



Proceedings of 3rd Seminar on Spatial Statistics and Its Applications



23-24 October 2019

In the Name of God



**Proceedings of the 3rd Seminar
on Spatial Statistics and Its Applications**

23-24 October 2019

University of Zanjan
Zanjan, Iran.

Proceedings of the 3rd Seminar on Spatial Statistics and Its Applications

Address: Faculty of Sciences, Department of Statistics
University of Zanjan, Zanjan, Iran

Phone: 024 – 33054042

E-mail: spatial3@znu.ac.ir

Website: spatial3.znu.ac.ir

Preface

In the Name of GOD

It is our great pleasure to welcome the participants to the Third Seminar on Spatial Statistics and Its Applications (Spatial3), held at the University of Zanjan from 23-24 October 2019. We hope this seminar provides an excellent forum for scientific communications and collaborations between professors, researchers, and experts in spatial statistics, in Iran, and promote communications and collaborations between the researchers in this area and those from industry and other organizations. With no doubt, we could not organize this seminar without the kind collaboration of our colleagues with Iranian Statistical Society and the Center of Excellence in Analysis of Spatio-Temporal Correlated Data at Tarbiat Modares University in Tehran, Iran. On behalf of the organizing and scientific committees, we would like to express our sincere appreciation to all our colleagues and students in the department of statistics, and personnel of the University of Zanjan.

Secretary of the 3rd Seminar on
Spatial Statistics and Its Applications
October 2019
University of Zanjan

Scientefic Committee

- | | |
|---|---------------------------------------|
| 1- Dr. Ali Mohammadian Mosamam (Chairman) | University of Zanjan |
| 2- Dr. Ali Aghamohammadi | University of Zanjan |
| 3- Dr. Hossein Baghishani | Shahrood University of Technology |
| 4- Dr. Omid Karimi | Semnan University |
| 5- Dr. Afshin Fallah | Emam Khomini International University |
| 6- Dr. Abbas Rasouli | University of Zanjan |
| 7- Dr. Mohsen Mohammadzadeh | Tarbiat Modarres University |

Excutive Committee

- | | |
|-------------------------------------|----------------------|
| 1- Dr. Ali Aghamohammadi (Chairman) | University of Zanjan |
| 2- Dr. Farazollah Mirzapour | University of Zanjan |
| 3- Dr. Abbas Rasouli | University of Zanjan |
| 4- Dr. Ali Mohammadian Mosamam | University of Zanjan |
| 5- Dr. Hadi Emami | University of Zanjan |
| 6- Dr. Omid Khademno | University of Zanjan |
| 7- Dr. Mehri Javanian | University of Zanjan |

Contents

Application of Multivariate Spatial Correlation Model in Seismic Hazard Analysis Considering Anisotropy (<i>Morteza Abbasnejad, Morteza Bastami, Afshin Fallah</i>)	1
Spatial Dependence of Extreme Damage Probabilities in the Lifeline Networks Using the Max-stable Process (<i>Shahin Borzoo, Morteza Bastami, Afshin Fallah</i>)	13
Sensitivity Analysis In Elliptical Spatial Linear Measurement Error Models (<i>Hadi Emami</i>)	20
A Spatial Heterogeneity Mixed Model with Skew-Elliptical Distributions (<i>Mohadeseh Alsadat Farzammehr</i>)	31
Seismic Hazard Analysis of Lifeline Networks Considering Anisotropy of Earthquake Ground-motion Intensity (<i>Alireza Garakaninezhad, Morteza Bastami</i>)	51
Clustering Hyperspectral Images Using Spatial Dependencies (<i>Mousa Golalizadeh, Hamideh Sadat Fatemi Ghomi</i>)	60
Utilization of Pearson Correlation to Grab Spatial Correlation in the Absence of Information (<i>Lida Kalhori Nadrabadi</i>)	69
Approximate Composite Marginal Likelihood Inference in SGLM Models with Skew Normal Latent Variables (<i>Omid Karimi, Fatemeh Hosseini</i>)	78
Analysis of Spatial Data with Non-Ignorable Missingness (<i>Mohsen Mohammadzadeh, Samira Zahmatkesh</i>)	88
Modeling of Spatially Clustered Survival HIV/AIDS Data in the Presence of Competing Risks Setting: A Bayesian Approach (<i>Somayeh Momenyan and Amir Kavousi</i>)	101

On Spatial Skew-Gaussian Process and Its Applications in Survival Analysis

(Kiomars Motarjem, Mohsen Mohammadzadeh) **109**



Application of Multivariate Spatial Correlation Model in Seismic Hazard Analysis Considering Anisotropy

Morteza Abbasnejad^{1*}, Morteza Bastami¹, Afshin Fallah²

¹International Institute of Earthquake Engineering and Seismology, Tehran, Iran.

²Imam Khomeini International University, Ghazvin, Iran.

Abstract:

Spatial correlation and cross-correlation of earthquake intensity measures (IMs) is an important subject in seismic hazard and risk assessment of spatially distributed lifeline systems. Spatial statistic proposes effective methods to investigate the spatial correlations of earthquake IMs. Among the different problems in this subject, current study focuses on spatial correlations of multiple earthquake IMs considering anisotropy. In this regard, latent dimension method is implemented in this study to construct valid cross-covariance matrix considering three different earthquake IMs : peak ground acceleration, peak ground velocity, and peak ground displacement. Data of three earthquake events occurred in Japan are used and results are presented.

Keywords: Spatial correlations, Earthquake, Latent dimensions, Anisotropy.

Mathematics Subject Classification (2010): 62P30, 62N01, 62H11.

1 Introduction

Spatially distributed lifeline network systems such as electrical power transmission system, water distribution network, gas distribution system, urban road network are critical systems which their operation is crucial for communities. Earthquake is one of the catastrophic events which treat the operation of the lifeline systems and safety of communities. Each of above mentioned networks composed of elements that are vulnerable to earthquake event and consequently distribution of vulnerable

*Speaker: m.abbasnejad@iiees.ac.ir

elements across the network cause vulnerability of network system.

In an engineering perspective, vulnerability of civil and structural elements relates to earthquake intensity measures (IM) like peak ground acceleration (PGA), peak ground velocity (PGV) and peak ground displacement (PGD). Earthquake engineers use ground motion prediction equations (GMPE) to predict earthquake IMs in locations of interested. GMPEs (see equation (1.1)) which are statistical models based on previously recorded IMs in past earthquake events are used to predict earthquake IMs as dependent variables considering independent variables like magnitude of earthquake event (M), closest distance between point of interest and fault rupture plan (R_{rup}), soil properties of point of interest (V_{s30}), etc.

$$\ln(Y_{i,j}) = \ln(\bar{Y}_{i,j}) + \delta_{i,j} \quad (1.1)$$

In equation (1.1), Y is earthquake intensity measure at location i given the earthquake event j , $\ln(\bar{Y}_{i,j})$ presents predicted value of interested earthquake IM (as defined in equation (1.2)) and $\delta_{i,j}$ is residual term (as defined in equation (1.3)).

$$\ln(\bar{Y}_{i,j}) = f(M, R_{rup}, V_{s30}, \dots) \quad (1.2)$$

$$\delta_{i,j} = \eta_j + \varepsilon_{i,j} \quad (1.3)$$

In equation (1.3) η_j known as inter-event residual which represents event to event variability and is equal for all points in an earthquake event j . Moreover, $\varepsilon_{i,j}$ known as intra-event residual in point i given the event j which represents within event variability. Considering n spatially distributed points, [Jayaram and Baker \(2008\)](#) showed that the vector of intra-event residuals of $\varepsilon_j = (\varepsilon_{1j}, \varepsilon_{2j}, \dots, \varepsilon_{nj})$ follows multivariate normal distribution with zero mean and covariance matrix of \mathbf{C}_e . More detail can be found in ([Jayaram and Baker, 2009](#)). Consequently, knowing covariance matrix of ε_j one can simulate correlated values of intra-event residuals at n spatially distributed points in earthquake event j .

Geostatistical tools are widely used in previous research works to define valid covariance matrix of intra-event residuals. In the case of univariate random field where one of different earthquake are of interest, semivariogram function is implemented in different research work and different predictive models are proposed based on recorded data in past earthquake events ([Boore et al., 2003](#); [Wang and Takada, 2005](#); [Goda and Hong, 2008](#); [Goda and Atkinson, 2009](#); [Jayaram and Baker, 2009](#); [Esposito and Iervolino, 2011](#)). In these research work exponential model of equation

(1.4) is fitted on empirical semivariogram values of intra-event residuals of earthquake IMs and consequent range values (b) are reported.

$$\gamma(h) = a \left[1 - \exp\left(-\frac{3h}{b}\right) \right] \quad (1.4)$$

In equation (1.4), $\gamma(h)$ is fitted semivariogram function, a is known as sill and b is known as range (the distance beyond which the values can be considered uncorrelated). Typical form of equation (1.4) is presented in Figure 1.

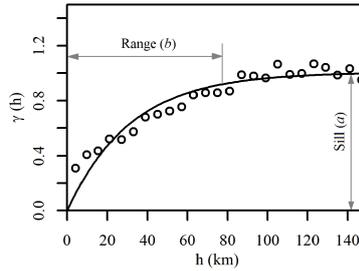


Figure 1: Sample of fitted exponential semivariogram form of equation (1.4)

Defining valid covariance matrix in the case of multivariate random fields is not as straightforward as univariate random fields. This is because that defined covariance matrix should be non-negative definite (Genton and Kleiber, 2015) while using above-mentioned method does not guarantees the covariance matrix of multivariate randomfield to be in valid form (Wang and Du, 2013). Considering multiple earthquake IMs, previous research works used statistical tools like *Linear Model of Coregionalization* (LMC) (Loth and Baker, 2013; Wang and Du, 2013) or *Principal Component Analysis* (PCA) (Markhvida et al., 2018).

All above mentioned researches are based on assumption of isotropy. Garakanezhad and Bastami (2017) using statistical test of Bowman and Crujeiras (2013) showed that assumption of isotropy of intra-event residuals of earthquake IMs is not valid in general. Consequently, it is important to consider anisotropy in spatial correlation and cross-correlation of earthquake IMs. The matter of spatial correlation of multivariate random fields is challenging issue in statistic and using above mentioned methods (LMC or PCA) will not lead us to construct valid cross-covariance matrixes for an anisotropic multivariate random field. For this reason, current study implements *Latent Dimensions Method* introduced by Apanasovich and Genton (2010) to construct valid cross-covariance functions for multivariate random field of PGA, PGV and PGD.

2 Spatial correlations in multivariate random fields

Considering $\{\varepsilon'(\mathbf{s}) = [\varepsilon'_1(\mathbf{s}), \dots, \varepsilon'_k(\mathbf{s})]'_{n \times k} : \mathbf{s} \in \mathbb{R}^2\}$ is second-ordered stationary multivariate random process, with $0_{n \times k}$ mean matrix and a marginal-variogram of:

$$2\gamma_{\alpha\alpha}(\mathbf{h}) = \mathbb{E} \left[(\varepsilon'_\alpha(\mathbf{s} + \mathbf{h}) - \varepsilon'_\alpha(\mathbf{s}))^2 \right], \quad \alpha = 1, \dots, k \quad (2.1)$$

then the cross-variogram can be defined as (Cressie, 1993):

$$2\gamma_{\alpha\beta}(\mathbf{h}) = \mathbb{E} [(\varepsilon'_\alpha(\mathbf{s} + \mathbf{h}) - \varepsilon'_\alpha(\mathbf{s}))(\varepsilon'_\beta(\mathbf{s} + \mathbf{h}) - \varepsilon'_\beta(\mathbf{s}))], \quad \alpha, \beta = 1, \dots, k. \quad (2.2)$$

In equations (2.1) and (2.2), \mathbf{h} is separation vector. Moreover, the cross-covariance can be defined as:

$$C_{\alpha\beta}(\mathbf{h}) = \text{Cov}(\varepsilon'_\alpha(\mathbf{s}), \varepsilon'_\beta(\mathbf{s} + \mathbf{h}))$$

$$= \mathbb{E} [(\varepsilon'_\alpha(\mathbf{s}) - \mathbb{E}[\varepsilon'_\alpha]) (\varepsilon'_\beta(\mathbf{s} + \mathbf{h}) - \mathbb{E}[\varepsilon'_\beta])], \quad \alpha, \beta = 1, \dots, k. \quad (2.3)$$

Empirical estimator of cross-covariance is as (Genton and Kleiber, 2015):

$$\hat{C}_{\alpha\beta}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{N(\mathbf{h})} \left\{ (\varepsilon'_\alpha(\mathbf{s}_i) - \bar{\varepsilon}'_\alpha) (\varepsilon'_\beta(\mathbf{s}_i) - \bar{\varepsilon}'_\beta) \right\} \quad (2.4)$$

where $N(\mathbf{h}) \equiv \{(i, j) : \mathbf{s}_i - \mathbf{s}_j = \mathbf{h}\}$ and $|N(\mathbf{h})|$ is the number of distinct elements of $N(\mathbf{h})$ and $\bar{\varepsilon}'_k = \frac{1}{n} \sum_{i=1}^n \varepsilon'_k(\mathbf{s}_i)$

3 Latent dimensions method for construction of valid cross-covariance functions

As it is discussed in previous section, considering anisotropy make it impossible to use models that are introduced in previous studies (Loth and Baker, 2013; Wang and Du, 2013; Markhvida et al., 2018) for construction of valid cross-covariance matrix of function. Apanasovich and Genton (2010) proposed an innovative approach based on latent dimensions method based on exiting covariance models of univariate random fields to define closed form cross-covariance function. The key idea is to represent a vectors components as points in k dimensional space and converting multivariate problem to multidimensional univariate one. In this regard each component α of multivariate random field $\varepsilon'(\mathbf{s})$ considered as a point of univariate random field in

k dimensional space. Based on these latent dimensions, $C_{\alpha\beta}(\mathbf{s}_1, \mathbf{s}_2) : \mathbf{s}_1, \mathbf{s}_2 \in \mathbb{R}^n$ becomes as $C((\mathbf{s}_1, \xi_\alpha), (\mathbf{s}_2, \xi_\beta))$, a covariance of a univariate random field which its arguments are from \mathbb{R}^{n+k} space instead of \mathbb{R}^n . In the case of current study n equals to 2 (because data have been recorded in 2 dimensional space) and k considered as 1, so a 2-dimensional three-variate random field of normalized intra-event residuals of PGA, PGV and PGD is represented as a 3-dimensional univariate random field. Consequently using valid covariance models of univariate random fields, the covariance matrix is guaranteed to be non-negative definite. In current study, the cross covariance function form of equation (3.1) which is proposed by [Apanasovich and Genton \(2010\)](#) based on latent dimension method is implemented.

$$C_{\alpha\beta}(\mathbf{h}) = C(\mathbf{h}, v_{\alpha\beta} - \mathbf{\Gamma}_\xi \mathbf{h}) = \frac{\sigma_{\alpha\beta}}{|v_{\alpha\beta} - \gamma \omega^T \mathbf{h}| + 1} \exp \left\{ -\frac{a \|\mathbf{h}\|}{(|v_{\alpha\beta} - \gamma \omega^T \mathbf{h}| + 1)^{1/2}} \right\} \quad (3.1)$$

In equation (3.1) $\alpha, \beta = 1, 2, 3$ represent ε'_{PGA} , ε'_{PGV} and ε'_{PGD} , normalized intra-event residuals of PGA, PGV and PGD respectively which are the components of multivariate random field in original space, $\mathbf{h} = \mathbf{s}_i - \mathbf{s}_j$ is relative position vector of points i and j , $\sigma_{\alpha\beta}$ is variance parameter and $\omega^T = \{\omega_1, \omega_2\}^T$ is a 2-dimensional vector such that $\omega^T \omega = 1$. In this equation, $\arctan(\omega_2/\omega_1)$ shows the anisotropy direction and $\gamma \geq 0$ defines anisotropy ratio.

