{2}\mathbf{z}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}\right\} d\mathbf{z} \\ &= \frac{1}{|2\pi(\boldsymbol{\Omega} + \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}^T)|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}_0^T\left(\boldsymbol{\Omega} + \mathbf{F}\boldsymbol{\Sigma}\mathbf{F}^T\right)^{-1}\mathbf{x}_0\right\}, \end{aligned} \tag{2.4}$$

where $\mathbf{x} = (\mathbf{x}_1^T, \dots, \mathbf{x}_S^T)^T$ is a pS-variate vector; $\mathbf{z} = (\mathbf{z}_1^T, \dots, \mathbf{z}_S^T)^T$ is a qS-variate vector; symbols $\boldsymbol{\nu} = \boldsymbol{\iota} \otimes \boldsymbol{\mu}$, $\mathbf{F} = \mathbf{I}_S \otimes \boldsymbol{\Gamma}$, $\boldsymbol{\Omega} = \mathbf{I}_S \otimes \boldsymbol{\Psi}$, $\mathbf{x}_0 = \mathbf{x} - \boldsymbol{\nu}$ are introduced to improve readability; and $\boldsymbol{\Sigma}$ is the correlation matrix of $\operatorname{vec}\mathbf{Z}$, such that

$$\operatorname{corr}(Z_{sj}, Z_{lk}) = 0, \quad \text{if } j \neq k, \ j, k = 1, \dots, q; \tag{2.5}$$

$$\operatorname{corr}(Z_{sj}, Z_{lj}) = \exp\left\{-\frac{d_{sl}}{\lambda_j}\right\}, \quad j = 1, \dots, q; \tag{2.6}$$

with unknown parameters $\lambda_j > 0$, $j, k = 1, \dots, q$ and d_{sl} the distance between locations s and l.

Other correlation models for $\boldsymbol{\Sigma}$ can be specified instead of the exponential covariance model (2.6), for instance the Matérn covariance function, see Gneiting, Genton, and Guttorp (2007). We start however with a simple exponential model with the intention of extending it to a more sophisticated covariance model.

2.2. Mixed-Scale Manifest Variables

The SGLLVM was first introduced by Wang and Wall (2003) under the name of generalized common spatial factor model. Wang and Wall (2003) adopted a Bayesian approach to estimation of parameters with a Markov chain Monte Carlo technique to approximate the integral in the posterior distribution. We briefly present the model here and propose the Laplace approach to approximate the integral in the model likelihood, which is less computationally expensive and faster than Markov chain Monte Carlo techniques. The use of a Laplace approximation is new in SGLLVM, though this approach is used in spatial generalized mixed effects models (Shun and McCullagh 1995) and in GLLVM (Huber, Ronchetti, and Victoria-Feser 2004).

Without loss of generality we can consider $X_{s1}|\mathbf{Z}_s, \ldots, X_{sp_1}|\mathbf{Z}_s$ to be Bernoulli distributed, $X_{s,p_1+1}|\mathbf{Z}_s, \ldots, X_{sp_2}|\mathbf{Z}_s$ to be Poisson distributed and $X_{s,p_2+1}|\mathbf{Z}_s, \ldots, X_{sp}|\mathbf{Z}_s$ to be continuous and generated by a normal distribution. Then

Bernoulli ($k = 1, \ldots, p_1$):

$$g_i(x_{si} \mid \mathbf{z}_s; \mu_i, \boldsymbol{\gamma}_i, \phi_i) = \frac{\exp(x_{si}\mu_i + x_{si}\boldsymbol{\gamma}_i^T \mathbf{z}_s)}{1 + \exp(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s)} \tag{2.7}$$

$$\text{i.e.} \quad b_i\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big) = \log\big\{\, 1 + \exp\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big)\big\};$$

$$c_i(x_{si}, \phi_i) = 0 \quad \text{with } \phi_i = 1,\ a_i(\phi_i) = 1;$$

Poisson ($k = p_1 + 1, \ldots, p_2$):

$$g_i(x_{si} \mid \mathbf{z}_s; \mu_i, \boldsymbol{\gamma}_i, \phi_i) = \exp\big\{x_{si}\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}\big) - \exp\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big)\big\}/x_{si}! \tag{2.8}$$

$$\text{i.e.} \quad b_i\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big) = \exp\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big);$$

$$c_i(x_{si}, \phi_i) = -\log x_{si}! \quad \text{with } \phi_i = 1,\ a_i(\phi_i) = 1;$$

Normal ($k = p_2 + 1, \ldots, p$):

$$g_i(x_{si} \mid \mathbf{z}_s; \mu_i, \boldsymbol{\gamma}_i, \phi_i) = \sqrt{\frac{1}{2\pi\phi_i}} \exp\left\{-\frac{1}{2\phi_i}\big(x_{si} - \mu_i - \boldsymbol{\gamma}_i^T \mathbf{z}_s\big)^2\right\} \tag{2.9}$$

$$\text{i.e.} \quad b_i\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big) = 0.5\big(\mu_i + \boldsymbol{\gamma}_i^T \mathbf{z}_s\big)^2;$$

$$c_i(x_{si}, \phi_i) = -\frac{x_{si}^2}{2\phi_i} - 0.5\log(2\pi\phi_i) \quad \text{with } a_i(\phi_i) = \phi_i.$$

This description can be extended to any other distribution belonging to the linear exponential family. The probability mass function of vec $\mathbf{X}$ in the SGLLVM can be written as

$$f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Gamma}, \phi) = \frac{1}{|2\pi\boldsymbol{\Sigma}|^{1/2}} \int_{\mathbb{R}^{qS}} \left\{\prod_{i=1}^{p}\prod_{s=1}^{S} g_i(x_{si} \mid \mathbf{z}_s; \mu_i, \boldsymbol{\gamma}_i, \phi_i)\right\} \exp\big\{-0.5\mathbf{z}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}\big\}\, d\mathbf{z}, \tag{2.10}$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_p)^T$ is loadings matrix, $\boldsymbol{\Sigma}$ is correlation matrix of vec $\mathbf{Z}$ defined by (2.5) and (2.6), and $\phi = (\phi_{p_2+1}, \ldots, \phi_p)^T$, $\phi_1, \ldots, \phi_{p_2}$ being identically ones in (2.7) and

(2.8). The multivariate integral (2.10) exists in closed form when all the manifest variables are conditionally normal given normally distributed latent variables; in other cases it necessitates numerical approximation.

The likelihood induced by (2.10) can also be viewed as one observation likelihood of GLLVM with qS correlated latent variables. Indeed, $\mathbf{\Sigma}$ is not the $qS \times qS$ identity matrix but contains positive non-diagonal entries as specified by (2.5) and (2.6). As shown in Huber (2004, pp. 29–32), the likelihood (2.10) is identified up to an oblique rotation of the latent variables $\mathbf{Z}$, i.e., in order to have a unique solution, $qS(qS-1)/2$ elements of the loadings matrix $\mathbf{\Gamma}$ in (2.10) should be constrained to zero (for oblique rotation, see Bartholomew, Knott, and Moustaki 2011, p. 35). Then the SGLLVM (2.10) is identified up to the sign of γ_{11} because of the particular form of its loadings matrix $\mathbf{I}_S \otimes \mathbf{\Gamma}$ which has $pSqS - pqS$ zero elements (it is easy to check that $pS(qS-q) > qS(qS-1)/2$). This particular form of loadings matrix becomes evident if we consider the fundamental relationship to the likelihood function (2.10): $\zeta\{E(\mathrm{vec}\,\mathbf{X})|\,\mathrm{vec}\,\mathbf{Z}\} = \boldsymbol{\iota} \otimes \boldsymbol{\mu} + (\mathbf{I}_S \otimes \mathbf{\Gamma})\mathbf{z}$, where the function $\zeta\{\cdot\}$ is a function from $\mathbb{R}^p$ to $\mathbb{R}^p$ with component functions $\zeta_i\{\cdot\}$, $i = 1, \ldots, p$, being monotonic link functions. We found a similar approach to the identification of spatiotemporal GLLVM in the recent work by Lopes, Gamerman, and Salazar (2011).

Huber, Ronchetti, and Victoria-Feser (2004) introduced the Laplace approximation in simple GLLVM without spatially correlated latent variables. Their work is applicable in our case when the integral dimension qS increases together with the exponent order pS of the integrated function. To apply the Laplace approximation to the multidimensional integral in (2.10) we rewrite Equation (2.10) as

$$f(\mathbf{x}) = \int_{\mathbb{R}^{qS}} \exp\big\{p\, \mathcal{Q}(\mathbf{\Gamma}, \phi, \mathbf{z}, \mathbf{x})\big\}\, d\mathbf{z}, \tag{2.11}$$

where

$$\mathcal{Q}(\mathbf{\Gamma}, \phi, \mathbf{z}, \mathbf{x}) = \frac{1}{p}\left[\left\{\sum_{i=1,s=1}^{p,S} \log\, g_i(x_{si}|\mathbf{z}_s)\right\} - \frac{\mathbf{z}^T\mathbf{\Sigma}^{-1}\mathbf{z}}{2} - \frac{1}{2}\log|2\pi\,\mathbf{\Sigma}|\right] \tag{2.12}$$

and

$$\log\, g_i(x_{si}|\mathbf{z}_s) = \frac{x_{si}\mu_i + x_{si}\boldsymbol{\gamma}_i^T\mathbf{z}_s - b_i(\mu_i + \boldsymbol{\gamma}_i^T\mathbf{z}_s)}{a_i(\phi_i)} + c_i(x_{si}, \phi_i).$$

Applying the Laplace approximation to the density (2.11), we obtain

$$f(\mathbf{x}) = \left(\frac{2\pi}{p}\right)^{q/2} \big|\mathbf{U}(\hat{\mathbf{z}})\big|^{-1/2} \exp\big\{p\mathcal{Q}(\mathbf{\Gamma}, \phi, \hat{\mathbf{z}}, \mathbf{x})\big\}\big\{1 + O\big(p^{-1}\big)\big\}, \tag{2.13}$$

where

$$\mathbf{U}(\hat{\mathbf{z}}) = \frac{\partial^2\mathcal{Q}(\mathbf{\Gamma}, \phi, \mathbf{z}, \mathbf{x})}{\partial\mathbf{z}\partial\mathbf{z}^T}\bigg|_{\mathbf{z}=\hat{\mathbf{z}}} = -\frac{1}{p}\left\{\sum_{i=1}^{p} \frac{1}{\phi_i} \frac{\partial^2 b_i(\mu_i + \boldsymbol{\gamma}_i^T\mathbf{z}_s)}{\partial\mathbf{z}\partial\mathbf{z}^T}\bigg|_{\mathbf{z}=\hat{\mathbf{z}}} + \mathbf{\Sigma}^{-1}\right\}, \tag{2.14}$$

and $\hat{\mathbf{z}}$ is the unique maximum of $\mathcal{Q}(\mathbf{\Gamma}, \phi, \mathbf{z}, \mathbf{x})$.