Normalized intra-events residuals are defined as (3.2) assuming that the standard deviation of intra-event residuals are independent of location ([Jayaram and Baker, 2009](#); [Du and Wang, 2013](#); [Garakaninezhad and Bastami, 2017](#)) and consequently they have unite standard deviation.

$$\varepsilon'_{ij} = \frac{\varepsilon_{ij}}{\sigma} = \frac{\ln(Y_{ij}) - \ln(\overline{Y_{ij}})}{\sigma} \quad (3.2)$$

Instead of using latent dimension, $\xi_1 = \{\xi_{11}, \xi_{12}, \xi_{13}\}^T$ it is possible to treat with latent distances $v_{\alpha\beta} = \xi_{1\alpha} - \xi_{1\beta}$, $\alpha, \beta = 1, 2, 3$. The larger latent distance $v_{\alpha\beta}$ indicates the smaller cross-correlation between components α and β . The vector ω and the parameter γ determine anisotropy direction and anisotropy extend (ratio) respectively. Moreover, latent distance parameter v determine the amount of correlation between different components (variable) so that larger value of v shows smaller amount of correlation. In this regard v equals to 0 for marginal-covariance models.

Table 1: Information of earthquake events used in this study

Earthquake Name	Year	Magnitude	Location	Number of Selected Records	Fault Mechanism
Chuetsu	2007	6.8	Japan	613	Reverse
Iwate	2008	6.9	Japan	367	Reverse
Tottori	2000	6.6	Japan	414	Strike-Slip

4 Application of latent dimensions method on earthquake data

In current study, the method proposed by [Apanasovich and Genton \(2010\)](#) is implemented on normalized intra-event residuals of earthquake IMs PGA, PGV and PGD. Recorded data of three earthquake events occurred in Japan are implemented in this study. Selected earthquake events have been recorded in enough record stations so that they are suitable for spatial analysis. Table 1 presents brief information of implemented earthquake events. Intra-event residuals of earthquake IMs are calculated using [Campbell and Bozorgnia \(2014\)](#) ground motion prediction equations for PGA and PGV and [Campbell and Bozorgnia \(2008\)](#) ground motion prediction equation for PGD. Calculated intra-event residuals are normalized using equation (3.2).

Non-linear least square method is implemented in current study to estimate parameters of model of equation (3.1). For this purpose, model of equation (3.1) is fitted to estimated marginal and cross-covariance values of intra-event residuals of PGA, PGV and PGD which is calculate for different lag distances and different directions using equation (2.4). Contour plot of fitted marginal and cross-covariance models to normalized intra-event residuals of PGA, PGV and PGD of selected earthquake events are presented in Figures 2, 3 and 4. Moreover, estimated parameter values of implemented model (equation (3.1)) are presented in Table 2 through Table 7.

Azimuth in Table 2 through Table 7 is clockwise angle between maximum range direction (anisotropy angle) and north baseline and is calculated based on $\arctan(\omega_1/\omega_2)$. Also the values in parentheses shows 95% confidence intervals. In these tables latent distance (v) is considered as 0 for marginal-covariance models. This value is estimated using non-linear least square method for cross-covariance models of PGA-PGV and PGV-PGD. v value for cross-covariance models of PGA-PGD are calculated based on estimated v value of cross-covariance models of PGA-PGV and PGV-PGD considering $v_{13} = v_{12} + v_{23}$ ([Apanasovich and Genton, 2010](#)). Larger latent distance values indicates the smaller cross-correlation between components

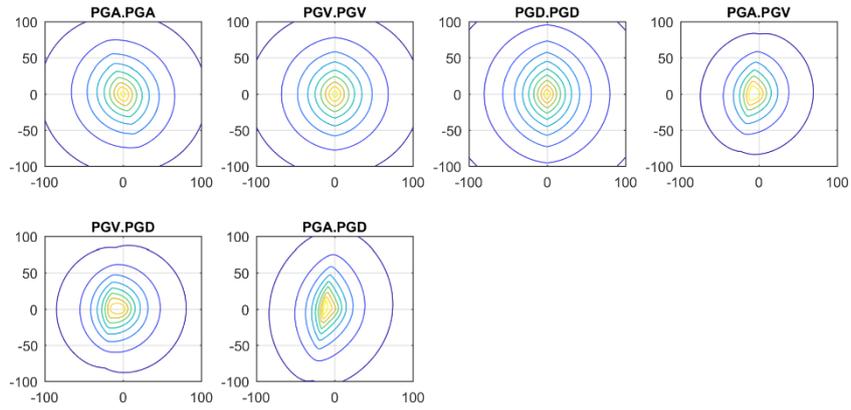


Figure 2: Contour plot of fitted marginal and cross-covariance models of the Chuetsu earthquake. (values on horizontal and vertical axes are lag distance (h) in E-W and N-S directions) in km.

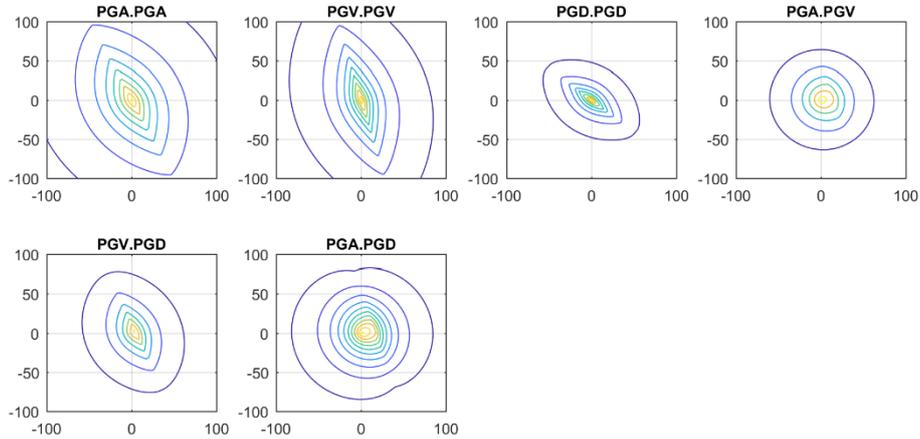


Figure 3: Contour plot of fitted marginal and cross-covariance models of the Iwate earthquake. (values on horizontal and vertical axes are lag distance (h) in E-W and N-S directions) in km.

(Apanasovich and Genton, 2010) and considering ν values in mentioned tables it can be concluded that PGA and PGD with average ν value of 1.128 has smaller correlations than PGA-PGV with average ν value of 0.0.508 and PGV-PGD with average ν value of 0.473.

Parameter α relates to maximum rang of correlations. Smaller α value shows larger maximum rang and larger α value shows smaller maximum correlation rang. Table 8 presents average value of parameter α in different marginal and cross-covariance models. As can be seen in this table, cross-covariance models has generally lower maximum correlation rang than marginal-covariance models. It should be noted that more earthquake events analysis are needed to conclude more accurately.

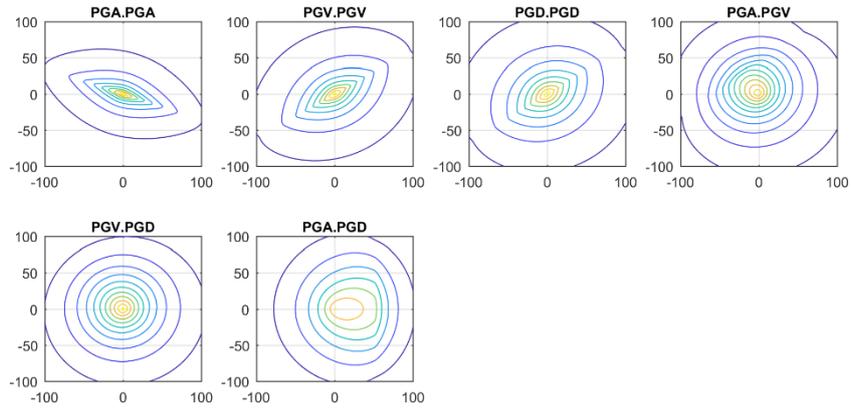


Figure 4: Contour plot of fitted marginal and cross-covariance models of the Tottori earthquake. (values on horizontal and vertical axes are lag distance (h) in E-W and N-S directions) in km.

Table 2: Parameter estimation of marginal-covariance model of intra-event residuals of PGA

Earthquake		ν	α	γ	σ	Azimuth ^a	R^2	RMSE
			0.021	0.012	1.00			
Chuetsu	0	(0.019, 0.023) ^b	0.015	(0.004, 0.019)	(0.96, 1.04)	165.1	0.92	0.079
Iwate	0	(0.009, 0.020)	0.021	(0.011, 0.047)	(0.89, 1.01)	154.2	0.83	0.109
Tottori	0	(0.016, 0.027)	0.021	(0.008, 0.016)	(0.94, 1.01)	109.7	0.95	0.062

^a Azimuth is calculated based on $\arctan(\omega_1/\omega_2)$

^b Values in parentheses indicates 95% confidence interval

Table 3: Parameter estimation of marginal-covariance model of intra-event residuals of PGV

Earthquake		ν	α	γ	σ	Azimuth	R^2	RMSE
			0.021	0.010	1.00			
Chuetsu	0	(0.017, 0.024)	0.017	(0.001, 0.019)	(0.93, 1.06)	0.0	0.84	0.110
Iwate	0	(0.010, 0.024)	0.020	(0.034, 0.089)	(0.99, 1.11)	164.6	0.90	0.104
Tottori	0	(0.018, 0.023)	0.020	(0.028, 0.050)	(0.95, 1.03)	58.0	0.94	0.065

Anisotropy direction is another matter that can be investigated using implemented method. Taking into consideration of Figures 2, 3 and 4, it can be concluded that different marginal and cross-covariance models has almost same direction of anisotropy, albeit assessed earthquake events have different amount of anisotropy which can be inferred by parameter γ .

Table 4: Parameter estimation of marginal-covariance model of intra-event residuals of PGD

Earthquake Name	v	α	γ	σ	Azimuth	R^2	RMSE
Chuetsu	0	0.018 (0.016, 0.020)	0.013 (0.006, 0.020)	1.14 (1.08, 1.19)	0.0	0.93	0.084
Iwate	0	0.034 (0.023, 0.046)	0.090 (0.049, 0.131)	0.95 (0.89, 1.01)	128.7	0.87	0.099
Tottori	0	0.020 (0.018, 0.022)	0.019 (0.010, 0.0268)	1.00 (0.95, 1.04)	55.2	0.95	0.068

Table 5: Parameter estimation of cross-covariance model of intra-event residuals of PGA and PGV

Earthquake Name	v	α	γ	σ	Azimuth	R^2	RMSE
Chuetsu	0.376	0.029 (0.023, 0.034)	0.037 (0.014, 0.060)	1.07 (0.87, 1.27)	8.1	0.94	0.091
Iwate	0.569	0.033 (0.026, 0.040)	0.015 (-0.003, 0.033)	0.83 (0.63, 1.03)	163.8	0.79	0.075
Tottori	0.58	0.029 (0.024, 0.034)	0.017 (0.003, 0.031)	0.95 (0.85, 1.05)	53.1	0.73	0.113

Table 6: Parameter estimation of cross-covariance model of intra-event residuals of PGV and PGD

Earthquake Name	v	α	γ	σ	Azimuth	R^2	RMSE
Chuetsu	0.490	0.031 (0.027, 0.034)	0.029 (0.021, 0.037)	0.68 (0.44, 0.91)	8.1	0.97	0.044
Iwate	0.230	0.026 (0.019, 0.033)	0.049 (0.021, 0.077)	0.83 (0.73, 0.92)	157.5	0.85	0.089
Tottori	0.699	0.031 (-0.016, 0.078)	0.006 (-0.021, 0.033)	0.98 (-2.03, 3.98)	53.1	0.70	0.115

Table 7: Parameter estimation of cross-covariance model of intra-event residuals of PGA and PGD

Earthquake Name	v	α	γ	σ	Azimuth	R^2	RMSE
Chuetsu	0.866	0.026 (0.021, 0.030)	0.066 (0.054, 0.077)	1.4 (1.31, 1.49)	0.0	0.95	0.096
Iwate	1.24	0.062 (0.047, 0.076)	0.059 (0.007, 0.109)	1.28 (1.11, 1.45)	171.9	0.70	0.123
Tottori	1.279	0.021 (0.016, 0.025)	0.025 (0.013, 0.036)	0.92 (0.80, 1.03)	0.0	0.70	0.105

Conclusion

Spatial correlation of earthquake intensity measures (IMs) are of interest especially when spatially distributed life-line systems of portfolio of buildings is considered.

Table 8: Average value of parameter α in different marginal and cross-covariance models

PGA	PGV	PGD	PGA-PGV	PGV-PGD	PGA-PGD
0.019	0.019	0.024	0.031	0.029	0.036

Different models for marginal and cross-correlation of earthquake IMs have been proposed in previous studies. All of these models are based on the assumption of isotropy. Recent studies show that the assumption of isotropy is not valid in general and consequently appropriate models should be implemented in this regard. On the other hand, spatial correlations of multivariate random fields is a challenging issue in statistics especially when anisotropy is to be involved. The latent dimensions method, which is proposed by [Apanasovich and Genton \(2010\)](#), presents an effective way to overcome this challenge and to construct a valid covariance matrix. This method is implemented in the current study to investigate spatial correlations of multiple earthquake IMs considering anisotropy. Recorded data of three earthquake events in Japan are implemented in the current study and marginal and cross-covariance models are fitted to the considered data. Conducted investigations confirm that the assumption of isotropy is not valid generally. Moreover, anisotropy properties of marginal and cross-covariance models including anisotropy direction and maximum range of correlations have been studied in the current study.

Acknowledgement

The authors acknowledge financial support from the International Institute of Earthquake Engineering and Seismology. Also, the authors would like to acknowledge Dr. Garakaninezhad for his valuable comments.

References

- Apanasovich, T., Genton, M. (2010), Cross-Covariance Functions for Multivariate Random Fields Based on Latent Dimensions, *Biometrika*, **97**, 15-30.
- Boore, D. M., Gibbs, J. F., Joyner, W. B., Tinsley, J. C., Ponti, D. J. (2003), Estimated Ground Motion from the 1994 Northridge, California, Earthquake At the Site of the Interstate 10 and la Cienega Boulevard Bridge Collapse, East Los Angeles, California, *Bulletin of the Seismological Society of America*, **93**, 2737-2751.

Bowman, A. W. and Crujeiras, R. M. (2013), Inference for Variograms, *Computational Statistics and Data Analysis*, **66**, 19-31.

Campbell, K. W., Bozorgnia, Y. (2008), Nga Ground Motion Model for the Geometric Mean Horizontal Component of pga, pgv, pgd and 5% Damped Linear Elastic Response Spectra for Periods Ranging from 0.01 to 10 s, *Earthquake Spectra*, **24**, 139-171.

Campbell, K. W., Bozorgnia, Y. (2014), Nga-west2 Ground Motion Model For the Average Horizontal Components of pga, pgv, and 5% Damped Linear Acceleration Response Spectra, *Earthquake Spectra*, **30**, 1087-1115.

Cressie, N. (1993). *Statistics for Spatial Data*, John Wiley, New York.

Du, W., Wang, G. (2013), Intra-Event Spatial Correlations for Cumulative Absolute Velocity, Arias Intensity, and Spectral Accelerations Based on Regional Site Conditions, *Bulletin of the Seismological Society of America*, **103**, 1117-1129.

Esposito, S., Iervolino, I. (2011), Pga and pgv Spatial Correlation Models Based on European Multievent datasets, *Bulletin of the Seismological Society of America*, **101**, 2532-2541.

Garakaninezhad, A., Bastami, M. (2017), A Novel Spatial Correlation Model Based on Anisotropy of Earthquake Ground-Motion Intensity, *Bulletin of the Seismological Society of America*, **107**, 2809-2820.

Genton, M. G., Kleiber, W. (2015), Cross-Covariance Functions for Multivariate Geostatistics, *Statistical Science*, **30**, 147-163.

Goda, K., Hong, H. P. (2008), Spatial Correlation of Peak Ground Motions and Response Spectra, *Bulletin of the Seismological Society of America*, **98**, 354-365.

Goda, K., Atkinson, G. M. (2009), Probabilistic Characterization of Spatially Correlated Response Spectra for Earthquakes in Japan, *Bulletin of the Seismological Society of America*, **99**, 3003-3020.

Jayaram, N., Baker, J. W. (2008), Statistical Tests of the Joint Distribution of Spectral Acceleration Values, *Bulletin of the Seismological Society of America*, **98**, 2231-2243.

Jayaram, N., Baker, J. W. (2009), Correlation Model For Spatially Distributed Ground-Motion Intensities, *Earthquake Engineering and Structural Dynamics*, **38**, 1687-1708.

Loth, C., Baker, J. W. (2013), A Spatial Cross-Correlation Model of Spectral Accelerations at Multiple Periods, *Earthquake Engineering and Structural Dynamics*, **42**, 397-417.

Markhvida, M., Ceferino, L., Baker, J. W. (2018), Modeling Spatially Correlated Spectral Accelerations at Multiple Periods Using Principal Component Analysis and Geostatistics, *Earthquake Engineering and Structural Dynamics*, **47**, 1107-1123.

Wang, G., Du, W. (2013), Spatial Cross-Correlation Models for Vector Intensity Measures (pga, i a, pgv, and sas) Considering Regional Site Conditions, *Bulletin of the Seismological Society of America*, **103**, 3189-3204.

Wang, M., Takada, T. (2005), Macrospatial Correlation Model of Seismic Ground Motions, *Earthquake spectra*, **21**, 1137-1156.



Spatial Dependence of Extreme Damage Probabilities in the Lifeline Networks Using the Max-stable Process

Shahin Borzoo^{1*}, Morteza Bastami¹, Afshin Fallah²

¹International Institute of Earthquake Engineering and Seismology, Tehran, Iran.

²Department of Statistics, Imam Khomeini International University, Ghazvin, Iran.

Abstract:

In this paper, the max-stable process is used to model the spatial dependence of the probabilities of exceedance for different damage states. Also, the extreme value theory is used to select the extreme scenarios for seismic hazard assessment of the transportation network of Tehran. The probabilities of different damage states are investigated. The extremal coefficient function of Brown-Resnick model is adopted to model the spatial dependence of extreme value of damage probabilities. The results showed that the probabilities of exceedance for different damage states are spatially dependent and it should be considered to assess the performance of lifeline network.

Keywords: Max-stable process, Brown-Resnick model, Spatial dependence, Extreme values, Damage probabilities.

Mathematics Subject Classification (2010): 86A32, 60G70, 62H11.

1 Introduction

In the seismic hazard assessment of lifeline networks, the Monte Carlo method is used to simulate the input seismic scenarios. Since a large part of simulated scenarios is not important due to the small values of their intensities at intended sites the produced scenarios cause high computational costs. Hence the scenarios that cause large intensities are more important than other scenarios. These scenarios

*Speaker: s.borzoo@iiees.ac.ir

are so-called extreme scenarios that have large return period and a small probability of occurrence. Since extreme scenarios are such that their occurrence will have irreparable consequences, analysis of lifeline networks based on these scenarios is recommended. For this reason, the extreme value theory is proposed to select extreme scenarios. In extreme value theory two common extreme distributions are used to model the extreme and rare events; generalized extreme value distribution for block maxima and generalized Pareto distribution for peaks over thresholds. In this study, the seismic analysis of selected lifeline network has been done using the extreme scenarios. The ground motion intensity measures of each scenario have been calculated by a ground motion model. In the next step, the intensity measures have deduced the probabilities of exceedance for different damage states using fragility curves. Due to the wide dimension of lifeline networks, the deterministic and probabilistic seismic hazard assessment methods are not applicable. It is suggested to use the methods that consider the spatial dependence of response values of a network in important nodes of it. Since the damage probabilities are deduced from extreme scenarios, it would be possible to use the dependent models based on the extreme value theory to analyze the seismic performance of lifeline network. In this paper, the max-stable process has been identified and used to model the spatial extreme values. In the second section, the max-stable models have been detailed. In the third section, a part of the transportation network of Tehran has been investigated using the Brown-Resnick model. In the last section, the discussion and results have been presented.

2 Max-stable model

The main purpose of modeling the spatial extreme values is the risk estimation in spatial cases. Max-stable models are used to model multivariate extreme values. For example, using the precipitation values, $y(x)$, in a catchment area, $\chi \subset R^2$ the probabilities of exceedance of precipitation for the selected catchment area is

$$\Pr \left\{ \int_{\chi} Y(x) dx > z_{crit} \right\}, \quad (2.1)$$

where z_{crit} is an extreme level. It is clear that estimating the above probability is more difficult than estimating in a univariate or multivariate model, as it requires the knowledge of examining the distribution of the random variable $y(x)_{x \in \chi}$ which are spatially correlated. If Z_1, Z_2, \dots is an independent sequence of a random process

$\{Z(x) : x \in \chi\}$ and there are $a_n > 0$ and $b_n \in R$ for $n \geq 1$ such that

$$\frac{\max_{i=1, \dots, n} Z_i - b_n}{a_n} \stackrel{d}{=} Z, \quad (2.2)$$

and $\{Z(x) : x \in \chi\}$ is the max-stable process (Ribatet , 2013). Investigating the distribution of multivariate extreme values with max-stable models is one of the issues under consideration in the technical literature that is still under development.

2.1 Max-stable process models

The first model of the max-stable family is the Smith model (Smith , 1990). In this model the maximum value in each $x \in \chi$ is selected as:

$$Z(x) = \max_{i \geq 1} \zeta_i \varphi(x - U_i; 0, \Sigma) \quad x \in \chi, \quad (2.3)$$

where $\{(\zeta_i, U_i) : i \geq 1\}$ are points of Poisson process on $(0, \infty) \times R^d$ with intensity measure $\zeta^{-2} d\zeta du$ and $\varphi(\cdot, 0, \Sigma)$ denotes the multivariate Normal distribution with mean 0 and the covariance matrix Σ . This process is rarely used due to the lack of flexibility in modeling. The second model of the max-stable process was presented by Schlather (2002). This process is sometimes referred to as the maximum Gaussian process

$$Z(x) = \sqrt{2\pi} \max_{i \geq 1} \zeta_i \max\{0, W_i(x)\} \quad x \in \chi, \quad (2.4)$$

where $\{W_i(x) : x \in \chi\}$ are independent replicates of a standard Gaussian process with correlation function ρ . It is notable that the scale factor $\sqrt{2\pi}$ is required to $\sqrt{2\pi} E[\max\{0, W_i(x)\}] = 1, x \in \chi$. The third model of this process is Brown-Resnick. This model is

$$Z(x) = \max_{i \geq 1} \zeta_i \exp\{W_i(x) - \gamma(x)\}, \quad x \in \chi, \quad (2.5)$$

where $\{W_i(x) : x \in \chi\}$ are independent replicates of a Gaussian process having stationary increments and variogram $\gamma(h) = Var\{W(x+h) - W(x)\}/2$ (Brown and Resnick , 1977). The last model of this section is known as the extremal-t2 that generalizes the Schlather model. It is presented firstly in multivariate topics by Nikoloulopoulos et al. , (2009) and statistical modeling of spatial extreme by Davison et al. , (2012) and Ribatet and Sedki , (2013). Finally, Opitz , (2012) presents the spectrum properties of this process as

$$Z(x) = c_v \max_{i \geq 1} \zeta_i \max\{0, W_i(x)\}^v, \quad x \in \chi, \quad (2.6)$$

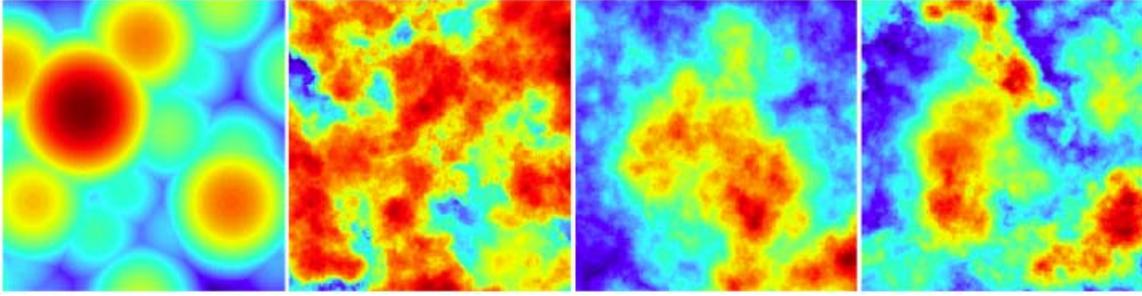


Figure 1: different models of the max-stable process. Left to right: Smith model, Schlather model, Brown-Resnick model, extremal-t2 model (Ribatet , 2013).

where $\nu \geq 1$ and $\{W_i(x) : x \in \chi\}$ are independent replicates of a standard Gaussian process with correlation function ρ and $c_\nu = \sqrt{\pi} 2^{-(\nu-2)/2} \Gamma(\frac{\nu+1}{2})^{-1}$ where Γ is Gamma function. As can be seen in mentioned models, the Schlather model is a special form of extremal-t2 model with $\nu = 1$. In Figure 1 the defined max-stable models are presented. As expected, the Smith process has provided synthetic surfaces that have created its inflexibility as previously mentioned. Other processes produce sample surfaces with wavy contours. As it is shown in figure 1 the Schlather model offers a larger area of high value areas than other models. Although the max-stable theory has been well developed until decades, no specific application of this process has been seen in modeling spatial values. The main reason was the lack of closed-form solution for likelihood functions of this process.

3 Spatial dependence of extreme values using the max-stable process

In geostatistics, the variogram is calculated using

$$\gamma(x_1 - x_2) = \frac{1}{2} E[\{Z(x_1) - Z(x_2)\}^2] \quad x_1, x_2 \in \chi, \quad (3.1)$$

to define the spatial dependence of normal random variables. Since the high-order moments of extreme values (such as variance and even mean) might not even exist, in spatial analysis of extreme values this model of variogram is not applicable. One convenient way to summarize the dependence structure of the max-stable process is through its extremal coefficient function (Schlather and Tawn , 2003).

$$P_r[Z(x_1) \leq z, Z(x_2) \leq z] = P_r[Z(x_1) \leq z]^{\theta(x_1-x_2)}, \quad (3.2)$$

where θ is extremal coefficient function and a measure of spatial dependence as $\theta = 1$ shows complete dependence and $\theta = 2$ shows independence (Ribatet, 2013). The extremal coefficient function of different models of the max-stable process is presented in Table 1.

Table 1: Different models of extremal coefficient function.

$\theta_{(x_1-x_2)} = 2\Phi \left\{ \frac{\sqrt{(x_1-x_2)^T \Sigma^{-1} (x_1-x_2)}}{2} \right\}$	Smith model
$\theta_{(x_1-x_2)} = 1 + \sqrt{\frac{1-\rho_{(x_1-x_2)}}{2}}$	Schalter model
$\theta_{(x_1-x_2)} = 2\Phi \sqrt{\frac{\sigma^2 \{1-\rho_{(x_1-x_2)}\}}{2}}$	extremal-t2 model
$\theta_{(x_1-x_2)} = 2\Phi \left(\sqrt{\frac{\gamma_{(x_1-x_2)}}{2}} \right)$	Brown-Resnick model

4 Spatial dependence in the lifeline network

In this paper, a part of Tehran transportation system which includes 26 bridges is modeled. In order to simulate seismic scenarios, the Mosha, Niavaran, Tehran north, Ray north, Ray south, Kahrizak, Ghamsar and Pishva faults are considered. In the first step, 10000 seismic scenarios are simulated using the Monte Carlo method. Using the extreme value theory and return level of Generalized Pareto distribution, 33 extreme scenarios are adopted. The seismic analysis of network using these 33 seismic scenarios is done. In the next step, the ground motion intensity measures of extreme scenarios are obtained using ground motion prediction equations. exceedance of damage probabilities and the combined discrete damage states probabilities are calculated using the fragility curves. The fragility curve states the relation between seismic intensity and structure functionality. The fragility curve of bridge k indicates the exceedance probability of damage for different damage states. The form of fragility curve is

$$P(DS_i \geq ds | IM) = \phi \left(\frac{\ln(IM) - \ln(IM_i)}{\beta_i} \right), \quad (4.1)$$

where ϕ is the normal cumulative distribution function, IM_i is the median value of ground motion in damage state i and β_i is dispersion factor (lognormal standard deviation). Accordingly, damage probabilities of five damage states including, no damage, slight damage, moderate damage, extensive damage, complete damage, is calculated using ground motion intensities and fragility curves. The Brown-Resnick model is used to investigate the spatial dependence. By calculating the extremal

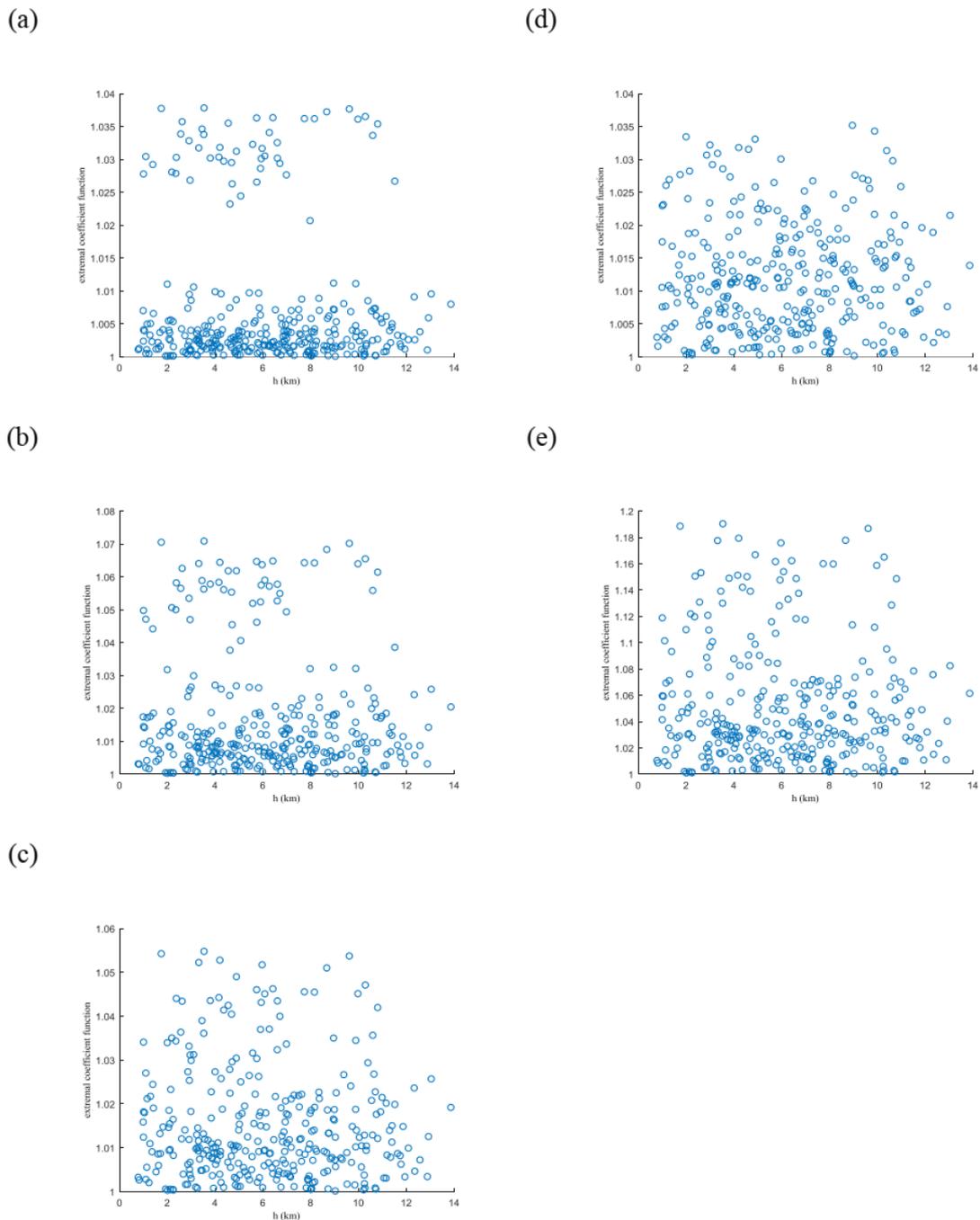


Figure 2: Extremal coefficient function plot for different damage levels for first extreme scenario, (a) no damage, (b) slight damage, (c) moderate damage, (d) extensive damage, (e) complete damage.

coefficient function, the plot of $\theta - h$ is presented for different scenarios and different damage states. As it can be seen in Figure 2, the extremal coefficient function values are near to 1; consequently the probabilities of different damage states are dependent. Also, in lower damage states, damage probabilities are more dependent. The range of the dependence is more than the furthest distance between sites of network.