Then the approximate log-likelihood function for (2.10) is

$$\tilde{\ell}(\boldsymbol{\mu}, \boldsymbol{\Gamma}, \phi|\mathbf{x}) = -\frac{1}{2}\log\left|\mathbf{U}(\hat{\mathbf{z}})\right| - \frac{\hat{\mathbf{z}}^T\boldsymbol{\Sigma}^{-1}\hat{\mathbf{z}}}{2} - \frac{1}{2}\log|2\pi\boldsymbol{\Sigma}| + \frac{q}{2}\log\left(\frac{2\pi}{p}\right) + \left\{\sum_{i=1,s=1}^{p,S} \frac{x_{si}\mu_i + x_{si}\boldsymbol{\gamma}_i^T\hat{\mathbf{z}}_s - b_i(\mu_i + \boldsymbol{\gamma}_i^T\hat{\mathbf{z}}_s)}{a_i(\phi_i)} + c_i(x_{si}, \phi_i)\right\}. \quad (2.15)$$

Score functions for the approximated log-likelihood function and derivatives $\partial^2 b_i(\mu_i + \boldsymbol{\gamma}_i^T\mathbf{z}_s)/\partial\mathbf{z}\partial\mathbf{z}^T$ can be found in Huber (2004, pp. 75–82).

3. SGLLVM WITH MIXTURE OF GAUSSIAN LATENT VARIABLES

In this section we construct the multivariate distribution of latent variables with non-normal multivariate non-overlapping marginal distributions and propose to estimate the SGLLVM with the resulting distribution of latent variables via the pairwise likelihood.

3.1. ASSUMPTIONS ON THE LATENT VARIABLES DISTRIBUTION

Similarly to Verbeke and Lesaffre (1996), we assume the marginal distribution of $\mathbf{Z}_s,\ s = 1, \dots, S$, to be a mixture of K Gaussian distributions

$$\sum_{r=1}^{K}\pi_r N_q(\boldsymbol{\alpha}_r, \mathbf{I}_q), \quad \text{with } \pi_K = 1 - \sum_{r=1}^{K-1}\pi_r. \quad (3.1)$$

We also set $\boldsymbol{\alpha}_K = -\sum_{r=1}^{K-1}\pi_r\boldsymbol{\alpha}_r/\pi_K$ in order to satisfy $E(\mathbf{Z}_s) = \mathbf{0}$ and $\pi_1 > \pi_2 > \cdots > \pi_{K-1} > \pi_K$ for identification of (3.1). An additional constraint, as in Montanari and Viroli (2010a), can be put on $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_K, \pi_1, \dots, \pi_K$, in order to have $\mathrm{var}(\mathbf{Z}_s) = \sum_{r=1}^{K}\pi_r(\mathbf{I}_q + \boldsymbol{\alpha}_K\boldsymbol{\alpha}_K^T) = \mathbf{I}_q$. In this work, for the sake of computational speed, we prefer standardizing the estimated loadings. If one accepts to increase the computational burden, then it is possible to increase the flexibility of density (3.1) allowing $K-1$ covariance matrices to be of the form $\sigma_r^2\mathbf{I}_q$ with $\sigma_r^2,\ r = 2, \dots, K$, additional parameters to estimate.

The qS-variate distribution of $\mathrm{vec}\,\mathbf{Z}$ with non-overlapping q-variate margins (3.1) is constructed with the "linkage" tool proposed by Li, Scarsini, and Shaked (1996). Consider the transformation $\boldsymbol{\Theta}_F : \mathbb{R}^q \to [0,1]^q$ defined by the relationship

$$\boldsymbol{\Theta}_F(z_{s1}, \dots, z_{sq}) = \left[F_1(z_{s1}), F_{2|1}(z_{s2}|z_{s1}), \dots, F_{q|1,2,\dots,q-1}(z_{sq}|z_{s1}, z_{s2}, \dots, z_{s,q-1})\right], \quad (3.2)$$

where F_1 is the marginal cumulative distribution function of Z_{s1}; $F_{i+1|1,2,\dots,i}(\cdot|z_{s1}, z_{s2}, \dots, z_{si})$ is the conditional cumulative distribution function of $Z_{s,i+1}$ given $Z_{s1} = z_{s1},\ Z_{s2} = z_{s2}, \dots,\ Z_{si} = z_{si}$. Notice that if

$$\mathbf{U}_s^T = (U_{s1}, U_{s2}, \dots, U_{sq}) = \boldsymbol{\Theta}_F(Z_{s1}, Z_{s2}, \dots, Z_{sq}), \quad (3.3)$$

then $U_{s1}, U_{s2}, \ldots, U_{sq}$ are independent uniform [0, 1] random variables. The joint distribution of

$$\left(\mathbf{U}_1^T, \mathbf{U}_2^T, \ldots, \mathbf{U}_S^T\right) = \left[\boldsymbol{\Theta}_F\left(\mathbf{Z}_1^T\right), \boldsymbol{\Theta}_F\left(\mathbf{Z}_2^T\right), \ldots, \boldsymbol{\Theta}_F\left(\mathbf{Z}_S^T\right)\right] \tag{3.4}$$

is called linkage corresponding to $(\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_S^T)$.

Now let $(\mathbf{U}_1^T, \mathbf{U}_2^T, \ldots, \mathbf{U}_S^T)$ be a qS-dimensional Gaussian copula (Nelsen 2006) with correlation matrix

$$\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{I}_q & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} & \ldots & \boldsymbol{\Sigma}_{1S} \\ \boldsymbol{\Sigma}_{12} & \mathbf{I}_q & \boldsymbol{\Sigma}_{23} & \ldots & \boldsymbol{\Sigma}_{2S} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{1S} & \boldsymbol{\Sigma}_{2S} & \boldsymbol{\Sigma}_{3S} & \ldots & \mathbf{I}_q \end{pmatrix}, \tag{3.5}$$

where the $q \times q$-matrix $\boldsymbol{\Sigma}_{sl} = \text{diag}[\exp(-d_{sl}/\lambda_j)]_{j=1,\ldots,q}$. It follows from (3.4) and Li, Scarsini, and Shaked (1996, pp. 28–30) that the vector $(\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_S^T)$ is such that $\mathbf{Z}_s \sim \sum_{r=1}^K \pi_r N_q(\boldsymbol{\alpha}_r, \mathbf{I}_q)$ and $\text{corr}(\mathbf{Z}_s, \mathbf{Z}_l) = \boldsymbol{\Sigma}_{sl}$. The probability density function of $(\mathbf{Z}_1^T, \mathbf{Z}_2^T, \ldots, \mathbf{Z}_S^T)$ is then

$$\begin{aligned} f(\mathbf{z}) = {} & \frac{\varphi_{\boldsymbol{\Sigma}}(v_{11}, v_{12}, \ldots, v_{Sq})}{\varphi(v_{11})\varphi(v_{12})\cdots\varphi(v_{Sq})} \\ & \times f_1(z_{11}) f_{2|1}(z_{12}|z_{11}) \cdots f_{q|1,2,\ldots,q-1}(z_{Sq}|z_{S1}, z_{S2}, \ldots, z_{S,q-1}), \end{aligned} \tag{3.6}$$

where $\varphi_{\boldsymbol{\Sigma}}$ is the density of the qS-variate centered Gaussian distribution with correlation matrix $\boldsymbol{\Sigma}$ as covariance ($\boldsymbol{\Sigma}$ should be positive definite); φ is the univariate standard Gaussian density; $v_{si} = \Phi^{-1}[F_{i|1,2,\ldots,i-1}(z_{si}|z_{s1}, z_{s2}, \ldots, z_{s,i-1})]$; Φ^{-1} is the inverse cumulative distribution function of the univariate standard Gaussian distribution; and $f_{i+1|1,2,\ldots,i}(\cdot|z_{s1}, z_{s2}, \ldots, z_{si})$ is the density corresponding to the cumulative distribution function $F_{i+1|1,2,\ldots,i}$. The weakly stationarity of (3.6) with univariate margins is discussed in Kazianka and Pilz (2011), and is valid also in our case because of Gaussian copula properties.