Consequently, in order to model the dependence of damage probabilities, a spatial extreme model should be used. This model helps to consider the extreme probabilities of the extreme scenarios and it is useful to estimate the risk and functionality of lifeline network under extreme seismic scenarios.

Conclusion

A part of the transportation network of Tehran and their damage probabilities for different damage states is studied. Due to the wide dimension of this network, the necessity of considering the spatial dependence is investigated. In this paper, to model the spatial dependence of extreme values, the max-stable theory is used. The dependency of damage probabilities is investigated using the extremal coefficient function of Brown-Resnick model. The results show that in the selected network, the probabilities of damage for five damage states are dependent. Hence considering the dependency of damage probabilities is necessary for seismic analyzing of lifeline network. Moreover, the results indicate in lower damage states, the dependency is higher than other damage states.

References

- Brown, B. M., Resnick, S. I. (1977), Extreme values of independent stochastic processes, *J. Appl. Probab.*, **14**, 732-739.
- Davison, A. C., Padoan, S. A. and Ribatet, M. (2012), Statistical modeling of spatial extremes, *Stat. Sci.*, **27**, 161-186.
- Nikoloulopoulos, A. K., Joe, H. and Li, H. (2009), Extreme value properties of multivariate t copulas, *Extremes*, **12**, 129-148.
- Opitz, T. (2012), Extremal-t process: Elliptical domain of attraction and a spectral representation, *Submitted, Prepr. <http://arxiv.org/abs/1207.2296>*.
- Ribatet, M. (2013) Spatial extremes: Max-stable processes at work, *J. la Société Française Stat.*, **154**, 156-177.
- Ribatet, M., Sedki, M. (2013), Extreme value copulas and max-stable processes, *J. la Société Française Stat.*, **1**, 138-150.
- Schlather, M. (2002), Models for stationary max-stable random fields, *Extremes*, **5**, 33-44.
- Schlather, M., Tawn, J. A. (2003), A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, **90**, 139-156.
- Smith, R. L. (1990), Max-stable processes and spatial extremes, *Unpubl. Manuscr.*



Sensitivity Analysis In Elliptical Spatial Linear Measurement Error Models

Hadi Emami*

Department of Statistics, Faculty of Science, University of Zanjan, Zanjan, Iran.

Abstract:

In this paper, we consider estimation and inference procedures in spatial linear models when some of the covariates are measured with errors. It is assumed that the additive error distributed according to the law belonging to the class of elliptically contoured distributions. The development of the corrected score function with the family of elliptical distributions is the basis for derivation of the estimators. For sensitivity analysis, the local influence approach is used for assessing influence of small perturbations on the parameter estimates. A simulation study is presented illustrating the good performance of the proposed approach, including the robustness property for the heavier tail models.

Keywords: Geostatistic, Global influence, Spatial variability, Elliptical distribution, Measurement error.

Mathematics Subject Classification (2010): 62H11, 62J20, 62F99.

1 Introduction

Sensitivity analyses and diagnostic techniques for the ordinary linear regressions model have received a great deal of attention in statistical literature since the seminal work of [Cook \(1977\)](#) and others including [Cook and Weisberg \(1982\)](#) and [Poon and Poon \(1999\)](#). In spatial linear regression models, diagnostic results are quite rare; among them [Uribe-Opazo et al \(2012\)](#) applied diagnostic techniques to assess the sensitivity of the maximum likelihood estimators, covariance functions and linear

*Speaker: h.emami@znu.ac.ir

predictor to small perturbations in the Gaussian spatial linear model assumptions. [De Bastiani et al \(2015\)](#) use the local influence methodology to assess the sensitivity of the maximum likelihood estimators to small perturbations in the elliptical spatial linear model. [Lachos et al \(2017\)](#) developed local and global influence measures on the basis of the conditional expectation of the complete-data loglikelihood function in spatial linear regression models when censoring is present. Recently [De Bastiani et al \(2018\)](#) presented global diagnostics techniques to assess the influence of observations on spatial linear mixed models.

However these researchers in their study assumed that all the observation are correctly observed. But this assumption can be violated, and the measurement errors may crept into and the usual statistical tools tend to loose their validity, see [Fuller \(1987\)](#). Fields, such as agriculture, medicine, engineering, psychology, education, and finance are some disciplines presenting situations where covariates are contaminated by measurement errors. For this reason there has been extensive research in the measurement error problem. So, when the problem of measurement error occur in the model data, then the main subject is how to obtain the consistent estimates of regression parameters. In spatial linear measurement error models solving such problems are quite rare and no results are so far available in the literature. So in this paper we fill this gap with a more general assumption. In general, most of the researchers assumed that the covariate matrix of the measurement error models is normally distributed. However, such assumptions may not be valid in many practical situations. It happens particularly if the covariate distribution has heavier tails. Since the elliptical distribution contains a lot of distributions, defining biased estimators for the elliptical error distribution in such models would be a valuable asset for the researchers of this topic. By employing the corrected score function of [Nakamura \(1990\)](#) we concentrate on spatial linear regression models and derive the corrected maximum likelihood estimator of the parameters by relaxing the assumption of normality of errors of covariates.

The paper unfolds as follows: In section 2 we explain the spatial linear measurement error model. In section 3 the maximum corrected likelihood estimation of parameters is described. Section 4 reviews concepts of the local influence approach. Section 5 contains a simulation study, to illustrate the methodology developed in this paper.

2 The model description

Consider a Gaussian stochastic process $\{Y(s), s \in \mathcal{D}\}$, where \mathcal{D} is a subset of \mathcal{R}^d , the d -dimensional Euclidean space. It supposes that data $Y(s_1), \dots, Y(s_n)$ of this process are collected at known sites (locations) s_i , for $i = 1, \dots, n$ where s_i is a d -dimensional vector of spatial site coordinates, and generated from the model,

$$Y(s_i) = \mu(s_i) + \epsilon(s_i)$$

with the deterministic and stochastic terms, $\epsilon(s_i)$ and $\mu(s_i)$ respectively, may depend on the spatial location at where $Y(s_i)$ is collected. The Gaussian stochastic errors have zero mean, $E(\epsilon(s_i)) = 0$ and the variation between spatial points is determined by a covariance function $C(s_i, s_j) = cov(\epsilon(s_i), \epsilon(s_j))$. It assumed each family of covariance functions $C(s_i, s_j)$, is fully specified by a q -dimensional parameter vector $\phi = (\phi_1, \dots, \phi_q)^\top$. For example, with $q = 3$, the Matern is a covariance function particularly attractive given by

$$C(d_{ij}) = \begin{cases} \frac{\phi_2}{2^{k-1}\Gamma(k)} (d_{ij}/\phi_3)^k K_k(d/\phi_3) & d_{ij} > 0 \\ \phi_1 + \phi_2 & d_{ij} = 0 \end{cases}$$

where $d_{ij} = \|s_i - s_j\|$ being the Euclidean distance between the points s_i and s_j , $\phi_j > 0$, $j = 1, 2, 3$ and $K_k(u) = \int_0^\infty x^{k-1} e^{-u(x+x^{-1})} dx$ is the modified Bessel function of the third kind of order k , with $k > 0$ fixed.

In linear spatial models for some known functions of s_i , $x_1(s_i), \dots, x_p(s_i)$ the mean of stochastic process is

$$\mu(s_i) = \sum_{j=1}^p X_j(s_i) \beta_j = X_i^\top \beta$$

where $X_i = (x_1(s_i), \dots, x_p(s_i))$ and $\beta_1 \dots \beta_p$ are unknown parameters to be estimated. In this paper we consider the case where the covariate X_i is measured with additive errors, i.e., we cannot observe X_i but we can observe $W_i = (w_1(s_i), \dots, w_p(s_i))$ with

$$W_i = X_i + \delta_i$$

where the spatial measurement errors $\delta_i = (\delta_1(s_i), \dots, \delta_p(s_i))$ are i.i.d from $El_p(0, \Lambda)$ with Λ a $p \times p$ known matrix. δ_i 's are independent of X_i and $\epsilon = (\epsilon(s_1), \dots, \epsilon(s_n))$. When W_i is identically equal to a null vector, implying that $\Lambda = 0$ and consequently that X_i is fixed and is measured without any measurement error, we get the classical

spatial linear regression model. Equivalently a matrix spatial linear measurement error model can be written as

$$\begin{aligned} Y &= \mathbf{X}\beta + \epsilon \\ \mathbf{W} &= \mathbf{X} + \delta \quad \delta = (\delta_1, \dots, \delta_n)^\top \end{aligned} \quad (2.1)$$

where matrices \mathbf{X} , \mathbf{W} and δ are the $n \times p$ full rank matrices with i th row X_i^\top , W_i^\top and δ_i^\top respectively. In this notation, $E(\epsilon) = 0$ and the covariance matrix of ϵ is $\Sigma_\phi = [(\sigma_{ij})]$, where σ_{ij} is proportional to $C(s_i, s_j)$. A particular parametric form for the non-singular covariance matrix Σ_ϕ is

$$\Sigma_\phi = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R} \quad (2.2)$$

where ϕ_1 is a measurement error variance or a nugget effect, ϕ_2 is defined as sill, \mathbf{I}_n is the $n \times n$ identity matrix and $\mathbf{R} = \mathbf{R}(\phi_3) = [(r_{iu})]$, is an $n \times n$ symmetric matrix with diagonal elements $r_{ii} = 1$, for $i = 1, \dots, n$ and ϕ_3 is a function of the range of the model. This parametric form occurs for several isotropic processes, where $C(s_i, s_j)$ is defined via the function $C(d_{ij}) = \phi_2 r_{ij}$. In the covariance functions $C(d_{ij})$, the variance of the stochastic process ϵ is $C(0) = \phi_1 + \phi_2$ and the variogram can be defined as $\gamma(d) = C(0) - C(d)$.

3 Maximum Corrected Likelihood Estimation (CLE)

For our model (2.1) it is easily seen that the loglikelihood is

$$\mathcal{L}(\theta; \mathbf{X}, Y) = \frac{-1}{2} \log |\Sigma_\phi| - \frac{1}{2} (Y - \mathbf{X}\beta)^\top \Sigma_\phi^{-1} (Y - \mathbf{X}\beta) \quad (3.1)$$

with $\theta = (\beta^\top, \phi^\top)^\top$, $\phi = (\phi_1, \phi_2, \phi_3)$ and $\Sigma_\phi = \phi_1 \mathbf{I}_n + \phi_2 \mathbf{R}$ as in (2.2). In cases that measurement error is negligible, we replace \mathbf{X} by \mathbf{W} in (3.1) so that

$$U(\theta; \mathbf{W}, Y) = \frac{\partial}{\partial \theta} \mathcal{L}(\theta; \mathbf{W}, Y)$$

which is typically called naive score function. Hence, the expectations of $U(\theta; \mathbf{W}, Y)$ with respect to Y evaluated at the true parameter value θ_0 , typically, are not equal to zero. Furthermore, in general, estimators obtained using naive score functions are not consistent (Nakamura , 1990).

The approach developed to obtain the corrected score functions corresponds to find a score function whose expectation with respect to the distribution of \mathbf{W} coincides

with the unobserved score (a function of \mathbf{X}). Denote by E^* the conditional expectation (with respect to the distribution of \mathbf{W}) given \mathbf{X} , and Y . The corrected log-likelihood function for our model, denoted by $\mathcal{L}^*(\theta; \mathbf{W}, Y)$, should satisfy the condition in

$$E^*(U^*(\theta; \mathbf{W}, Y)) = U(\theta; \mathbf{X}, Y) \quad (3.2)$$

where $U^*(\theta; \mathbf{W}, Y) = \frac{\partial}{\partial \theta} \mathcal{L}^*(\theta; \mathbf{W}, Y)$. Hence, under the conditions for the model defined by (2.1) and using property 3.1 in Riquelmea et al (2015), we have that

$$E^*(\mathbf{W}^\top \Sigma_\phi^{-1} \mathbf{W}) = \mathbf{X}^\top \Sigma_\phi^{-1} \mathbf{X} - 2\varphi(0)' \text{tr}(\Sigma_\phi^{-1}) \Lambda \quad (3.3)$$

Given the matrix Λ , it follows from (3.2) and (3.3), that the unique corrected log-likelihood function for the model defined in (9) and (10), is given by

$$\mathcal{L}^*(\theta; \mathbf{W}, Y) = \frac{-1}{2} \log |\Sigma_\phi| - \frac{1}{2} \left((y - \mathbf{W}\beta)^\top \Sigma_\phi^{-1} (y - \mathbf{W}\beta) + 2\varphi(0)' \text{tr}(\Sigma_\phi^{-1}) \beta^\top \Lambda \beta \right) \quad (3.4)$$

Making use of the corrected log-likelihood functions given in (3.4) it is possible to obtain the corresponding corrected scores as:

$$U^*(\beta; \mathbf{W}, Y) = \frac{\partial}{\partial \beta} \mathcal{L}^*(\theta; \mathbf{W}, Y) = \mathbf{W}^\top \Sigma_\phi^{-1} (Y - \mathbf{W}\beta) - 2\varphi(0)' \text{tr}(\Sigma_\phi^{-1}) \Lambda \beta \quad (3.5)$$

and

$$U^*(\phi; \mathbf{W}, Y) = \frac{\partial}{\partial \phi} \mathcal{L}^*(\phi; \mathbf{W}, Y) \quad \text{with} \quad (3.6)$$

$$\begin{aligned} \frac{\partial}{\partial \phi_j} \mathcal{L}^*(\phi; \mathbf{W}, Y) = & -\frac{1}{2} \text{tr}(\Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j}) + \frac{1}{2} (Y - \mathbf{W}\beta)^\top \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} \Sigma_\phi^{-1} (Y - \mathbf{W}\beta) \\ & + \varphi(0)' \text{tr}(\Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} \Sigma_\phi^{-1}) \beta^\top \Lambda \beta \end{aligned} \quad (3.7)$$

where $\dot{\Sigma}_{\phi_j} = \frac{\partial \Sigma}{\partial \phi_j}$. From (3.5)-(3.7), we can verify that the corrected score functions satisfy the conditions given in (3.2). Let θ_0 be the true parameter value and E^+ denote the expectation respect to random vector Y and $E = E^+ E^*$ denote global expectation then

$$E(U^*(\beta_0; \mathbf{W}, Y)) = E^+ E^*(U^*(\beta_0; \mathbf{W}, Y)) = E^+ \left(\mathbf{X}^\top \Sigma_\phi^{-1} (Y - \mathbf{X}\beta) \right) = 0$$

and

$$E(U^*(\phi_{0j}; \mathbf{W}, Y)) = E^+ E^*(U^*(\phi_{0j}; \mathbf{W}, Y)) = -\frac{1}{2} \text{tr}(\Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j}) + \frac{1}{2} E^+ \left(\epsilon^\top \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} \Sigma_\phi^{-1} \epsilon \right) = 0$$

This indicates that the corrected score functions are unbiased. The corrected observed information matrix is

$$\mathbf{I}^*(\theta; \mathbf{W}, Y) = -\frac{\partial^2}{\partial\theta\partial\theta^\top}\mathcal{L}(\theta; \mathbf{W}, Y) = \begin{pmatrix} \ddot{\mathcal{L}}_{\beta\beta} & \ddot{\mathcal{L}}_{\beta\phi} \\ \ddot{\mathcal{L}}_{\beta\phi}^\top & \ddot{\mathcal{L}}_{\phi\phi} \end{pmatrix}$$

where

$$\begin{aligned} \ddot{\mathcal{L}}_{\beta\beta} &= \mathbf{W}^\top \Sigma_\phi^{-1} \mathbf{W} + 2\varphi(0)'tr(\Sigma_\phi^{-1})\Lambda \\ \ddot{\mathcal{L}}_{\beta\phi_j} &= \mathbf{W}^\top \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} \Sigma_\phi^{-1} (Y - \mathbf{W}\beta) - 2\varphi(0)'tr(\Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} \Sigma_\phi^{-1})\Lambda\beta \\ \ddot{\mathcal{L}}_{\phi_i\phi_j} &= -\frac{1}{2}tr(\Sigma_\phi^{-1} \dot{\Sigma}_{\phi_i} \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} - \Sigma_\phi^{-1} \ddot{\Sigma}_{\phi_{ij}}) + \frac{1}{2}\alpha^\top \Sigma_\phi^{-1} \mathbf{M} \Sigma_\phi^{-1} \alpha \\ &\quad + \varphi(0)'tr(\Sigma_\phi^{-1} \mathbf{M} \Sigma_\phi^{-1})\beta^\top \Lambda \beta \end{aligned}$$

with $\alpha = Y - \mathbf{W}\beta$ and $\mathbf{M} = \dot{\Sigma}_{\phi_i} \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} - \ddot{\Sigma}_{\phi_{ij}} + \dot{\Sigma}_{\phi_j} \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_i}$. The corrected fisher information matrix is

$$\mathbf{I}(\theta; \mathbf{W}, Y) = E^+ E^* (\mathbf{I}^*(\theta; \mathbf{W}, Y)) = \begin{pmatrix} \mathbf{X}^\top \Sigma_\phi^{-1} \mathbf{X} & 0 \\ 0^\top & -\frac{1}{2}tr(\Sigma_\phi^{-1} \dot{\Sigma}_{\phi_i} \Sigma_\phi^{-1} \dot{\Sigma}_{\phi_j} - \Sigma_\phi^{-1} \ddot{\Sigma}_{\phi_{ij}}) \end{pmatrix}$$

3.1 Corrected score estimators

Since, the equation $U^*(\phi; \mathbf{W}, Y) = 0$ does not lead to an explicit solution for ϕ a common practice is to maximize the concentrated log-likelihood obtained as follows. From (3.4), given Σ_ϕ the corrected estimator for β is

$$\hat{\beta} = (\mathbf{W}^\top \Sigma_\phi^{-1} \mathbf{W} + 2\varphi(0)'tr(\Sigma_\phi)\Lambda)^{-1} \mathbf{W}^\top \Sigma_\phi^{-1} Y \quad (3.8)$$

By substituting (3.8) into the corrected log-likelihood function, we obtain a concentrated corrected log-likelihood

$$\mathcal{L}_c(\theta, \mathbf{W}, Y) = \frac{-1}{2} \log|\Sigma_\phi| - \frac{1}{2} \left((y - \mathbf{W}\hat{\beta})^\top \Sigma_\phi^{-1} (y - \mathbf{W}\hat{\beta}) + 2\varphi(0)'tr(\Sigma_\phi^{-1})\hat{\beta}^\top \Lambda \hat{\beta} \right)$$

which must be maximized numerically with respect to ϕ_j 's. Given the CLE of ϕ , $\hat{\phi}$ say, the CLE of β is

$$\hat{\beta} = (\mathbf{W}^\top \hat{\Sigma}_\phi^{-1} \mathbf{W} + 2\varphi(0)'tr(\hat{\Sigma}_\phi)\Lambda)^{-1} \mathbf{W}^\top \hat{\Sigma}_\phi^{-1} Y$$

where $\hat{\Sigma}_\phi = \Sigma_\phi$. Let ξ_k be a p -vector with 1 at the k th position and zero elsewhere, then $\hat{\beta}_k = \xi_k \hat{\beta}$ has the standard error s_k and t -value $t_k = \hat{\beta}_k / s_k$, where $s_k^2 =$

$$\xi_k^\top (\mathbf{W}^\top \hat{\Sigma}_\phi^{-1} \mathbf{W} + 2\varphi(0)' \text{tr}(\hat{\Sigma}_\phi) \Lambda)^{-1} \xi_k.$$

4 Local influence diagnostics

Local influence is a method of sensitivity analysis for assessing the influence of small perturbations in a general statistical model. The general idea is to give every individual its own weight in the calculation of the parameter estimates and to investigate how these estimates depend on the weights, locally around the equal-weight case, which is the ordinary maximum likelihood case. Let $\mathcal{L}(\theta|\omega)$ denote any perturbed version of $\mathcal{L}(\theta)$, depending on an r dimensional vector ω of weights, which is assumed to belong to an open subset Ω of R^r , and such that the original log-likelihood $\mathcal{L}(\theta)$ is obtained for $\omega = \omega_0$. Let $\hat{\theta}_\omega$ and $\hat{\theta}$ denote the maximum likelihood estimates under the model defined by $\mathcal{L}(\theta)$ and $\mathcal{L}(\theta|\omega)$, respectively, and assume that there is an $\omega_0 \in \Omega$ representing no perturbation, such that $\mathcal{L}(\theta) = \mathcal{L}(\theta|\omega)$ for all θ . Cook (1977) proposed to measure the distance between $\hat{\theta}_\omega$ and $\hat{\theta}$ by the so-called likelihood displacement, defined by

$$LD(\omega) = 2[\mathcal{L}(\hat{\theta}) - \mathcal{L}(\hat{\theta}|\omega)]$$

From this perspective, a graph of $LD(\omega)$ versus ω contains essential information on the influence of case weight perturbations. Because evaluation of $LD(\omega)$ for all ω is practically unfeasible, Cook (1977) proposed to study the local behavior of $LD(\omega)$ around ω_0 , which can be performed by evaluating the normal curvature C_i of $LD(\omega)$ at ω_0 in the direction of some unit vector h . The normal curvature in the direction h takes the form,

$$C_h = 2|h^\top \ddot{\mathbf{F}} h| \quad (4.1)$$

where $\ddot{\mathbf{F}} = \Delta^\top \mathbf{I}^{*-1} \Delta$ and $\Delta = \frac{\partial \mathcal{L}^*(\theta)}{\partial \theta \partial \omega^\top}$ are both evaluated at $\theta = \hat{\theta}$ and $\omega = \omega_0$. There are several ways in which (4.1) can be used to study the influence graph, each corresponding to a specific choice of the unit vector h . One evident choice corresponds to the perturbation of the i th weight only (case-weight perturbation). This is obtained by taking h equal to the vector h_i which contains zeros everywhere except on the i th position, where there is a one. Another important direction is determined by h_{max} , which corresponds to the maximal normal curvature C_{max} . The most influential elements of the data may be identified by looking the components of the vector h_{max} , which is relatively large. Furthermore, h_{max} is just the eigenvector corresponding to the largest eigenvalue, (C_{max}) , of $\ddot{\mathbf{F}}$. Although Cook's method works well for a lot of statistical models, several shortcomings remain to be resolved. Therefore,

Poon and Poon (1999) proposed the invariant conformal normal curvature. This curvature provides a measure of local influence ranging from 0 to 1, with objective bench-marks to judge largeness. According to the arguments of Poon and Poon (1999), the corresponding conformal normal curvatures of C_h for our considered model may be expressed as:

$$B_h = \frac{h^\top \ddot{\mathbf{F}} h}{\sqrt{\text{tr}(\ddot{\mathbf{F}})^2}} = \frac{h^\top \Delta^\top \mathbf{I}^{*-1} \Delta h}{\sqrt{\text{tr}(\Delta^\top \mathbf{I}^{*-1} \Delta)^2}}$$

4.1 Case-weight perturbation

For case weight perturbation we assign an arbitrary weighting vector $\omega = (\omega_1, \dots, \omega_n)^\top$ with $0 \leq \omega_i \leq 1$ to the individual corrected likelihood function, that is

$$\begin{aligned} \mathcal{L}^*(\theta|\omega) &= \sum_{i=1}^n \omega_i \mathcal{L}_i^*(\theta) \\ &= \frac{-1}{2} \log |\Sigma_\phi| - \frac{1}{2} \left((y - \mathbf{W}\beta)^\top \mathbf{P}_\omega \Sigma_\phi^{-1} (y - \mathbf{W}\beta) + 2\varphi(0)' \text{tr}(\mathbf{P}_\omega \Sigma_\phi^{-1}) \beta^\top \Lambda \beta \right) \end{aligned} \quad (4.2)$$

where $\mathbf{P}_\omega = \text{diag}(\omega_1, \dots, \omega_n)$. By taking $\omega_i = 0$ and $\omega_j = 1$ for $j \neq i$ that the information for the i th individual is excluded from the corrected log-likelihood function so that the non-perturbation vector will be $\omega_0 = \mathbf{1}_n = (1, \dots, 1)^\top \in R^n$. Differentiating $\mathcal{L}^*(\theta|\omega)$ given in (4.2), we obtain

$$\Delta_\beta = \frac{\mathcal{L}^*(\theta|\omega)}{\partial \beta \partial \omega^\top} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} = \mathbf{W}^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \hat{e}) - 2\varphi(0)' \Lambda \hat{\beta} \mathbf{1}_n^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1})$$

and

$$\Delta_{\phi_j} = \frac{\mathcal{L}^*(\theta|\omega)}{\partial \phi_j \partial \omega^\top} \Big|_{\theta=\hat{\theta}, \omega=\omega_0} = \frac{1}{2} \hat{e}^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \dot{\Sigma}_{\hat{\phi}_j} \Sigma_{\hat{\phi}}^{-1} \hat{e}) + \varphi(0)' \hat{\beta}^\top \Lambda \hat{\beta} \mathbf{1}_n^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \dot{\Sigma}_{\hat{\phi}_j} \Sigma_{\hat{\phi}}^{-1})$$

Thus we have

$$\Delta = \begin{pmatrix} \mathbf{W}^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \hat{e}) - 2\varphi(0)' \Lambda \hat{\beta} \mathbf{1}_n^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1}) \\ \frac{1}{2} \hat{e}^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \dot{\Sigma}_{\hat{\phi}_1} \Sigma_{\hat{\phi}}^{-1} \hat{e}) + \varphi(0)' \hat{\beta}^\top \Lambda \hat{\beta} \mathbf{1}_n^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \dot{\Sigma}_{\hat{\phi}_1} \Sigma_{\hat{\phi}}^{-1}) \\ \vdots \\ \frac{1}{2} \hat{e}^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \dot{\Sigma}_{\hat{\phi}_q} \Sigma_{\hat{\phi}}^{-1} \hat{e}) + \varphi(0)' \hat{\beta}^\top \Lambda \hat{\beta} \mathbf{1}_n^\top \text{diag}(\Sigma_{\hat{\phi}}^{-1} \dot{\Sigma}_{\hat{\phi}_q} \Sigma_{\hat{\phi}}^{-1}) \end{pmatrix}$$

Table 1: The sample performance of the corrected estimators for two cases with different distributions of measurement error. The results of naive estimators are in the parentheses.

<i>Distribution</i>	$t_5(0, \Lambda)$		$t_{25}(0, \Lambda)$		$N(0, \Lambda)$	
	<i>I</i>	<i>II</i>	<i>I</i>	<i>II</i>	<i>I</i>	<i>II</i>
<i>mean</i>						
$\hat{\beta}_1$	1.969(1.873)	1.901(1.429)	2.002(1.916)	2.007(1.546)	1.988(1.943)	2.089(1.585)
$\hat{\beta}_2$	2.986(2.805)	2.723(2.246)	3.001(2.894)	3.045(2.281)	2.999(2.897)	3.014(2.359)
$\hat{\phi}_1$	4.900(4.799)	4.902(4.541)	4.899(4.812)	4.901(4.791)	4.990(4.900)	4.920(4.814)
$\hat{\phi}_2$	9.799(9.736)	9.799(9.493)	9.830(9.790)	9.801(9.501)	9.899(9.800)	9.861(9.672)
$\hat{\phi}_3$	14.899(14.802)	14.900(14.685)	14.990(14.870)	14.901(14.650)	14.996(14.842)	14.957(14.799)
<i>mse</i>						
$\hat{\beta}_1$	0.071(0.081)	0.140(0.383)	0.065(0.080)	0.122(0.271)	0.063(0.082)	0.118(0.237)
$\hat{\beta}_2$	0.073(0.098)	0.198(0.629)	0.064(0.075)	0.124(0.578)	0.065(0.073)	0.111(0.472)
$\hat{\phi}_1$	0.010(0.092)	0.010(0.104)	0.010(0.022)	0.015(0.0310)	0.010(0.021)	0.011(0.091)
$\hat{\phi}_2$	0.040(0.096)	0.049(0.116)	0.040(0.064)	0.055(0.092)	0.040(0.058)	0.051(0.088)
$\hat{\phi}_3$	0.010(0.087)	0.019(0.091)	0.0104(0.031)	0.013(0.066)	0.010(0.019)	0.011(0.057)

5 Simulation Study

Here to demonstrate the performance of estimators we carry out an simulation study. The data are generated from model (2.1) with $s_i \sim U(0, 100)$, $n = 100$, $\beta = (2, 3)$, the i th column of \mathbf{X} are from $\sim N(0, 1)$, $i = 1, 2$, $\epsilon \sim N(0, \Sigma_\phi)$ where Σ_ϕ is assumed to be Gaussian spatial covariance matrix with $\phi = (5, 10, 15)$ which is given by

$$C(d_{ij}) = \begin{cases} \phi_2 \exp(-(d_{ij}/\phi_3)^2) & d_{ij} > 0 \\ \phi_1 + \phi_2 & d_{ij} = 0 \end{cases}$$

The rows of measurement error matrix δ are generated from $t_2(0, \Lambda)$. For spatial covariance matrix Λ two case are considered. Case (I): $\Lambda = \text{diag}(0.2^2, 0.2^2)$ and Case (II): $\Lambda = \text{diag}(0.5^2, 0.5^2)$. We investigate the corrected score estimator vector of $\theta = (\beta, \phi)$ and it's usual naive estimators ignoring measurement error in \mathbf{W} . The simulation study was conducted using the R software with 10^5 repetition. Table 1 represent the corrected score and naive estimators of θ for different diagonal values of matrix Λ . It can be seen that with increasing measurement errors variances (diagonal elements of Λ), the naive estimators become more biased and the corrected score estimators have better performance than the naive estimators. Another conclusion is that heavier tail model have significant more impact on the bias of the naive estimators of θ than lighter tail models. To evaluate the effectiveness of the local influence measures derived in this paper, we consider three models of the above simulated data set correspond to 3 distribution of the measurement error with two

outliers and a leverage point forced in each models. Before we proceeded, we changed the responses at spatial response points 2 and 62 by adding $+2sd(Y(s_i))$ for these points. To have high leverage for spatial response 2 we also changed the 2th row of \mathbf{W} to vector (3, 5). Figure 1 gives the case weight perturbation for three models. Cases 2 and 60 are found to be influential on estimate θ and case 2 is more influential in each models. In all three situation the outlier screening in Fig.1 is successful in identifying both outliers.