It would be tempting to specify the conditional distribution of $\mathbf{Z}_s|\mathbf{Z}_l$ to be normal, take $\mathbf{Z}_s$ and $\mathbf{Z}_l$ to be marginally distributed as mixture of Gaussians, compute the joint distribution of $(\mathbf{Z}_s^T, \mathbf{Z}_l^T)^T$ and then use the pairwise likelihood approach. This approach proves to be incorrect at the level of distributional assumptions. Indeed, let $Z_s, Z_l \sim \pi_1 N(\mu_1, 1) + \pi_2 N(\mu_2, 1)$ be a univariate mixture of Gaussians. Then the joint density of (Z_s, Z_l) is

$$f(z_s, z_l) = \frac{\varphi_{\boldsymbol{\Sigma}_{sl}}(v_{s1}, v_{l1})}{\varphi(v_{s1})\varphi(v_{l1})} f(z_s) f(z_l) \tag{3.7}$$

and the conditional density of $Z_s|Z_l$ is

$$f(z_s|z_l) = \frac{\varphi_{\boldsymbol{\Sigma}_{sl}}(v_{s1}, v_{l1})}{\varphi(v_{s1})\varphi(v_{l1})} f(z_s),$$

which evidently cannot be a Gaussian density if $f(z_s)$ is a mixture of Gaussian densities.

3.2. PAIRWISE LIKELIHOOD

The density (3.6) is multi-modal and highly multidimensional, hence plugging it in the likelihood (2.10) instead of $\exp\{-0.5\mathbf{z}^T\boldsymbol{\Sigma}^{-1}\mathbf{z}\}/|2\pi\boldsymbol{\Sigma}|^{1/2}$ increases the computational difficulty of an already "unfriendly" multidimensional integral. Usual integral approximation methods such as Gaussian quadrature, Laplace or adaptive Gaussian quadrature perform poorly because of the multi-modality of the integrand. In spatiotemporal GLLVM with multivariate Gaussian latent variables, Zhu, Eickhoff, and Yan (2005) used Monte Carlo integration, simulating the latent scores with a Metropolis–Hastings algorithm that necessitates a burn-in period. A similar approach could be useful in our case. We decided, however, to implement the pairwise likelihood approach (Cox and Reid 2004; Varin, Reid, and Firth 2011) that seems less demanding and time-consuming at the simulations level than the approach of Zhu, Eickhoff, and Yan (2005).

The pairwise log-likelihood function for the SGLLVM model is

$$P\ell(\boldsymbol{\theta}|\mathbf{x}) = \sum_{(s,l)\in\mathcal{R}} \log\left[\int_{\mathbb{R}^{2q}} f(\mathbf{x}_s,\mathbf{x}_l,\mathbf{z}_s,\mathbf{z}_l;\boldsymbol{\theta})\,d\mathbf{z}_s\,d\mathbf{z}_l\right]$$

$$= \sum_{(s,l)\in\mathcal{R}} \log\left[\int_{\mathbb{R}^{2q}} g(\mathbf{x}_s|\mathbf{z}_s;\boldsymbol{\theta})g(\mathbf{x}_l|\mathbf{z}_l;\boldsymbol{\theta})f(\mathbf{z}_s,\mathbf{z}_l;\boldsymbol{\theta})\,d\mathbf{z}_s\,d\mathbf{z}_l\right], \tag{3.8}$$

where $\mathcal{R}$ is a subset of all possible pairwise neighbors; $g(\mathbf{x}_s|\mathbf{z}_s;\boldsymbol{\theta})$ and $g(\mathbf{x}_l|\mathbf{z}_l;\boldsymbol{\theta})$ are the conditional densities of $\mathbf{X}_s|\mathbf{Z}_s$ and $\mathbf{X}_l|\mathbf{Z}_l$ respectively; $f(\mathbf{z}_s,\mathbf{z}_l;\boldsymbol{\theta})$ is the joint density of $(\mathbf{Z}_s^T,\mathbf{Z}_l^T)^T$ given by Equation (3.7); and $\boldsymbol{\theta}$ is the overall parameters' vector $\boldsymbol{\theta} = (\boldsymbol{\mu}^T, (\operatorname{vec}\boldsymbol{\Gamma})^T, (\operatorname{diag}\boldsymbol{\Psi})^T, \boldsymbol{\lambda}^T, \boldsymbol{\pi}^T, \boldsymbol{\alpha}^T)^T$ with $\boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\alpha}$ being vectors of respectively all $\lambda_j,\ j = 1,\ldots,S$; $\pi_r,\ r = 1,\ldots,K-1$; and $\boldsymbol{\alpha}_r,\ r = 1,\ldots,K-1$. The identification of (3.8) can be shown similarly to the identification discussed in Section 2.2.

The dimension of the integral in (3.8) is decreased from qS to $2q$ at the price of efficiency loss (Cox and Reid 2004). The integral still needs to be computed numerically and in this work we try two approaches, the first one being the multidimensional adaptive algorithm implemented in the R-package "cubature" (Johnson and Narasimhan 2009). Unfortunately, the accuracy of the integral approximation over a multidimensional rectangular area decreases dramatically with higher dimensions. As a possible alternative, the second considering an integral approximation approach is the Monte Carlo EM-algorithm (Booth and Hobert 1999; Eickhoff, Zhu, and Amemiya 2004) based on Monte Carlo integration, the accuracy of which does not depend on the dimension of the integral. We briefly describe these two approaches in Sections 3.3 and 3.4 below.

Model selection, including the choice of the number of components K in the mixtures of normals, is an important issue of our pairwise likelihood approach to non-Gaussian SGLLVM. Instead of using a visualization argument, as we do here, one can use the composite Akaike selection criterion (cAIC) or composite Bayesian selection criterion (cBIC) (Varin and Vidoni 2005; Gao and Song 2010) to select among alternative models. Notice that if the SGLLVM pairwise likelihood with normal latent variables is preferred by the selection criterion, then it is preferable to use a Laplace approximation of the SGLLVM full likelihood instead of pairwise likelihood for model estimation.

3.3. ADAPTIVE INTEGRATION

The adaptive algorithm for the numerical calculation of multiple integrals over a hyper-rectangle is based on work by Genz and Malik (1980) and Berntsen, Espelid, and Genz (1991). This algorithm is programmed in C by Steven G. Johnson for the R-package "cubature" and uses the following integration rule:

$$\int_H f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s, \mathbf{z}_l)\, d\mathbf{z}_s \mathbf{z}_l \backsimeq R\big[f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s, \mathbf{z}_l)\big] = \sum_{j=1}^{L} w_j f\big(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s^{(j)}, \mathbf{z}_l^{(j)}\big), \tag{3.9}$$

where H is a rectangular region of integration with the grid containing a total of L points; $\mathbf{z}_s^{(j)}, \mathbf{z}_l^{(j)}$ are the evaluation points and w_j are the corresponding weights. The basic idea of the adaptive algorithm is to divide the region of integration into subregions, compute the integral and error estimations for each subregion and then reiterate this procedure for regions with the highest estimated errors until the overall error is less or equal to a predefined quantity. The details on the algorithm rules and its efficient implementation can be found in Berntsen, Espelid, and Genz (1991). As our integral is over $\mathbb{R}^{2q}$, we perform the change of variables $z_{si} = t_{si}/(1 - t_{si}^2)$ and compute the integral

$$\int_{[-1,1]^{2q}} f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{t}_s, \mathbf{t}_l) J\, d\mathbf{t}_s\, d\mathbf{t}_l, \tag{3.10}$$

where $\mathbf{t}_s = (t_{s1}, \dots, t_{sq})^T$ and $J = \prod_{j=1}^{q}[(1 + t_{sj}^2)(1 + t_{lj}^2)/\{(1 - t_{sj}^2)^2(1 - t_{lj}^2)^2\}]$ is the Jacobian of the change of variables.

Plugging the approximation (3.9) in (3.8) with change of variables (3.10) we obtain the approximate pairwise log-likelihood:

$$P\tilde{\ell}(\boldsymbol{\theta}|\mathbf{x}) = \sum_{(s,l)\in\mathcal{R}} \log R\big[f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s, \mathbf{z}_l)\big]. \tag{3.11}$$

3.4. PAIRWISE MONTE CARLO EM-ALGORITHM

The EM-algorithm for GLLVM estimated with maximum likelihood was proposed by Sammel, Ryan, and Legler (1997), and the pairwise likelihood EM-algorithm was introduced by Varin, Høst, and Skare (2005) for spatial generalized linear mixed effects models. Similarly to both works, we consider the latent variable values as missing data and use Monte Carlo integration to approximate the expectation step. In details, for our problem, this gives:

Monte Carlo Expectation step

At the tth iteration of the EM-algorithm, the expectation step evaluates the sum of conditional expectations:

$$\begin{aligned} Q\big(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\big) &= \sum_{(s,l)\in\mathcal{R}} E\big\{\log f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s, \mathbf{z}_l, ; \boldsymbol{\theta})|\mathbf{x}_s, \mathbf{x}_l, \boldsymbol{\theta}^{(t)}\big\} \\ &= \sum_{(s,l)\in\mathcal{R}} \int_{\mathbb{R}^{2q}} f\big(\mathbf{z}_s, \mathbf{z}_l|\mathbf{x}_s, \mathbf{x}_l; \boldsymbol{\theta}^{(t)}\big)\, d\mathbf{z}_s\, d \log f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s, \mathbf{z}_l; \boldsymbol{\theta})\mathbf{z}_l, \end{aligned} \tag{3.12}$$

where $\boldsymbol{\theta}^{(t)}$ is the value of the parameter $\boldsymbol{\theta}$ at the tth iteration of the EM-algorithm. In practice we approximate the E-step with

$$\tilde{Q}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right) = \sum_{(s,l)\in\mathcal{R}} \frac{1}{m^{(t)}} \sum_{d=1}^{m^{(t)}} f\left(\mathbf{z}_s^{(td)}, \mathbf{z}_l^{(td)}|\mathbf{x}_s, \mathbf{x}_l; \boldsymbol{\theta}^{(t)}\right) \log f\left(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s^{(td)}, \mathbf{z}_l^{(td)}; \boldsymbol{\theta}^{(t)}\right), \tag{3.13}$$

where $m^{(t)}$ is the size of the simulation and $\mathbf{z}_s^{(td)}$ is the dth ($d = 1, \dots, m^{(t)}$) realization of latent variable at location s, i.e. we have $m^{(t)}$ realizations $(\mathbf{z}_s^{(td)T}, \mathbf{z}_l^{(td)T})^T$ issued from

$$f\left(\mathbf{z}_s, \mathbf{z}_l|\mathbf{x}_s, \mathbf{x}_l; \boldsymbol{\theta}^{(t)}\right) = a g\left(\mathbf{x}_s|\mathbf{z}_s; \boldsymbol{\theta}^{(t)}\right) g\left(\mathbf{x}_l|\mathbf{z}_l; \boldsymbol{\theta}^{(t)}\right) f\left(\mathbf{z}_s, \mathbf{z}_l; \boldsymbol{\theta}^{(t)}\right), \tag{3.14}$$

with a normalizing constant a as proposed in multivariate rejection sampling by Geweke (1996).

Step 1. Simulate $(\mathbf{z}_s^{(td)T}, \mathbf{z}_l^{(td)T})^T$ from (3.7) and, independently of it, ω from a uniform distribution on [0, 1].

Step 2. If $\omega < g(\mathbf{x}_s|\mathbf{z}_s^{(td)}; \boldsymbol{\theta}^{(t)}) g(\mathbf{x}_l|\mathbf{z}_l^{(td)}; \boldsymbol{\theta}^{(t)})/\tau$, where $\tau = \sup_{\mathbf{z}_s,\mathbf{z}_t} g(\mathbf{x}_s|\mathbf{z}_s; \boldsymbol{\theta}^{(t)}) g(\mathbf{x}_l|\mathbf{z}_l; \boldsymbol{\theta}^{(t)})$, accept $(\mathbf{z}_s^{(td)T}, \mathbf{z}_l^{(td)T})^T$ as issued from (3.14), otherwise return to Step 1.

Maximization step

$$\text{Find} \quad \boldsymbol{\theta}^{(t+1)} = \arg\max_{\boldsymbol{\theta}} \tilde{Q}\left(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)}\right). \tag{3.15}$$

Choice of simulation size

It is well known (Booth and Hobert 1999; Eickhoff, Zhu, and Amemiya 2004 and references therein) that the simulation size should be increased with the number of iterations. In our case, as the parameters space is high-dimensional, we use the slightly modified approach of Eickhoff, Zhu, and Amemiya (2004), based on measuring the distance between consecutive estimated (pairwise) likelihood values, while the method of Booth and Hobert (1999) uses the distance between consecutive parameters estimates.

Let $Q_t = Q(\boldsymbol{\theta}^{(t)}|\boldsymbol{\theta}^{(t-1)})$. For a predefined $\delta > 0$ we want to bound $|Q_t - Q_{t-1}| < \delta$ using the inequality

$$|Q_t - Q_{t-1}| \le |\tilde{Q}_t - Q_t| + |\tilde{Q}_{t-1} - Q_{t-1}| + |\tilde{Q}_t - \tilde{Q}_{t-1}|. \tag{3.16}$$

The behavior of the first two terms is similar. Consider the first term. Notice that $E|\tilde{Q}_t - Q_t|^2 \approx V^{(t)}/m^{(t)}$, where

$$V^{(t)} = \operatorname{var}\left\{\sum_{(s,l)\in\mathcal{R}} \log f\left(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s^{(t)}, \mathbf{z}_l^{(t)}; \boldsymbol{\theta}^{(t)}\right)\right\}$$

can be estimated using the same Monte Carlo approximation of the integral as

$$\hat{V}_t = \frac{1}{m^{(t)} - 1} \sum_{(s,l)\in\mathcal{R}} \left\{\sum_{d=1}^{m^{(t)}} \log^2 f\left(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s^{(td)}, \mathbf{z}_l^{(td)}; \boldsymbol{\theta}^{(t)}\right) - \left(\frac{1}{m^{(t)}} \sum_{d=1}^{m} \log f\left(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s^{(td)}, \mathbf{z}_l^{(td)}; \boldsymbol{\theta}^{(t)}\right)\right)^2\right\}. \tag{3.17}$$

Notice that (3.17) overestimates $V^{(t)}$ because it does not take into account the dependence of $f(\mathbf{x}_s, \mathbf{x}_l, \mathbf{z}_s^{(t)}, \mathbf{z}_l^{(t)}; \boldsymbol{\theta}^{(t)})$ across locations. In computations we found $\tilde{V}^{(t)} = \hat{V}^{(t)}/\sqrt{|\mathcal{R}|}$ more appropriate than $\hat{V}^{(t)}$ ($|\mathcal{R}|$ denotes the number of elements in $\mathcal{R}$). Let $\epsilon > 0$ be a predetermined probability value and $\delta_1 = \delta/3$. Then by the Tchebychev–Markov inequality,

$$\Pr\{|\tilde{Q}_t - Q_t| > \delta_1\} \le \frac{\tilde{V}^{(t)}}{m^{(t)}\delta_1^2} = \epsilon, \tag{3.18}$$

hence we chose $m^{(t)} \ge \tilde{V}^{(t)}/(\epsilon\delta_1^2)$ to be sure that the probability is less than ϵ.

Then at the tth iteration:

1. Obtain $\tilde{Q}_{t-1}$ and $\boldsymbol{\theta}^{(t-1)}$ with the simulation size $m^{(t-1)}$;
2. Compute $\tilde{V}_{t-1}$;
3. Compute the smallest integer $m^{(t)} \ge \tilde{V}^{(t-1)}/(\epsilon\delta_1^2)$;
4. Compute $\tilde{Q}_t$;
5. If $|\tilde{Q}_t - \tilde{Q}_{t-1}| < \delta_1$, stop and report $\boldsymbol{\theta}^{(t)}$ as the maximum pairwise likelihood estimator; otherwise return to Step 1.

4. IMPLEMENTATION AND SIMULATIONS

4.1. METHODS

We implemented four methods: we call GFA the spatial Gaussian Factor Analysis Maximum Likelihood estimator computed by maximizing the logarithm of (2.4); LAML the Laplace Approximated Maximum Likelihood estimator of SGLLVM obtained by maximizing (2.15) first with respect to $\mathbf{z}$ in order to obtain $\hat{\mathbf{z}}$, and then with respect to $\boldsymbol{\mu}$, $\boldsymbol{\Gamma}$, ϕ, alternating until convergence; the APL-Adaptive Pairwise Log-likelihood estimator of SGLLVM is the result of the maximization of (3.11); finally, the Monte Carlo EM Pairwise Log-likelihood estimator named MCEMPL and described previously in Section 3.4. The maximization with an R optimization routine for differentiable functions `nlminb` is used to compute the GFA and LAML estimators. APL and MCEMPL estimators are based on the sum of integral approximations which induces irregularities, therefore we use the downhill-simplex optimization method for non-differentiable functions implemented in `optim`, which is another optimization routine available in R (R Development Core Team 2011). The GFA computations are implemented entirely in R. The other three estimators are implemented in C programming language and called from the R interface.

4.2. SIMULATIONS FOR GFA, LAML AND APL ESTIMATORS

We explore the performance of GFA, LAML and APL on finite samples by simulating 300 samples of size 100 issued from a SGLLVM with four manifest variables and one latent: one conditionally Bernoulli with loading $\gamma_1 = 2$ and location parameter $\mu_1 = 0.7$; and three manifests conditionally normal with loadings $(\gamma_2, \gamma_3, \gamma_4) = (1, 2, 3)$, zero location parameters $(\mu_2, \mu_3, \mu_4) = (0, 0, 0)$ and $(\psi_2, \psi_3, \psi_4) = (1, 1, 1)$ as uniquenesses.

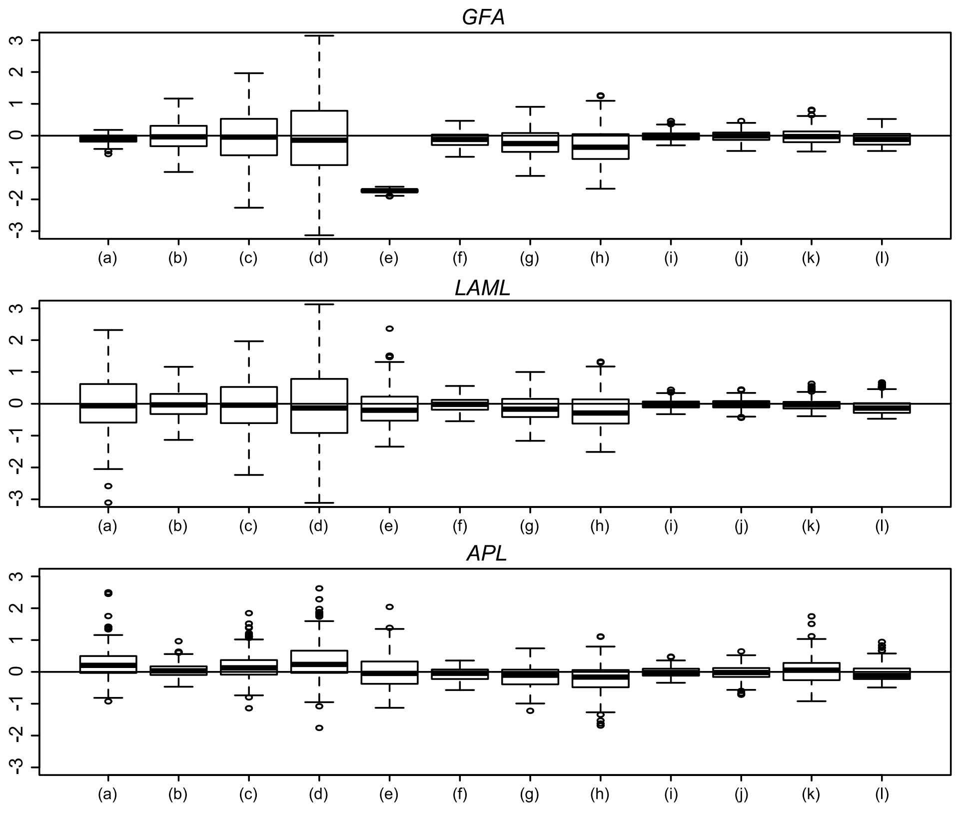

Figure 2. GFA, LAML and APL estimates for the SGLLVM with latent variable issued from $N(0, 1)$: (a)–(d) $\hat{\mu}_i - \mu_i$; (e)–(h) $\hat{\gamma}_i - \gamma_i$, $i = 1, 2, 3, 4$, (a) and (d) correspond to the binary manifest variable; (i)–(k) $\hat{\psi}_i - \psi_i$, $i = 2, 3, 4$; (l) $\hat{\lambda} - \lambda$.