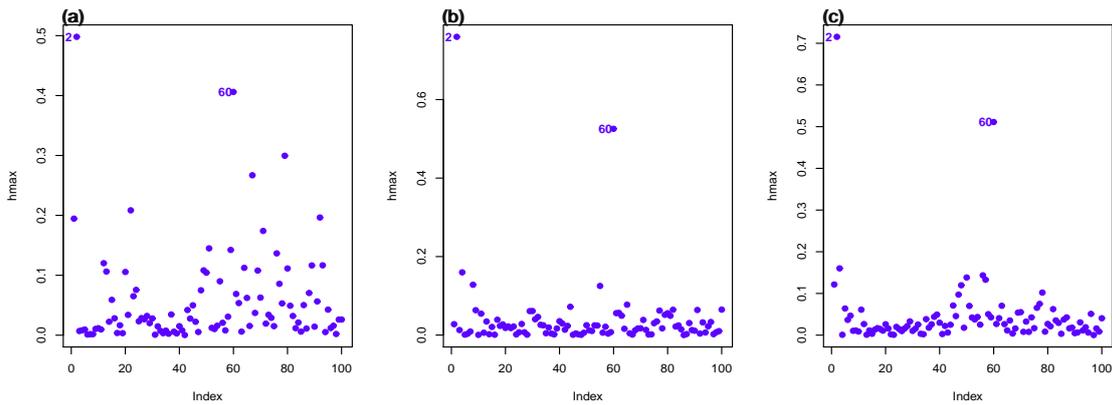


Figure 1: Plot of h_{max} under case- weight perturbation assuming (a) t_5 , (b) t_{25} and (c) *normal* distribution for measurement error

Conclusion

It is not unusual to have measurement error in spatial linear models. We have use the idea of the corrected likelihood method of Nakamura (1990) with to obtain a better estimators for the model parameters in spatial linear measurement error model. We assumed a general assumption for the covariates of the model, which is distributed according to the law belonging to the class of elliptically contoured distributions. The simulation study showed that the proposed estimators are also performing better than the naive estimators (ignoring measurement error). Here we assumed the Λ is known and if it is unknown extending our diagnostics is an area of future research.

References

- Cook, RD. (1977), Detection of influential observations in linear regression, *Technometrics*, **19**, 15-18.

Cook RD., Weisberg S. (1982), *Residuals and influence in regression*, Chapman & Hall, New York.

De Bastiani, F., Uribe-Opazo, M., Cysneiros, A., Galea, M. (2015), Influence diagnostics in elliptical spatial linear models, *Test*, **24**, 322-340.

De Bastiani, F., Uribe-Opazo, M., Cysneiros, A., Galea, M. (2018), Case-deletion diagnostics for spatial linear mixed models, *Spatial Statistics*, **28**, 284-303.

Fuller, W. A. (1987). *Measurement error models*, New York, Wiley.

Lachos, VH., Matos, LA., Barbosa TS., Garay, AM., Dey. DK. (2017), Influence diagnostics in spatial models with censored response, *Environmetric*, **28**, 441-456.

Nakamura, T. (1990), Corrected score functions for error in variables models: Methodology and application to generalized linear models, *Biometrika*, **77**, 127-37.

Poon W, Poon YS, (1999), Conformal normal curvature and assessment of local influence, *Stat Soc Ser B*, **61**, 51-61.

Riquelmea M., H. Bolfarine, and M. Galea. (2015), Robust linear functional mixed models, *Multivariate Analysis*, **134**, 82-98.

Uribe-Opazo MA, Borssoi JA, Galea M. (2012), Influence diagnostics in gaussian spatial linear models, *Appl Stat*, **39**, 615-630.



A Spatial Heterogeneity Mixed Model with Skew-Elliptical Distributions

Mohadeseh Alsadat Farzammehr*

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

Abstract:

In the analysis of most spatial heterogeneity processes in econometrics studies, observations present skewed distributions. Usually, a single transformation of the data is used to approximate normality and to model the transformed data with normal assumption. However, such an assumption may not be appropriate as panel data do often have non-normality feature in these studies. This work relaxes the normality assumption of a spatial mixed model which allows for spatial heterogeneity. The inference procedure is performed under a Bayesian mixed modeling approach based mixed-effects model with a multivariate skew-elliptical distribution, which includes the skew-t, skew-normal, t-student, and normal distributions as special cases. We fit our models to yearly non-life insurance consumption observed between 1998 and 2002 located in a spatial panel of 103 Italian provinces in order to assess its determinants, in the light of the empirical literature. Different model comparison criteria and analysis of the posterior distribution of some parameters suggest that the proposed model outperforms standard ones used in the literature.

Keywords: Panel data, Linear mixed model, Spatial heterogeneity, Multivariate skew-elliptical distributions, Bayesian hierarchical approach, MCMC.

Mathematics Subject Classification (2010): 62H11, 91B72.

1 Introduction

Panel data are observations aggregated on a cross section over several time periods. In the case of cross sections of spatial units (such as regions, provinces or countries),

*Speaker: ma.farzammehr@gmail.com

panel data are referred to as spatial panels. The spatial panel data analysis has received a great deal of attention mostly in spatial econometrics studies.

Analysis of economic data that explicitly incorporate spatial information can be approached from spatial heterogeneity structure. Spatial heterogeneity might arise due to differences among different spaces or among different time points. Fundamentally, spatial heterogeneity is introduced by allowing the model to have heterogeneous effects in space and time. Chesson (1985) discussed the distinctions between three types of variations in these models. They are between pure spatial variation, which is constant from one time period to another, pure temporal variation, which is constant across space, and pure spatio-temporal variation, for which spatial and temporal variation can occur together. Note that the first two types consider spatial and temporal perspectives separately and hence ignore the possibility of spatio-temporal variation in spatial heterogeneity literature.

This paper studies the impacts of spatial heterogeneity that involves spatial-temporal variations. This is achieved by incorporating spatial sampling units (space-specific effects) into the structure of models. In such a setting, four types of space-specific effects as presented by Zheng *et al.* (2008) are considered. The models underlying the different specifications of the space-specific effects are referred to as (i) Homogeneous model (*homogeneous*), (ii) Heterogeneous model (*heterogeneous*), (iii) Fixed-effects model (*fixed*), and (iv) Random-effects model (*random*). The explanation of these models is given in Section 2.1.

Spatial panel data usually show features like skewness in practice. The most routinely adopted strategy in empirical econometric models is that the response and the explanatory variables are transformed so that classical models that are based on the normality assumption can be applied. Although using a transformation to handle the departure from normality may lead to reasonable empirical results, this work is often inappropriate and restrictive. Frequently, a suitable alternative theoretical model (that can directly handle skewness) can perform better than data transformation. Moreover, there are several limitations of using data transformation, including reduced information, no guarantee of joint normality, difficulty in interpreting the transformed variables, and no general transformation (i.e. transforms used for one particular data may not be suitable for a different data). In view of this, we propose new spatial heterogeneity panel models that provide greater flexibility in the distributional assumption for error terms to reduce the impact of the unrealistic normality assumption.

The multivariate skew-elliptical distributions are gaining popularity as useful tools for analyzing a variety of datasets that exhibit non-normal features. These

distributions provide flexible alternatives to traditional normal and t-student models with additional features such as asymmetry and heavy tails. The connections between the various multivariate skew-t distributions are analogues to those between the skew-normal distributions discussed in many studies (see, for example, the papers by Azzalini, 2005, Arellano-Valle and Genton, 2005, Arellano-Valle and Azzalini, 2006 and the recent monograph by Lee and McLachlan, 2013 and Azzalini, 2014).

In the statistical literature, several extensions of standard linear mixed models have been proposed based on replacing the normality assumption for the errors. Further, mixed models provide a convenient framework to model the spatial heterogeneity using space-specific effects. To the best of our knowledge, under the framework of using various mixed-effects models, there has been little work done on discussing the simultaneous impacts induced by spatial heterogeneity and non-normality, which are inherent features of spatial panel data.

Dealing with panel analyses on sets of spaces, the spatial heterogeneity perspective should always be considered for, as further methodological issues arise with respect to panels of spaces, where the relevant markets in spaces are well delimited by heterogeneity structure affecting panels studies. In this way, insurance market is widely believed to be important for the sound economic development of a country. Here, we analyze the consumption of non-life insurance, measured as premiums per capita, across Italys 103 provinces in the period 1998-2002. In order that, we approach the empirical investigation of insurance consumption from a new perspective, an intermediate one between existing panel studies and dataset analyses on household income, wealth and consumption surveys. Note that, data availability for Italian insurance premiums limits the analysis to the provincial aggregation level.

Overall, the objective of this paper is to estimate symmetric and asymmetric models parameters. In order that, we follow the Bayesian hierarchical approach described in Jara *et al.* (2008), in which all of the inferences were carried out through Markov chain Monte Carlo (MCMC) algorithm for drawing inferences in spatial heterogeneity panel model with multivariate skew-elliptical distributions for error terms. An important advantage of this modeling alternative is that the models can be easily fitted in freely available software R under OpenBUGS (Lunn *et al.*, 2009) and that the computational effort is equivalent to the one necessary to fit the normal version of the model. The rest of the paper is organized as follow. Section 2 presents the spatial heterogeneity panel model and Section 3 explains multivariate skew-elliptical distributions. In Section 4, we discuss about model selection criteria that may be used to evaluate the performance of our models. Bayesian analysis of

non-life insurance data set is presented in Section 5. The main goal of the paper is to investigate the determinants of the development of the non-life insurance market for Italy. Finally, Section 6 includes concluding remarks.

2 The Model

We begin by defining a regression model for the panel data that can be expressed as,

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + \lambda_{it} + \epsilon_{it}, \quad (2.1)$$

for $i = 1, \dots, N$ space units, and $t = 1, \dots, T$ time periods. In the above, we let y_{it} denote the response variable at space i and time t , \mathbf{x}_{it} is a known $1 \times K$ vector of explanatory variables, and $\boldsymbol{\beta}$ is a $K \times 1$ vector of fixed but unknown regression parameters. The variable ϵ_{it} denotes an independently and identically distributed (iid) error term for space i and time t with zero mean and variance σ^2 . Depending on the application, there are several commonly used assumptions for λ_{it} . One commonly adopted assumption is $\lambda_{it} = \alpha_i$ which is used for models with unobserved space-specific effects only.

Using (2.1), we can specify the simple model on panel data in matrix form as follows,

$$\mathbf{y}_t = \mathbf{X}_t\boldsymbol{\beta} + \mathbf{u}_t, \quad (2.2)$$

where $\mathbf{y}_t = (y_{1t}, \dots, y_{Nt})'$ and $\mathbf{X}_t = (\mathbf{x}_{1t}, \dots, \mathbf{x}_{Nt})'$, respectively, denote the vector of the response variables and the matrix of the explanatory variables in all spaces at time t . Thus, $\mathbf{u}_t = (u_{1t}, \dots, u_{Nt})'$ is the model error component involving the sum of two disturbances. One of these disturbances is the $N \times 1$ vector of $\boldsymbol{\alpha}_t$ for which, with appropriate specifications, we can obtain the four types of spatial panel models described in Zhang *et al.* (2008) (referred to as *homogeneous*, *heterogeneous*, *fixed* and *random*). This allows the model to account for differences among different spaces which is referred to as spatial heterogeneity structure. The other disturbance in \mathbf{u}_t is the vector of the remainder disturbances $\boldsymbol{\epsilon}_t = (\epsilon_{1t}, \dots, \epsilon_{Nt})'$ in each period.

Let $\mathbf{y} = (\mathbf{y}'_1, \dots, \mathbf{y}'_T)'$ and $\mathbf{X} = (\mathbf{X}'_1, \dots, \mathbf{X}'_T)'$ be a concatenated form of \mathbf{y}_t and \mathbf{X}_t , respectively, such that they denote the response and explanatory variables, respectively, where \mathbf{y} denotes an $NT \times 1$ vector of the response variable and \mathbf{X} is a $NT \times K$ matrix of the non-stochastic exogenous regressors. In a similar manner,

the (concatenated form of the) disturbance vector can be written as

$$\mathbf{u} = \boldsymbol{\alpha} + \boldsymbol{\epsilon}, \quad (2.3)$$

where $\boldsymbol{\alpha} = (\boldsymbol{\iota}_T \otimes \mathbf{I}_N)\boldsymbol{\alpha}_t$, ($\boldsymbol{\iota}_p$ is a $p \times 1$ vector of ones, \mathbf{I}_p is a $p \times p$ identity matrix and the symbol \otimes denotes the Kronecker product) and $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}'_1, \dots, \boldsymbol{\epsilon}'_T)'$ denotes the $NT \times 1$ idiosyncratic error vector.

As pointed out before, the common assumption for the spatial panel models is a normal distribution for \mathbf{u} . As the normality assumption may be unrealistic when the responses are skewed, it will be valuable to explore multivariate skew-elliptical distributions as an alternative to the normal distribution to account for skewness in the data. Here, for the *homogeneous*, *heterogeneous*, *fixed* and *random* space-specific effects, we consider the skew-elliptical distributions for the remainder error term.

To implement the Bayesian inference for the aforementioned models, we need to specify the prior distribution for all unknown parameters in the model (2.2) in order to obtain the posterior distributions of the model parameters. A popular choice is to consider proper conditionally non-informative conjugate priors as suggested by Gelman and Hill (2006). Here we assign conjugate weakly-informative priors to obtain well defined and proper posteriors as in Jara *et al.* (2008) methodology. Note that, a family of prior distributions $p(\varphi)$ is conditionally conjugate for φ if the conditional posterior distribution $p(\varphi|y)$ is also in that class. In addition to these parameters, we also need to specify the prior distribution and derive the posterior distribution for different specifications of unobserved effects $\boldsymbol{\alpha}_t$.

In methodological papers and also when applying spatial models to real datasets, one needs to decide whether the unobserved space-specific effect α_i is to be treated as a random effect or a fixed effect. Traditionally, α_i is called a random effect when it is treated as a random variable and a fixed effect when it is treated as a parameter to be estimated for each cross-sectional observation i . When α_i is referred to an unobserved random effect, it is often assumed that they are uncorrelated with the \mathbf{x}_{it} ; that is, it is assumed that $E(\alpha_i|\mathbf{x}_{i1}, \dots, \mathbf{x}_{iT}) = 0$.

2.1 Spatial Heterogeneity Structure

In this paper, we suppose that the α_i are unobserved effects and that they can be the same or vary with spatial and temporal variation depending on the chosen specifications. We will consider all four spatial-temporal type models as specified in Zheng *et al.* (2008). The technical details are given in the following sections.

2.1.1 Homogeneous Model

In the case of the homogeneous model, the unobserved space-specific effects at time t are assumed to be space-homogeneous and time-homogeneous with $\boldsymbol{\alpha}_t = \alpha \boldsymbol{\iota}_N$, where α is an overall mean. The prior distribution for $\boldsymbol{\alpha}_t$ can be specified as follows,

$$\alpha \sim N(\alpha_0, \gamma_{\alpha_0}), \quad (2.4)$$

Thus the full conditional distribution of α is,

$$\alpha | \mathbf{y}_t, \boldsymbol{\beta}, \sigma_\epsilon^2, \delta_\epsilon \sim N(A_\alpha^{-1} a_\alpha, A_\alpha^{-1}), \quad (2.5)$$

where $A_\alpha = (NT/\sigma_\epsilon^2 + 1/\gamma_{\alpha_0})^{-1}$ and $a_\alpha = \sum_{t=1}^T \boldsymbol{\iota}'_N (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) / \sigma_\epsilon^2 + \alpha_0 / \gamma_{\alpha_0}$.