Univariate scores of the latent variable are simulated at $S = 100$ regularly spaced locations on the interval $[0, 5]$ and correlated through locations according to the exponential model (2.5) and (2.6) with $\lambda = 0.6$. Their marginal distributions are taken to be: (1) symmetric unimodal normal $N(0, 1)$, and (2) asymmetric bimodal mixture of normals $0.7N(-1.5, 1) + 0.3N(3.5, 1)$. For each simulated data set we estimate the coefficients of the GFA, LAML and APL. We report the parameters estimates in Figures 2 and 3.

The GFA estimators of the loadings (top panel of Figure 2) exhibit biases. This is expected for the loading γ_1 corresponding to a conditionally Bernoulli manifest variable. However, even the loadings estimators $\hat{\gamma}_2$, $\hat{\gamma}_3$ and $\hat{\gamma}_4$ corresponding to conditionally normal manifest variables are slightly biased due to the wrong distributional specification of only one manifest variable. The estimation of the spatial correlation parameter λ also presents a slight bias.

In SGLLVM with true normal latent variables, the LAML estimators (middle panel of Figure 2) of the loadings $\gamma_1, \gamma_3, \gamma_4$ and the spatial correlation parameter λ are biased. This is due to the low dimension of the latent variable. As shown by Huber, Ronchetti, and Victoria-Feser (2004) in a simple GLLVM context, the convergence of LAML for higher-dimensional latent variables is much better.

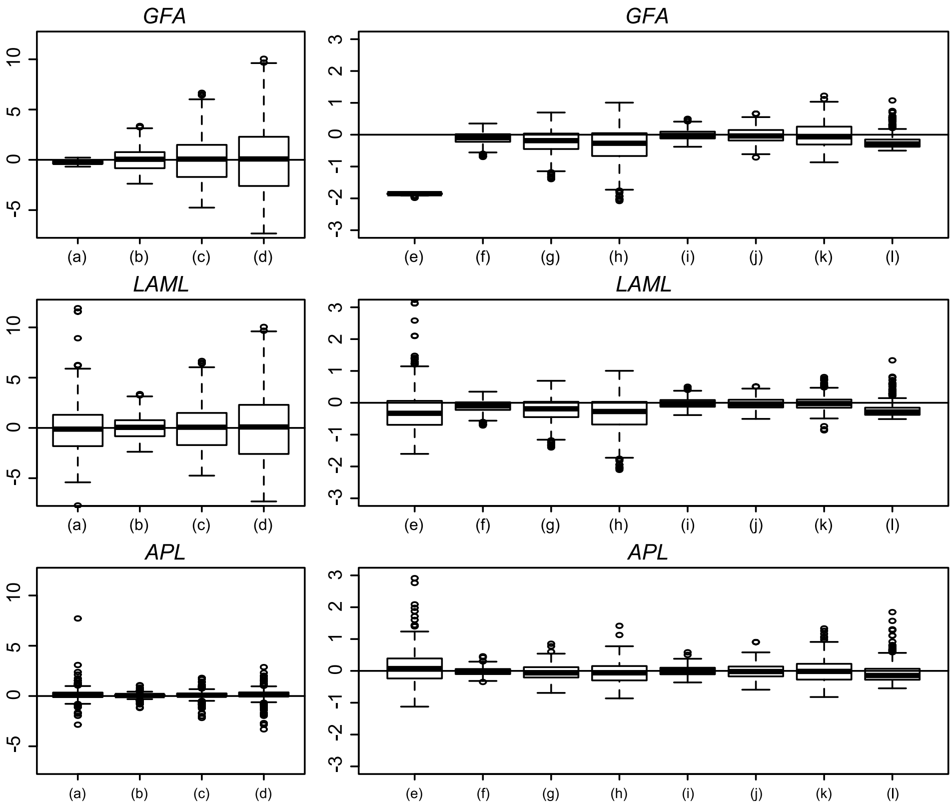

Figure 3. GFA, LAML and APL estimates for the SGLLVM with latent variable issued from $0.7N(-1.5, 1) + 0.3N(3.5, 1)$: (a)–(d) $\hat{\mu}_i - \mu_i$; (e)–(h) $\hat{\gamma}_i - \gamma_i$, $i = 1, 2, 3, 4$, (a) and (d) correspond to the binary manifest variable; (i)–(k) $\hat{\psi}_i - \psi_i$, $i = 2, 3, 4$; (l) $\hat{\lambda} - \lambda$.

The APL estimators of loadings and λ behave well (bottom panel of Figure 2). However, APL mean estimates are slightly biased.

The GFA, LAML and APL estimations of SGLLVM with latent variable issued from a mixture of Gaussians are shown in Figure 3. The GFA and LAML estimation of loadings $\gamma_2, \gamma_3, \gamma_4$ and uniquenesses ψ_2, ψ_3, ψ_4 corresponding to conditionally normal manifest variables is not sensitive to the incorrect specification of the latent variable distribution. An analogous conclusion for a simple GLLVM can be found in Ma and Genton (2010) and Irincheeva, Cantoni, and Genton (2012). However, the wrong specification of the latent variable distribution increases biases in the estimation of γ_1 and λ compared to the corresponding boxplots in Figure 2. In case of GFA, biases are due to the wrong distributional assumptions on both the latent variables and the manifest variable x_{s1} corresponding to the loading γ_1. In GFA the latent variables are assumed to be normally distributed and the manifest variables are supposed to be conditionally normal given the latent variables. The LAML estimator is also biased because of the incorrect normality assumption on the latent variable. LAML can be developed for cases of non-normal latent variable, for example in case of a latent variable issued from a skewed Student distribution. As already noticed in Section 2.2 when applying formula (2.14), the necessary condition for the LAML construc-

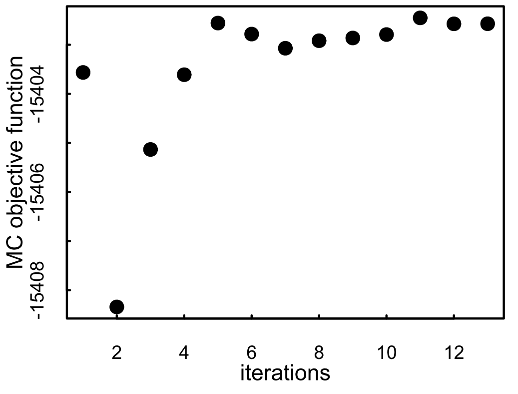

Figure 4. The values of the Monte Carlo approximated objective function of the EM-algorithm versus the iteration number.

tion is the unimodality of the latent variable distribution. This condition cannot be satisfied when the latent variable marginal distribution is a mixture of normals. APL is nearly unbiased for all the parameters.

From a computational point of view, the easiest and fastest estimators are however GFA and LAML. In the described context the computation of APL can take up to 5 hours depending on resources and initial points. In simulations we were successful in taking LAML estimators as initial value for APL, and GFA as initial value for LAML. The initial value for GFA is computed with a simple factor analysis.

4.3. A Simulated Example for the MCEMPL Estimator

We illustrate the performance of the MCEMPL on a simulated example of size 50. The example is issued from a SGLLVM with four manifest variables (one conditionally Bernoulli and three conditionally normal) and one latent variable with distribution $0.7N(-1.5, 1) + 0.3N(3.5, 1)$. The parameter values are $\mu_1 = 0.7, \gamma_1 = 2$ for the binary manifest variable; $(\mu_2, \mu_3, \mu_4) = (0, 0, 0)$, $(\gamma_2, \gamma_3, \gamma_4) = (1, 2, 3)$ and $(\psi_2, \psi_3, \psi_4) = (1, 1, 1)$ for the conditionally normal manifest variables. The APL estimated parameters are taken as initial values for simulations in MCEMPL estimation. Both values of δ_1 and ϵ in (3.18) are taken to be 0.05, implying a 95% confidence of the Monte Carlo approximation $\tilde{Q}_t$ to be within the distance $\delta_1 = 0.05$ from Q_t. The initial simulation size is 50. After a burn-in iteration with this size, the required simulation size is estimated to be 6202. We perform 13 iterations with simulation sizes increasing till 6202 as follows: $100, 200, 300, \ldots, 1000, 2000, \ldots, 5000, 6202$. The required simulation size after each iteration is about 6400, but the convergence criterion $|\tilde{Q}_t - \tilde{Q}_{t-1}| < \delta_1$ is met at the 13th iteration with absolute difference of 0.002. The Monte Carlo approximated values of the EM-algorithm objective function by iterations can be seen in Figure 4.

The GFA, LAML, APL and MCEMPL estimators of the parameters in the simulated data set with approximate 95% confidence intervals can be seen in Table 1. In this simulated sample, the MCEMPL and APL estimators are the best.