2.1.2 Heterogeneous Model

In the heterogeneous model, the unobserved space-specific effects are assumed to be space-homogeneous but time-inhomogeneous with $\boldsymbol{\alpha}_t = \alpha_t \boldsymbol{\iota}_N$, where α_t is a mean across spaces at time t . Suppose that $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_T)'$ denotes the vector of the means over time. In addition, we assume that

$$\boldsymbol{\alpha} \sim N_{NT}(\boldsymbol{\alpha}_0, \boldsymbol{\Gamma}_{\alpha_0}), \quad (2.6)$$

with $\boldsymbol{\Gamma}_{\alpha_0} = \gamma_{\alpha_0} \mathbf{I}_{NT}$, and thus the full conditional distribution of $\boldsymbol{\alpha}$ is given by,

$$\boldsymbol{\alpha} | \mathbf{y}, \boldsymbol{\beta}, \sigma_\epsilon^2 \sim N_{NT}(\mathbf{A}_\alpha^{-1} \mathbf{a}_\alpha, \mathbf{A}_\alpha^{-1}), \quad (2.7)$$

where $\mathbf{A}_\alpha = \boldsymbol{\Gamma}_{\alpha_0}^{-1} + \sigma_\epsilon^2 \mathbf{A}$ and $\mathbf{a}_\alpha = \boldsymbol{\Gamma}_{\alpha_0}^{-1} \boldsymbol{\alpha}_0 + \sigma_\epsilon^2 \mathbf{B}$, such that \mathbf{A} is a $T \times T$ diagonal matrix with diagonal elements $\mathbf{A}(t, t) = \boldsymbol{\iota}_N \boldsymbol{\iota}'_N$ and \mathbf{B} is a T dimensional vector with the t th element $\mathbf{B}(t) = \boldsymbol{\iota}'_N (\mathbf{y}_t - \tilde{\mathbf{X}}_t \boldsymbol{\beta})$ where $\tilde{\mathbf{X}}_t = \boldsymbol{\iota}_t \otimes \mathbf{X}_t$.

2.1.3 Fixed-effects Model

In this model, the unobserved space-specific effects are assumed to be time invariant and thus we can set $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\alpha}}$ with no time subscript. Note that, however, this vector is assumed to be time-homogeneous but space-inhomogeneous, that is, $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\alpha}} = (\alpha_1, \dots, \alpha_N)'$ in which α_i is a mean over time for space i . We use the following prior distribution for $\tilde{\boldsymbol{\alpha}}$,

$$\tilde{\boldsymbol{\alpha}} \sim N_N(\boldsymbol{\alpha}_0, \boldsymbol{\Gamma}_{\alpha_0}), \quad (2.8)$$

with $\mathbf{\Gamma}_{\alpha_0} = \gamma_{\alpha_0} \mathbf{I}_N$. Thus the full conditional distribution of $\tilde{\boldsymbol{\alpha}}$ is,

$$\tilde{\boldsymbol{\alpha}} | \mathbf{y}_t, \boldsymbol{\beta}, \sigma_\epsilon^2 \sim N_N(\mathbf{A}_{\tilde{\boldsymbol{\alpha}}}^{-1} \mathbf{a}_{\tilde{\boldsymbol{\alpha}}}, \mathbf{A}_{\tilde{\boldsymbol{\alpha}}}^{-1}), \quad (2.9)$$

where $\mathbf{A}_{\tilde{\boldsymbol{\alpha}}} = \mathbf{\Gamma}_{\alpha_0}^{-1} + T/\sigma_\epsilon^2 \mathbf{I}_N$ and $\mathbf{a}_{\tilde{\boldsymbol{\alpha}}} = \mathbf{\Gamma}_{\alpha_0}^{-1} \boldsymbol{\alpha}_0 + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) / \sigma_\epsilon^2 \boldsymbol{\iota}_N$.

2.1.4 Random-effects Model

In this case, the unobserved space-specific effects vector are assumed to be time-homogeneous but space-inhomogeneous with $\boldsymbol{\alpha}_t = \tilde{\boldsymbol{\alpha}} = (\alpha_1, \dots, \alpha_N)'$, where the α_i 's are iid random variables with $N(\alpha^*, \sigma_\alpha^2)$ distribution. Thus the full conditional distribution of $\tilde{\boldsymbol{\alpha}}$ is,

$$\tilde{\boldsymbol{\alpha}} | \mathbf{y}, \boldsymbol{\beta}, \sigma_\epsilon^2, \sigma_\alpha^2, \alpha^* \sim N_N(\mathbf{A}_{\tilde{\boldsymbol{\alpha}}}^{-1} \mathbf{a}_{\tilde{\boldsymbol{\alpha}}}, \mathbf{A}_{\tilde{\boldsymbol{\alpha}}}^{-1}), \quad (2.10)$$

where $\mathbf{A}_{\tilde{\boldsymbol{\alpha}}} = (1/\sigma_\alpha^2 + T/\sigma_\epsilon^2) \mathbf{I}_N$ and $\mathbf{a}_{\tilde{\boldsymbol{\alpha}}} = \alpha^* \boldsymbol{\iota}_N / \sigma_\alpha^2 + \sum_{t=1}^T (\mathbf{y}_t - \mathbf{X}_t \boldsymbol{\beta}) / \sigma_\epsilon^2 \boldsymbol{\iota}_N$. We use the following prior distributions for α^* and σ_α^2 ,

$$\alpha^* \sim N(\alpha_0^*, \gamma_{\alpha_0}^*), \quad \sigma_\alpha^2 \sim IG(\sigma_{\alpha_0}^2, \gamma_{\alpha_0}). \quad (2.11)$$

The full conditional distribution of α^* is then,

$$\alpha^* | \tilde{\boldsymbol{\alpha}}, \sigma_\alpha^2 \sim N(\alpha_n^*, \gamma_{\alpha_n}^*), \quad (2.12)$$

where $\alpha_n^* = \gamma_{\alpha_n}^* (\alpha_0^* / \gamma_{\alpha_0}^* + \boldsymbol{\iota}_N' \tilde{\boldsymbol{\alpha}} / \sigma_\alpha^2)$ and $\gamma_{\alpha_n}^* = (1/\gamma_{\alpha_0}^* + N/\sigma_\alpha^2)^{-1}$. Also, the full conditional distribution of σ_α^2 is,

$$\sigma_\alpha^2 | \tilde{\boldsymbol{\alpha}}, \alpha^* \sim IG(\sigma_{\alpha_0}^2 + N/2, \gamma_{\alpha_0} + (\tilde{\boldsymbol{\alpha}} - \alpha^* \boldsymbol{\iota}_N)' (\tilde{\boldsymbol{\alpha}} - \alpha^* \boldsymbol{\iota}_N) / 2). \quad (2.13)$$

By allowing for skewness in the error components, it provides great flexibility over the traditional spatial panel data model. The proposed panel model incorporates all types of spatial heterogeneity mentioned above and is referred to as a skew-elliptical spatial heterogeneity mixed model which can be fitted using the Bayesian estimation approach described in Jara *et al.* (2008).

3 Multivariate Skew-elliptical Distributions

In the following, we give the formal definition of the unrestricted multivariate skew-normal (uMSN) and unrestricted multivariate skew-t (uMST) distributions according to Lee and McLachlan (2013) formulations as special cases of the class of skew-

elliptical distributions by Arellano-Valle and Genton (2005). This type of multivariate skew-normal distribution was studied in Sahu *et al.* (2003). Suppose \mathbf{z}_0 and \mathbf{z}_1 are jointly normally distributed as

$$\begin{bmatrix} \mathbf{z}_0 \\ \mathbf{z}_1 \end{bmatrix} \sim N_{2n}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{bmatrix}\right). \quad (3.1)$$

Let \mathbf{z} denote a n -dimensional random vector. Then $\mathbf{z} = \boldsymbol{\mu} + \mathbf{\Delta}|\mathbf{z}_0| + \mathbf{z}_1$ defines the convolution-type stochastic representation of the unrestricted multivariate skew-normal distribution, and its density is given by

$$f(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Delta}) = 2^n \phi_n(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma}) \times \Phi_n(\mathbf{\Delta}'\mathbf{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu}); \mathbf{0}, \mathbf{\Lambda}) \quad (3.2)$$

where $\mathbf{\Omega} = \mathbf{\Sigma} + \mathbf{\Delta}\mathbf{\Delta}'$ and $\mathbf{\Lambda} = \mathbf{I}_n - \mathbf{\Delta}'\mathbf{\Omega}^{-1}\mathbf{\Delta} = (\mathbf{I}_n + \mathbf{\Delta}'\mathbf{\Sigma}^{-1}\mathbf{\Delta})^{-1}$. In the above, $\boldsymbol{\mu}$ is a $n \times 1$ vector of location parameter, $\mathbf{\Sigma}$ is a $n \times n$ diagonal positive definite scale matrix and $\mathbf{\Delta}$ is a $n \times n$ diagonal matrix of skewness parameters. Further, $\phi_n(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ and $\Phi_n(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$, respectively, denote the probability density function (pdf) and the cumulative distribution function (cdf) of a $N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$ random variable. Note that when $\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{\Sigma} = \mathbf{I}_n$; $\phi_n(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ and $\Phi_n(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma})$ will be denoted by $\phi_n(\cdot)$ and $\Phi_n(\cdot)$ respectively.

We shall follow the terminology of Lee and McLachlan (2013) and adopt the notation $\mathbf{z} \sim uSN_n(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{\Delta})$ if \mathbf{z} has the uMSN density. It is noted that if $\mathbf{\Delta} = \mathbf{0}$, the second part of (3.2) evaluates to 2^{-n} , and we again recover the multivariate normal density $\phi_n(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma})$. Then the density of \mathbf{z} reduces to the usual symmetric multivariate density, $N_n(\boldsymbol{\mu}, \mathbf{\Sigma})$.

Following Sahu *et al.*(2003), the uMSN distribution admits a convenient hierarchical representation. Let $\mathbf{w} = |\mathbf{z}_0|$, the hierarchical representation of (3.2) is given by

$$\begin{aligned} \mathbf{z}|\mathbf{w} &\sim N_n(\boldsymbol{\mu} + \mathbf{\Delta}\mathbf{w}, \mathbf{\Sigma}), \\ \mathbf{w} &\sim HN_n(\mathbf{0}, \mathbf{I}_N), \end{aligned} \quad (3.3)$$

where $HN_n(\mathbf{0}, \mathbf{\Sigma})$ represents the n -dimensional half-normal distribution with mean $\mathbf{0}$ and scale matrix $\mathbf{\Sigma}$.

Analogous to the skew-normal case, the uMST distribution has a similar stochastic representation to the uMSN distribution. Specifically, $\mathbf{z} = \boldsymbol{\mu} + \mathbf{\Delta}|\mathbf{z}_0| + \mathbf{z}_1$ has the uMST distribution, where conditional on the gamma variable r ,

$$\begin{bmatrix} \mathbf{z}_0 \\ \mathbf{z}_1 \end{bmatrix} \sim N_{2n}\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \frac{1}{r} \begin{bmatrix} \mathbf{I}_n & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{bmatrix}\right), \quad (3.4)$$

It follows that the density of \mathbf{z} is given by

$$f(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\Delta}, \tau) = 2^n t_n(\mathbf{z}; \boldsymbol{\mu}, \mathbf{\Omega}, \tau) T_n(\mathbf{c}(\mathbf{y}) \sqrt{\frac{\tau + n}{\nu + d(\mathbf{z})}}; \mathbf{0}, \mathbf{\Lambda}, \tau + n), \quad (3.5)$$

where $\mathbf{c}(\mathbf{z}) = \boldsymbol{\Delta}' \mathbf{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu})$ and $d(\mathbf{z}) = (\mathbf{z} - \boldsymbol{\mu})^T \mathbf{\Omega}^{-1}(\mathbf{z} - \boldsymbol{\mu})$. This density is expressed as the product of a multivariate t -student pdf and a multivariate t -student cdf. In fact, $T_n(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \tau)$, is the cdf of $t_n(\cdot; \boldsymbol{\mu}, \mathbf{\Sigma}, \tau)$. Here, the notation $\mathbf{z} \sim uST_n(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\Delta}, \nu)$ will be used. In the latter case, ν corresponds to the degrees of freedom parameters. When $\nu \rightarrow \infty$, the multivariate skew- t distribution approaches the multivariate skew-normal distribution and reduces to the multivariate normal distribution when $\boldsymbol{\Delta} = \mathbf{0}$.

Similar to the uMSN version, the uMST admits a convenient hierarchical representation,

$$\begin{aligned} \mathbf{z} | \mathbf{w} &\sim N_n(\boldsymbol{\mu} + \boldsymbol{\Delta} \mathbf{w}, \frac{1}{r} \mathbf{\Sigma}), \\ \mathbf{w} | r &\sim HN_n(\mathbf{0}, \frac{1}{r} \mathbf{I}_N), \\ r &\sim \text{gamma}(\frac{\tau}{2}, \frac{\tau}{2}). \end{aligned} \quad (3.6)$$

The MCMC computations for the skew-elliptical panel model will be derived based on the above spatial hierarchical Bayes models in expressions (3.3) and (3.6). These representations are important for obtaining the parameter estimates and will facilitate easy implementations in commonly used softwares such as R and OpenBUGS. Note that, for convenience, we use the simple stochastic representation of the skew-elliptical distributions when their means are not being modelled.

4 Model Selection

From a frequentist point of view, model assessment is based on the log likelihood. For Bayesian model selection, the Akaike Information Criterion (AIC) (Akaike, 1974), the Bayesian Information Criterion (BIC) (Schwarz, 1978) and the Hannan Quinn Criterion (HQC) (Hannan and Quinn, 1979) are used to assess the performance of asymmetric models and compare it to the symmetric models. They are

defined by $AIC = \hat{D} + 2P$, $BIC = \hat{D} + P \log(N)$, and $HQC = \hat{D} + 2P \log(\log(N))$, respectively, where \hat{D} is -2 times log likelihood evaluated at the maximum likelihood estimate, P is the number of parameters, and N is the sample size. It should be noted that a model with a smaller value of AIC, BIC, and HQC is preferred when comparing different fitted results.

5 Application

In this section, we illustrate the proposed spatial panel models by applying them to a dataset derived from records of insurance activities in Italy. Insurance is important for the economic development of a country and is usually classified into life and non-life. The life and non-life insurance consumption corresponds to different needs, so they usually are analyzed separately. Here, we focus on non-life insurance.

The economic rationale for purchasing non-life insurance is to get a (financial) protection for future losses by paying a premium, today, thus transferring future wealth from an indefinite to a definite state. According to economic theory, this branch of the insurance market is plays an important role in fostering the welfare and growth of the insurance industry and this is realized by protecting families and firms from the financial hardships caused by unexpected/unusual events such as fire, theft, disease, and accidents.

We analyze the consumption of non-life insurance across 103 Italian provinces in 1998-2002 in order to assess its determinants. The records include: real per-capita GDP[†] (rgdp), real per-capita bank deposits (bank), inhabitants density per square Km (den), real interest rate on lending to families and small enterprises (rirs), density of insurance agencies per 1000 inhabitants (agen), share of people with second grade schooling or more (school), share of value added, agriculture sector (vaagr), average number of family members (fam), judicial inefficiency index: average years to settle first degree of civil case (inef), survey result to the question "do you trust others?" (trust), real non-life insurance premiums per capita (ppcd), the province code (code) and the year of observation (year), reported in Millo and Carmeci (2010). These data are publicly available from `splm` package in R (Millo and Piras, 2012).

As outlined by Millo and Carmeci (2010), these data are challenging to analyze. In particular, it is difficult to estimate the determinants of the development of the nonlife insurance market for Italy and hence good proxies are needed. There is also a high spatial differentiation across the Italian provinces as well as a high degree of spatial correlation in insurance density. For this illustration, the explanatory

[†]Gross Domestic Product

variables are *rgdp*, *bank*, *den*, *rirs*, *agen*, *school*, *vaagr*, *fam*, *inef* and *trust* and the response variable is *ppcd*. For a discussion of the determinants of insurance, see the papers by Browne *et al.* (2000), Beck and Webb (2003), and Esho *et al.* (2004).