Table 1. Estimation results and approximate 95% confidence intervals (CI) on a simulated example of size 50 (standard errors for CI computation are obtained with improved numerical approximations of gradient and Hessian). Mean parameters are denoted by $(\mu_1, \mu_2, \mu_3, \mu_4)$, loadings by $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ and uniquenesses by (ψ_2, ψ_3, ψ_4) with true values in parentheses, μ_1 and γ_1 correspond to the binary manifest variable. The true latent variable distribution is $0.7N(-1.5, 1) + 0.3N(3.5, 1)$.

	GFA	GFA CI	LAML	LAML CI	APL	APL CI	MCEMPL	MCEMPL CI
$\mu_1(0.7)$	0.47	(0.21, 0.73)	0.31	(−1.48, 2.09)	0.30	(−0.70, 1.30)	0.50	(−0.50, 1.50)
$\mu_2(0)$	0.73	(−0.14, 1.60)	0.73	(−0.13, 1.59)	0.15	(−0.32, 0.62)	0.12	(−0.35, 0.59)
$\mu_3(0)$	1.64	(−0.01, 3.29)	1.64	(−0.01, 3.29)	0.35	(−0.57, 1.27)	0.37	(−0.55, 1.29)
$\mu_4(0)$	2.40	(−0.05, 4.85)	2.41	(−0.03, 4.85)	0.55	(−0.81, 1.91)	0.60	(−0.76, 1.96)
γ_1 (2)	0.18	(0.06, 0.30)	1.87	(0.75, 2.99)	1.76	(0.68, 2.84)	1.63	(0.55, 2.71)
$\gamma_2(1)$	1.13	(0.69, 1.57)	1.13	(0.70, 1.56)	0.94	(0.55, 1.33)	0.92	(0.53, 1.31)
$\gamma_3(2)$	2.35	(1.55, 3.15)	2.35	(1.56, 3.14)	1.97	(1.32, 2.62)	1.92	(1.27, 2.57)
$\gamma_4(3)$	3.34	(2.21, 4.47)	3.35	(2.23, 4.47)	2.77	(1.87, 3.67)	2.72	(1.82, 3.62)
$\psi_2(1)$	0.10	(−0.18, 0.38)	1.03	(0.73, 1.33)	1.15	(0.85, 1.45)	1.07	(0.77, 1.37)
$\psi_3(1)$	1.04	(0.70, 1.38)	1.03	(0.67, 1.39)	1.05	(0.62, 1.48)	1.02	(0.59, 1.45)
$\psi_4(1)$	1.07	(0.56, 1.58)	0.66	(0.20, 1.12)	0.81	(−0.01, 1.63)	0.73	(−0.09, 1.55)
$\lambda(0.6)$	0.59	(0.08, 1.10)	0.68	(0.35, 1.01)	0.32	(−0.14, 0.78)	0.33	(−0.13, 0.79)

Computationally, this simulated example, despite its low size, takes the same time as a third of all the simulations for GFA, LAML and APL. The simulations of distant observations are extremely time-consuming. The simulation acceptance ratio is very low for distant observations and drops down when the dimensionality of latent variables increases. For this reason, we do not fit MCEMPL to the NMMAPS data set in the following section. One possible solution for reducing the computational cost of MCEMPL is to include in the pairwise likelihood only neighbors within a chosen distance. As shown by Bevilacqua et al. (2012), the choice of this distance impacts the pairwise likelihood estimator and its variance.

5. AIR POLLUTION DATA ANALYSIS

Here we come back to our motivating example introduced in Section 1. Larger subsamples of NMMAPS were analyzed by Peng et al. (2005), Bell et al. (2005) and Welty and Zeger (2005), among others. All these articles explain the variable death.count, assumed conditionally Poisson distributed, by covariates such as trimmed mean of hourly ozone concentration, lags of ozone concentration, and possible confounders, such as temperature (tmpd), dew point temperature (dptp) or three-day lag temperature (rmtmpd).

We propose to assess the relationship among death counts, weather variables and pollution variables with a SGLLVM. It is widely known that high/low temperature is an important accelerating agent/inhibitor in multiple chemical reactions of human body, while high humidity and pollution usually increase the adaptation charge. Hence high/low temperature, high humidity and high pollution level are expected to be associated with high mortality, proportionally to the population size. The ensemble of these variables should form a latent variable associated with an exposure risk important for policy makers and preventive medicine.

Table 2. Estimation results and 95% confidence intervals (CI) of SGLLVM with $q = 1$ latent variable on NMMAPS data of size 93. Mean parameters are denoted by $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9)$, loadings by $(\gamma_1, \gamma_2, \gamma_3, \gamma_4, \gamma_5, \gamma_6, \gamma_7, \gamma_8, \gamma_9)$ and uniquenesses by $(\psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8, \psi_9)$. The numbering corresponds to the following order of the manifest variables: death.count, log.pop, tmpd, dptp, mxrh, mnrh, o3tmean, rmtmpd, rmdptp.

	GFA	GFA CI	LAML	LAML CI	APL	APL CI
μ_1	11.43	(8.62, 14.25)	2.43	(2.37, 2.49)	2.16	(−0.08,4.40)
μ_2	13.47	(13.30, 13.65)	13.46	(13.29, 13.63)	13.57	(10.67, 16.46)
μ_3	67.85	(66.16, 69.54)	67.87	(66.19, 69.54)	68.13	(61.73, 74.52)
μ_4	54.40	(52.11, 56.69)	54.47	(52.37, 56.57)	55.20	(51.42, 58.98
μ_5	85.28	(82.31, 88.25)	85.37	(82.66, 88.08)	86.61	(80.81, 92.41)
μ_6	46.78	(43.80, 49.76)	46.88	(44.12, 49.64)	48.31	(42.45, 54.16)
μ_7	11.23	(9.24, 13.22)	11.24	(9.44, 13.05)	11.75	(10.37, 13.14)
μ_8	−3.32	(−4.62, −2.02)	−3.47	(−4.36, −2.57)	−2.94	(−4.42, −1.46)
μ_9	−7.89	(−9.19, −6.59)	−8.13	(−8.25, −8.00)	−7.49	(−9.15, −5.82)
γ_1	−1.36	(−4.16, 1.44)	−0.02	(−0.08, 0.03)	0.95	(0.01, 1.89)
γ_2	−0.12	(−0.29, 0.05)	−0.07	(−0.22, 0.08)	0.97	(0.77, 1.16)
γ_3	5.48	(4.15, 6.81)	0.22	(−1.31, 1.74)	−0.77	(−0.89, −0.65)
γ_4	11.06	(9.55, 12.58)	0.80	(−1.09, 2.69)	−0.79	(−0.90, −0.69)
γ_5	9.66	(7.03, 12.29)	0.92	(−1.58, 3.42)	0.13	(0.06, 0.19)
γ_6	8.02	(5.28, 10.76)	0.74	(−1.75, 3.23)	1.49	(1.01, 1.97)
γ_7	−1.95	(−3.94, 0.03)	0.16	(−1.47, 1.79)	−0.80	(−1.00, −0.59)
γ_8	−3.71	(−4.85, −2.57)	−1.20	(−2.03, −0.38)	2.36	(−0.10, 4.81)
γ_9	−1.69	(−2.90, −0.47)	−1.69	(−1.81, −1.56)	2.31	(2.13, 2.49)
ψ_2	0.72	(0.52, 0.93)	0.68	(0.50, 0.86)	0.08	(0.07 , 0.09)
ψ_3	38.25	(30.17, 46.33)	68.04	(48.03, 88.06)	80.29	(77.03, 83.56)
ψ_4	0.57	(−12.71, 13.85)	106.69	(79.56, 133.82)	137.62	(109.93, 165.30)
ψ_5	117.64	(83.48, 151.80)	177.95	(133.50, 222.41)	223.62	(206.37, 240.88)
ψ_6	148.86	(105.91, 191.81)	184.63	(137.79, 231.46)	228.29	(209.08, 247.50)
ψ_7	92.22	(65.42, 119.01)	78.86	(59.73, 97.99)	106.53	(102.10, 110.96)
ψ_8	26.55	(19.20, 33.90)	19.17	(16.07,22.27)	49.92	(49.92, 49.93)
ψ_9	38.08	(27.11, 49.05)	0.37	(0.36, 0.38)	52.20	(−95.20, 199.59)
λ	0.14	(−0.30, 0.56)	0.02	(−0.19, 0.24)	9.03	(−13.70, 31.76)

We assume death.count to be conditionally Poisson given the latent variable. The other variables being continuous, we consider them to be conditionally normal. Locations of the cities are given in latitudes and longitudes. We applied the transformation latlong2grid of the R-package SpatialEpi for converting latitude/longitude coordinates to thousand kilometer-based grid coordinates. Then the coordinates were rescaled to be expressed in hundreds of kilometers. Finally, we fitted GFA, LAML and APL with $q = 1$ and $q = 2$ to the data. We had to drop the case $q = 3$, as it is revealed impossible to find initial parameter values such that the pairwise likelihood value is different from zero (when computing the APL estimator). GFA and LAML use the normality as distributional assumption on the latent variables, while in APL the latent variables are supposed to follow a mixture of normals.

The estimation results with approximate 95% confidence intervals are presented in Table 2 (with $q = 1$ latent variable at each location) and Table 3 (with $q = 2$ latent variables). For both $q = 1$ and $q = 2$, the GFA, LAML and APL estimated values are very different except for mean estimations.

Table 3. Estimation results and 95% confidence intervals (CI) of SGLLVM with $q = 2$ latent variables on NMMAPS data of size 93. Mean parameters are denoted by $(\mu_1, \mu_2, \mu_3, \mu_4, \mu_5, \mu_6, \mu_7, \mu_8, \mu_9)$, loadings for the first latent variable by $(\gamma_{11}, \gamma_{21}, \gamma_{31}, \gamma_{41}, \gamma_{51}, \gamma_{61}, \gamma_{71}, \gamma_{81}, \gamma_{91})$, for the second latent variable by $(\gamma_{12}, \gamma_{22}, \gamma_{32}, \gamma_{42}, \gamma_{52}, \gamma_{62}, \gamma_{72}, \gamma_{82}, \gamma_{92})$ and uniquenesses by $(\psi_2, \psi_3, \psi_4, \psi_5, \psi_6, \psi_7, \psi_8, \psi_9)$. The numbering corresponds to the following order of the manifest variables: death.count, log.pop, tmpd, dptp, mxrh, mnrh, o3tmean, rmtmpd, rmdptp.

	GFA	GFA CI	LAML	LAML CI	APL	APL CI
μ_1	11.52	(8.66, 14.38)	2.41	(2.34, 2.47)	2.43	(2.02, 2.85)
μ_2	13.47	(13.30, 13.65)	13.45	(13.28, 13.62)	13.48	(12.28, 14.67)
μ_3	67.79	(66.09, 69.49)	68.18	(66.61, 69.74)	68.20	(66.31, 70.03)
μ_4	54.40	(52.07, 56.72)	55.32	(53.46, 57.17)	55.35	(53.49, 57.21)
μ_5	85.33	(82.34, 88.31)	86.14	(83.80, 88.49)	86.17	(83.01, 92.33)
μ_6	46.83	(43.83, 49.82)	47.58	(45.36, 49.79)	47.60	(44.06, 51.14)
μ_7	11.20	(9.18, 13.22)	10.88	(8.94, 12.82)	10.91	(9.04, 12.78)
μ_8	−3.30	(−4.61, −1.98)	−4.12	(−4.28, −3.97)	−4.09	(−5.16, −3.03)
μ_9	−7.85	(−9.17, −6.54)	−8.23	(−8.27, −8.19)	−8.20	(−9.97, −6.43)
γ_{11}	−0.13	(−3.25, 3.01)	−0.02	(−0.04, −0.01)	0.02	(−0.04, 0.08)
γ_{21}	−1.25	(−5.97, 3.48)	−0.07	(−0.12, −0.03)	−0.11	(−0.25, 0.03)
γ_{31}	13.54	(11.13, 15.94)	0.25	(−0.17, 0.67)	0.73	(0.25, 1.21)
γ_{41}	−3.46	(−5.48, −1.43)	0.91	(0.41, 1.41)	2.42	(1.99, 3.57)
γ_{51}	−1.62	(−3.04, −0.21)	1.03	(0.40, 1.66)	2.72	(2.03, 3.41)
γ_{61}	−0.11	(−0.29, 0.07)	0.84	(0.24, 1.43)	2.23	(1.47, 2.99)
γ_{71}	8.52	(4.21, 12.83)	0.13	(−0.39, 0.66)	0.42	(0.13, 0.71)
γ_{81}	0.40	(−6.72, 7.52)	−1.30	(−1.34, −1.25)	−3.26	(−4.05, −2.47)
γ_{91}	−2.26	(−4.35, −0.16)	−1.72	(−1.73, −1.71)	−4.35	(−5.07, −3.63)
γ_{12}	−0.04	(−0.24, 0.16)	−0.03	(−0.05, −0.02)	−2.14	(−2.71, −1.57)
γ_{22}	7.09	(2.09, 12.08)	−0.02	(−0.06, 0.03)	0.11	(0.06, 0.16)
γ_{32}	12.03	(9.53, 14.53)	0.63	(0.19, 1.08)	6.19	(5.59, 6.78)
γ_{42}	−3.10	(−4.83, −1.37)	1.67	(1.14, 2.20)	15.86	(14.03, 17.69)
γ_{52}	−1.70	(−4.59, 1.19)	1.60	(0.93, 2.26)	15.17	(13.79, 16.55)
γ_{62}	8.16	(6.73, 9.59)	1.50	(0.87, 2.13)	14.24	(13.82, 15.28)
γ_{72}	1.35	(−6.48, 9.18)	−0.93	(−1.48, −0.38)	−8.38	(−9.71, −7.05)
γ_{82}	0.22	(−2.52, 2.96)	−1.13	(−1.17, −1.08)	−10.24	(−12.01, −8.47)
γ_{92}	−0.82	(−2.40, 0.76)	0.21	(0.20, 0.22)	2.23	(1.57, 2.89)
ψ_2	0.74	(0.52 , 0.96)	0.67	(0.48, 0.87)	0.70	(0.43, 0.97)
ψ_3	0.01	(0, 8.66)	58.16	(42.36, 73.97)	58.19	(43.06, 73.32)
ψ_4	0.19	(0, 8.80)	81.53	(56.12, 106.94)	81.56	(56.13, 106.99)
ψ_5	27.06	(15.99, 38.13)	130.84	(99.89, 161.79)	130.87	(100.05, 161.69)
ψ_6	70.77	(46.54, 94.99)	116.19	(92.45,139.94)	116.22	(92.69, 139.75)
ψ_7	87.09	(60.80, 113.37)	89.42	(62.14, 116.70)	89.45	(62.83, 116.07)
ψ_8	26.10	(18.37, 33.83)	0.60	(0.35, 0.85)	14.94	(13.21, 16.67)
ψ_9	38.53	(27.01, 50.06)	0.04	(0, 0.09)	13.15	(12.00, 14.31)
λ_1	0.01	(−0.45, 0.47)	0.003	(−26.40, 26.40)	0.03	(−2.71, 2.77)
λ_2	0.31	(−0.04, 0.66)	0.003	(−10.50, 10.50)	6.17	(1.51, 10.83)

In the case of $q = 1$ latent variable, all three estimations, by the presence of different signs in the loadings, make clear that one latent variable is not sufficient in this data set (opposite signs of loadings suggest that corresponding manifest variables measure something else than the considered latent variable, therefore another latent variable is necessary). The estimated marginal distribution of the latent variable is asymmetric unimodal corresponding to the univariate mixture of normals estimated approximately as $0.3N(0.37, 1) + 0.7N(-0.16, 1)$.

Table 3 presents the results of GFA, LAML and APL estimation of SGLLVM with $q = 2$ latent variables. Taking into consideration the simulation results, we interpret only the APL estimation. The APL first latent variable has small spatial correlation, while the second latent variable is strongly correlated in space.

Because of the different scales of manifest variables, the interpretation is not straightforward. The frequent in factor analysis technique of scaling the data before the analysis is not applicable to discrete data. To judge the importance of latent variables, correlations between manifest and latent variables should be computed. While correlations are easy to compute for conditionally normal manifest variables, there is no appropriate method for conditionally Poisson manifest variables. In addition, correlations for discrete random variables and continuous random variables are not strictly comparable. However, we compute the correlations for conditionally normal manifest variables and latent variables: $(\tilde{\gamma}_{21}, \tilde{\gamma}_{31}, \tilde{\gamma}_{41}, \tilde{\gamma}_{51}, \tilde{\gamma}_{61}, \tilde{\gamma}_{71}, \tilde{\gamma}_{81}, \tilde{\gamma}_{91}) \approx (-0.12, 0.07, 0.13, 0.14, 0.12, 0.03, -0.29, -0.71)$ and $(\tilde{\gamma}_{22}, \tilde{\gamma}_{32}, \tilde{\gamma}_{42}, \tilde{\gamma}_{52}, \tilde{\gamma}_{62}, \tilde{\gamma}_{72}, \tilde{\gamma}_{82}, \tilde{\gamma}_{92}) \approx (0.13, 0.63, 0.86, 0.79, 0.79, -0.66, -0.90, 0.37)$ (for the method, see Bartholomew, Knott, and Moustaki, 2011, p. 206). The role of the manifest variable death.count does not appear now to be as small as suggested by $\hat{\gamma}_{11} = 0.02$ and $\hat{\gamma}_{12} = -2.14$.

As a straightforward interpretation of the loadings is not available, we interpret the estimation results by looking at the relative importance of latent variables on manifest variables. We notice that the expected value of death.count depends on the latent variables exponentially, i.e. the expectation of death.count can be doubled by increasing $\gamma_{11}z_{s1} + \gamma_{12}z_{s2}$ by only $\log 2 \approx 0.3$, while the expectations of variables rmtmpd and rmdptp depend on $\gamma_{81}z_{s1} + \gamma_{82}z_{s2}$ and $\gamma_{91}z_{s1} + \gamma_{92}z_{s2}$ only linearly, even though the coefficients are high. Then, given the estimated values of γ_{11} and γ_{21}, the first latent variable has a small influence on death.count, but the influence of the second latent variable on death.count is quite important, suggesting this second latent variable as index of exposure risk is expected to be found. Both latent variables have very small influence on the variable o3tmean.

Finally, the first latent variable explains mostly the non-spatial variation of temperature lag and humidity, while the second latent variable explains the spatial variation of temperature lag, humidity, death counts and log of population size. In the considered data set, no conclusion can be reached on the impact of the ozone concentration (o3tmean) on mortality, because temperature and humidity are highly important confounders. Hence weather conditions are a more important source of variability than air pollution to explain all causes of mortality excluding accidents.

The joint probability density of the estimated latent variables, shown in Figure 5, is bimodal and slightly asymmetric; the group with negative values for both z_1 and z_2 corresponds to cities with a high combination of values for temperature lag, mortality and population.

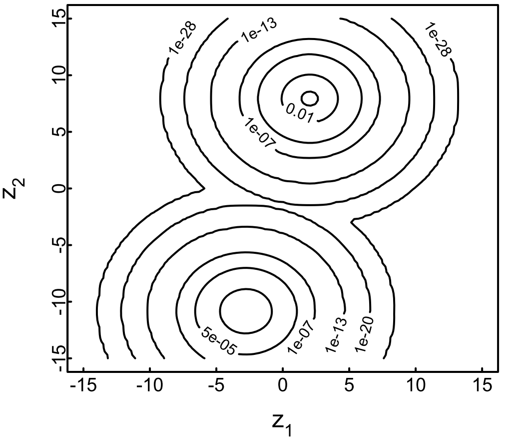

Figure 5. Contours of the APL estimated joint probability density function of the bivariate latent variables (i.e. $q = 2$) for the NMMAPS data.

6. DISCUSSION

In this article we proposed several estimation methods for spatial GLLVM. When the latent variables can be assumed to be normal, we introduced the Laplace approximation for the SGLLVM likelihood. In other cases we showed how to construct a multivariate distribution with Gaussian spatial dependence and predefined non-overlapping multivariate margins, which we took to be a mixture of normals with a common covariance matrix. This distribution has certainly its limitations, but is considerably more flexible than the Gaussian and offers the possibility of fitting marginally non-normal spatial data. We proposed a pairwise likelihood estimator for our SGLLVM and showed that an incorrect specification of the latent variable distribution affects the estimation of the parameters, in particular the parameter of spatial correlation. We believe our work is of interest for researchers and applications in multivariate spatial data and in latent variable modeling.

ACKNOWLEDGEMENTS

Genton's research was partially supported by NSF Grant DMS-1007504, and by Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

[Published Online August 2012.]

REFERENCES

Bardossy, A. (2006), "Copula-Based Geostatistical Models for Groundwater Quality Parameters," *Water Resources Research*, 42, 1–12.

Bartholomew, D. J., Knott, M., and Moustaki, I. (2011), *Latent Variable Models and Factor Analysis: A Unified Approach*, *John Wiley Series in Probability and Statistics*.

Bell, M., McDermott, A., Zeger, S., Samet, J., and Dominici, F. (2005), "Ozone and Short-Term Mortality in 95 US Urban Communities, 1987–2000," *Journal of the American Medical Association*, 292, 2371–2378.

Berntsen, J., Espelid, T. O., and Genz, A. (1991), "An Adaptive Algorithm for the Approximate Calculation of Multiple Integrals," *ACM Transactions on Mathematical Software*, 17, 437–451.

Bevilacqua, M., Gaetan, C., Mateu, J., and Porcu, E. (2012), "Estimating Space and Space-Time Covariance Functions: A Weighted Composite Likelihood Approach," *Journal of the American Statistical Association*, 107 (497), 268–280.

Booth, J. G., and Hobert, J. P. (1999), "Maximizing Generalized Linear Mixed Model Likelihoods With an Automated Monte Carlo EM Algorithm," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 61 (1), 265–285.

Brown, P. E., Diggle, P. J., and Henderson, R. (2003), "A Non-Gaussian Spatial Process Model for Opacity of Flocculated Paper," *Scandinavian Journal of Statistics*, 30 (2), 355–368.

Christensen, W. F., and Amemiya, Y. (2002), "Latent Variable Analysis of Multivariate Spatial Data," *Journal of the American Statistical Association*, 97 (457), 302–317.

Cox, D. R., and Reid, N. (2004), "A Note on Pseudo-Likelihood Constructed From Marginal Densities," *Biometrika*, 91, 729–737.

Eickhoff, J. C., Zhu, J., and Amemiya, Y. (2004), "On the Simulation Size and the Convergence of the Monte-Carlo EM Algorithm via Likelihood-Based Distances," *Statistics & Probability Letters*, 67, 161–171.

Gao, X., and Song, P. X.-K. (2010), "Composite Likelihood Bayesian Information Criteria for Model Selection in High-Dimensional Data," *Journal of the American Statistical Association*, 105 (492), 1531–1540.

Gelfand, A. E., Kottas, A., and MacEachern, S. N. (2005), "Bayesian Nonparametric Spatial Modeling With Dirichlet Process Mixing," *Journal of the American Statistical Association*, 100 (471), 1021–1035.

Genz, A., and Malik, A. (1980), "An Adaptive Algorithm for Numerical Integration Over an N-Dimensional Rectangular Region," *Journal of Computational and Applied Mathematics*, 6, 295–302.

Geweke, J. (1996), *Monte Carlo Simulation and Numerical, Integration*, Vol. 1. eds. H. M. Amman, D. A. Kendrick, and J. Rust. Amsterdam: Elsevier.

Gneiting, T., Genton, M. G., and Guttorp, P. (2007), "Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry," in *Statistical Methods for Spatio-Temporal Systems*, eds. B. Finkenstaedt, L. Held, and V. Isham, New York: Chapman & Hall/CRC, pp. 151–175.

Hogan, J. W., and Tchernis, R. (2004), "Bayesian Factor Analysis for Spatially Correlated Data, With Application to Summarizing Area-Level Material Deprivation From Census Data," *Journal of the American Statistical Association*, 99 (466), 314–324.

Huber, P. (2004). "Generalized Linear Latent Variables Models: Estimation, Inference and Empirical Analysis of Financial Data," unpublished doctoral dissertation, University of Geneva.

Huber, P., Ronchetti, E., and Victoria-Feser, M.-P. (2004), "Estimation of Generalized Linear Latent Variable Models," *Journal of the Royal Statistical Society, Series B*, 66, 893–908.

Irincheeva, I., Cantoni, E., and Genton, M. G. (2012), "Generalized Linear Latent Variable Models With Flexible Distribution of Latent Variables," *Scandinavian Journal of Statistics*, doi:*10.1111/j.1467-9469.2011.00777.x*.

Johnson, S. G., and Narasimhan, B. (2009), "Cubature: Adaptive Multivariate Integration Over Hypercubes" [Computer Software Manual].

Kazianka, H., and Pilz, J. (2011), "Bayesian Spatial Modeling and Interpolation Using Copulas," *Computers and Geosciences*, 37 (3), 310–319.

Li, H., Scarsini, M., and Shaked, M. (1996), "Linkages: A Tool for the Construction of Multivariate Distributions With Given Nonoverlapping Multvariate Margins," *Journal of Multivariate Analysis*, 56, 20–41.

Lopes, H. F., Gamerman, D., and Salazar, E. (2011), "Generalized Spatial Dynamic Factor Models," *Computational Statistics & Data Analysis*, 55, 1319–1330.

Ma, Y., and Genton, M. G. (2010), "Explicit Estimating Equations for Semiparametric Generalized Linear Latent Variable Models," *Journal of the Royal Statistical Society, Series B*, 72, 475–495.

Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus With Applications in Econometrics and Statistics*, New York: Wiley.

Minozzo, M., and Fruttini, D. (2004), "Loglinear Spatial Factor Analysis: An Application to Diabetes Mellitus Complications," *Environmetrics*, 15 (5), 423–434.

Montanari, A., and Viroli, C. (2010a), "Heteroscedastic Factor Mixture Analysis," *Statistical Modelling*, 10 (4), 441–460.

——— (2010b), "A Skew-Normal Factor Model for the Analysis of Student Satisfaction Towards University Courses," *Journal of Applied Statistics*, 43, 473–487.

Nelsen, R. B. (2006), *An Introduction to Copulas* (2nd ed.), Berlin: Springer.

Peng, R., Dominici, F., Pastor-Barriuso, R., Zeger, S., and Samet, J. (2005), "Seasonal Analyses of Air Pollution and Mortality in 100 U.S. Cities," *American Journal of Epidemiology*, 161 (6), 585–594.

R Development Core Team (2011). "R: A Language and Environment for Statistical Computing" [Computer software manual], Vienna, Austria. Available from *http://www.R-project.org*.

Reich, B. J., and Bandyopadhyay, D. (2010), "A Latent Factor Model for Spatial Data With Informative Missingness," *Annals of Applied Statistics*, 4 (1), 439–459.

Rue, H., and Held, L. (2005), *Gaussian Markov Random Fields: Theory and Applications*, New York: Chapman & Hall/CRC.

Sammel, M., Ryan, L., and Legler, J. (1997), "Latent Variable Models for Mixed Discrete and Continuous Outcomes," *Journal of the Royal Statistical Society, Series B*, 59, 667–678.

Shun, Z., and McCullagh, P. (1995), "Laplace Approximation of High Dimensional Integrals," *Journal of the Royal Statistical Society, Series B*, 57 (4), 749–760.

Varin, C., Høst, G., and Skare, Ø. (2005), "Pairwise Likelihood Inference in Spatial Generalized Linear Mixed Models," *Computational Statistics & Data Analysis*, 49 (4), 1173–1191.

Varin, C., Reid, N., and Firth, D. (2011), "An Overview of Composite Likelihood Methods," *Statistica Sinica*, 21, 5–42.

Varin, C., and Vidoni, P. (2005), "A Note on Composite Likelihood Inference and Model Selection," *Biometrika*, 92 (3), 519–528.

Verbeke, G., and Lesaffre, E. (1996), "A Linear Mixed-Effects Model With Heterogeneity in the Random-Effects Population," *Journal of the American Statistical Association*, 91 (433), 217–221.

Wang, F., and Wall, M. M. (2003), "Generalized Common Spatial Factor Model," *Biostatistics*, 4 (4), 569–582.

Welty, L., and Zeger, S. (2005), "Are the Acute Effects of PM10 on Mortality in NMMAPS the Result of Inadequate Control for Weather and Season? A Sensitivity Analysis Using Flexible Distributed Lag Models," *American Journal of Epidemiology*, 162 (1), 80–88.

Zhu, J., Eickhoff, J., and Yan, P. (2005), "Generalized Linear Latent Variable Models for Repeated Measures of Spatially Correlated Multivariate Data," *Biometrics*, 61 (3), 674–683.