Let us consider the simple panel regression model (2.1) for the total number of non-life insurance premiums years in Italian provinces. The main objective here is to investigate the link between non-life insurance premiums and some determinants of insurance consumption in the model. The panel results can be affected by some observable (income, wealth, etc) and unobserved effects (capital stock, attitudes towards litigation and etc) on non-life insurance premiums as will be discussed below.

Data availability for Italian insurance premiums limits the analysis to the provincial level. Due to this, we expect some heterogeneity at the provincial level. Alternatively, one may choose to eliminate that part of the heterogeneity that is associated with differences between provinces. The unobserved space-specific effects are captured by the α_i terms. Therefore, following the arguments in Section 2.1, we consider different specifications for the unobserved vector $\boldsymbol{\alpha}$ and then assess the spatial panel models based on the estimation results.

In the *homogeneous* model the provincial unobserved effects are the same across both time and provinces, such that $\boldsymbol{\alpha}$ can be defined as

$$\boldsymbol{\alpha}' = (\alpha \cdots \alpha \cdots \cdots \alpha \cdots \alpha), \quad (5.1)$$

which is a 515×1 vector. In the *heterogeneous* model the provincial unobserved effects are the same across the provinces but are different over time. In this case, $\boldsymbol{\alpha}$ has the following structure,

$$\boldsymbol{\alpha}' = (\alpha_1 \cdots \alpha_1 \cdots \cdots \alpha_5 \cdots \alpha_5). \quad (5.2)$$

On the other hand, in the *fixed* model and *random* model, the provincial unobserved effects are assumed to be the same over time but are different across provinces. Hence, in these cases, $\boldsymbol{\alpha}$ is defined as follows

$$\boldsymbol{\alpha}' = (\alpha_1 \cdots \alpha_{103} \cdots \cdots \alpha_1 \cdots \alpha_{103}) = (\tilde{\boldsymbol{\alpha}} \cdots \cdots \tilde{\boldsymbol{\alpha}}), \quad (5.3)$$

where, in the *random* setting, $\boldsymbol{\alpha}$ follows a specified distribution. For our model, multivariate skew-elliptical distributions are used for the α_i terms. The above-mentioned structures provide four different ways to introduce unobserved space-specific heterogeneity in the constant terms or random variables of the panel data model. We employ the approach described in Jara and *et al.* (2008) for the estimation of model

parameters. In fact, we follow the Bayesian methodology to perform the inference procedure. For $t = 1, \dots, T$, we assume ϵ_t follow an iid $uSN_N(0, \sigma_\epsilon^2 \mathbf{I}_N, \delta_\epsilon \mathbf{I}_N)$ and $uST_N(0, \sigma_\epsilon^2 \mathbf{I}_N, \delta_\epsilon \mathbf{I}_N, \nu)$ distributions corresponding to normal and t-student distributions. These models allow us to account for the spatial heterogeneity among different spaces with skew-elliptical distributions. We assign independent prior distributions for β , σ_ϵ^2 and δ_ϵ and obtain samples from the posterior distribution of the parameters using MCMC methods.

In Millo and Carmeci (2010) the log transformation of the data is used to approximate normality, and then the normal process is applied to model the transformed data. However, it can be observed from the histogram of log-transformed data in Figure 1(a) that the data do not appear normal even after log transformation. The histogram of the raw insurance consumption data in Figure 1(b) also shows that the data look asymmetric. This indicates that the normality assumption is not satisfied and that fitting a skew model to the data set seems more appropriate. Furthermore, we note that the skewness index for the raw non-life insurance premiums is 0.47.

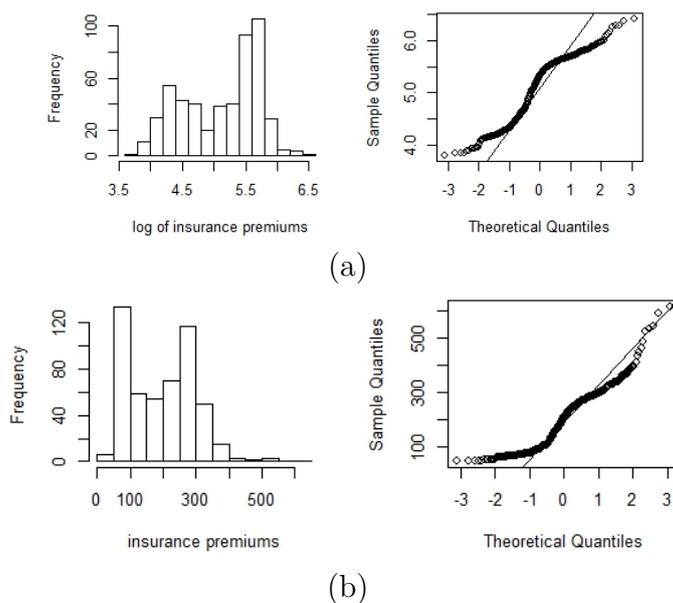


Figure 1: (a) Histogram and Q-Q plot of Italian insurance premiums with log-transformed non-life insurance data. (b) Histogram and Q-Q plot of Italian insurance premiums with raw non-life insurance data.

The histograms in Figure 2 display the distribution of the residuals obtained after fitting the four types of spatial panel models with normal assumption. Clearly, the distributions of the residuals seem to be positively skewed and heavy tailed. Since we use the proposed spatial heterogeneity panel models for the skewed data, it does not require the use of data transformation.

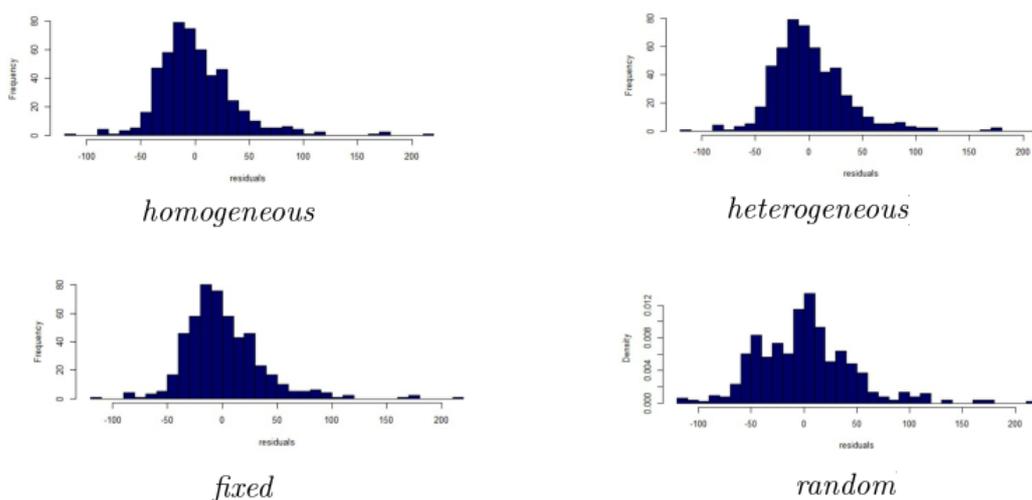


Figure 2: Non-life insurance consumption in Italy data. Histogram of the residuals, for each space-specific effect type, obtained after fitting the multivariate normal distribution.

Table 1: The posterior mean (standard deviation) and equitailed 95% credible intervals for the parameters in the normal spatial-temporal models fitted to the Italian non-life insurance data.

par.	homogeneous	heterogeneous	fixed	random
rgdp	1.216e-2(9.224e-4) (1.038e-2,0.01402)	1.128e-2(9.328e-4) (9.490e-3,1.315e-2)	1.231e-2(9.929e-4) (1.044e-2,1.426e-2)	1.217e-2(8.437e-4) (1.052e-2,0.01379)
bank	9.259e-3(9.931e-4) (7.248e-3,0.01109)	9.495e-3(9.478e-4) (7.592e-3,1.132e-2)	8.984e-3(1.069e-3) (6.788e-3,1.101e-2)	9.165e-3(1.023e-3) (7.153e-3,0.01112)
den	2.228e-2(5.869e-3) (1.080e-2,0.03381)	1.841e-2(5.343e-3) (7.505e-3,2.830e-2)	2.302e-2(5.437e-3) (1.232e-2,3.375e-2)	2.248e-2(5.732e-3) (1.043e-2,0.03369)
rirs	-9.458(2.004) (-1.341e+1,-5.457)	-1.111e+1(1.916) (-1.506e+1,-7.56)	-9.197(1.783) (-1.256e+1,-5.756)	-9.596(2.052) (-1.375e+1,-5.657)
agen	8.688(9.109) (-9.843,25.89)	7.390(9.144) (-1.095e+1,2.443e+1)	8.669(9.129) (-8.363,2.601e+1)	9.058(9.450) (-1.001e+1,27.29)
school	-4.492e-1(3.880e-1) (-1.189,0.2806)	1.655e-2 (3.781e-1) (-6.878e-1,7.835e-1)	-4.911e-1(4.002e-1) (-1.314,2.935e-1)	-4.773e-1(3.752e-1) (-1.170,0.2391)
vaagr	-3.738e-2(7.839e-1) (-1.539,1.514)	-4.606e-1(7.468e-1) (-1.953, 9.897e-1)	-9.179e-2(7.895e-1) (-1.662,1.345)	-5.056e-2(7.647e-1) (-1.499,1.575)
fam	-3.253e+1(6.856) (-4.509e+1,-18.62)	-2.359e+1(6.146) (-3.639e+1,-1.182e+1)	-3.246e+1(6.212) (-4.474e+1,-2.048e+1)	-3.213e+1(6.155) (-4.437e+1,-19.62)
inef	-3.889(1.380) (-6.742,-1.262)	-3.733(1.406) (-6.384,-9.338e-1)	-4.106(1.491) (-7.082,-1.065)	-3.977(1.517) (-6.957,-0.815)
trust	1.773e+1(7.057) (5.151,33.6)	1.245e+1(5.902) (1.309, 2.375e+1)	1.841e+1(7.025) (4.572,3.252e+1)	1.859e+1(6.464) (6.102,31.4)
σ_ϵ^2	1.295e+3(7.969e+1) (1.150e+3,1454)	1.141e+3(7.465e+1) (1.006e+3,1.296e+3)	1.277e+3(8.410e+1) (1.116e+3,1.449e+3)	1.298e+3(8.121e+1) (1.151e+3,1470)
- loglike	2576	2537	2564	2569.5
AIC	5170	5102	5156	5173
BIC	5102.18	5102.18	5156.18	5173.218
HQC	5150.507	5082.507	5136.507	5149.329

For the spatial heterogeneity panel models, details of implementation of the Gibbs sampler algorithm are as follows. For the skew-elliptical proposal distribu-

Table 2: The posterior mean (standard deviation) and equitailed 95% credible intervals for the parameters in the skew-normal spatial-temporal models fitted to the Italian non-life insurance data.

par.	homogeneous	heterogeneous	fixed	random
rgdp	1.192e-2(7.428e-4) (1.051e-2,1.332e-2)	1.215e-2(7.926e-4) (1.077e-2,1.372e-2)	1.196e-2(6.936e-4) (1.071e-2,1.341e-2)	1.197e-2(6.525e-4) (1.075e-2,0.01329)
bank	9.031e-3(7.732e-4) (7.509e-3,1.054e-2)	8.912e-3(8.199e-4) (7.292e-3,1.047e-2)	8.840e-3(7.374e-4) (7.344e-3,1.027e-2)	8.899e-3(7.837e-4) (7.356e-3,0.01042)
den	1.285e-2(4.298e-3) (4.518e-3,2.075e-2)	1.255e-2(4.500e-3) (3.798e-3,2.095e-2)	1.294e-2(4.426e-3) (4.143e-3,2.142e-2)	1.327e-2(4.276e-3) (4.705e-3,0.02175)
rirs	-1.048e+1(1.722) (-1.432e+1,-7.298)	-1.017e+1(1.847) (-1.346e+1,-6.624)	-1.064e+1(1.684) (-1.409e+1,-7.497)	-1.005e+1(1.628) (-1.307e+1,-6.854)
agen	-1.058(8.997) (-1.744e+1,1.646e+1)	-1.341(8.999) (-1.880e+1,1.662e+1)	-7.905e-2(8.897) (-1.771e+1,1.723e+1)	-1.002e-1(9.506) (-1.903e+1,18.31)
school	-8.819e-1(3.084e-1) (-1.515,-2.959e-1)	-8.851e-1(4.017e-1) (-1.710,-1.255e-1)	-1.015(3.222e-1) (-1.692,-4.504e-1)	-9.644e-1(3.598e-1) (-1.756,-0.271)
vaagr	4.866e-1(7.126e-1) (-8.814e-1,1.903)	4.478e-1(6.807e-1) (-9.469e-1,1.848)	3.935e-1(7.337e-1) (-1.064,1.788)	3.719e-1(6.696e-1) (-9.959e-1,1.685)
fam	-2.685e+1(5.192) (-3.692e+1,-1.616e+1)	-2.586e+1(6.149) (-3.737e+1,-1.413e+1)	-2.749e+1(6.645) (-3.988e+1,-1.586e+1)	-2.807e+1(6.277) (-4.011e+1,-15.81)
inef	-4.360(1.280) (-6.911,-1.99)	-4.305(1.244) (-6.678,-1.871)	-4.453(1.315) (-7.088,-1.680)	-4.515(1.321) (-6.987,-1.865)
trust	1.400e+1(4.970) (4.548,2.407e+1)	1.080e+1(5.834) (8.468e-1,2.354e+1)	1.573e+1(7.128) (3.061,3.143e+1)	1.454e+1(6.204) (6.891e-1,25.47)
σ_ϵ^2	3.074e+2(5.867e+1) (2.093e+2,4.436e+2)	3.027e+2(5.861e+1) (2.071e+2,4.326e+2)	2.845e+2(5.638e+1) (1.900e+2,4.127e+2)	2.945e+2(6.045e+1) (1.845e+2,422.2)
δ_ϵ	4.812e+1(2.507) (4.313e+1,5.286e+1)	4.788e+1(2.473) (4.307e+1,5.273e+1)	4.804e+1(2.584) (4.302e+1,5.299e+1)	4.795e+1(2.647) (4.249e+1,53.17)
- loglike.	2383.5	2378	2359.5	2366
AIC	4797	4786	4749	4768
BIC	4797.19	4786.19	4749.19	4768.23
HQC	4766.11	4765.11	4728.11	4742.94

tions, to proceed with Bayesian inference, we ran the Gibbs sampler algorithm with 20,000 iterations, discarded the 20,000 initial iterates and stored every 20th iteration. Our findings in Table 1 to 4 are rather diverse, with coefficients often changing sign and numerical values, which can be seen as evidence in favour of the need to investigate on skewness and unobserved heterogeneity in panel studies. Based on the AIC, BIC, and HQC, the results indicate that skew-t model may be the best fitting model and skew-normal model is next, supporting the contention of a departure from normality.

The population posterior mean, the corresponding standard deviation, and 95% credible interval for fixed-effects parameters, variance component of error and skewness component of error are presented in Table 1 to 4. We found that the parameter estimates among the four models are different from those obtained with symmetric and asymmetric distributions. The estimated results for the most model coefficients,

Table 3: The posterior mean (standard deviation) and equitailed 95% credible intervals for the parameters in the t-student spatial-temporal models fitted to the Italian non-life insurance data.

par.	homogeneous	heterogeneous	fixed	random
rgdp	1.156e-2(7.548e-4) (1.017e-2,1.306e-2)	1.105e-2(8.107e-4) (9.548e-3,1.264e-2)	1.189e-2(6.992e-4) (1.050e-2,1.317e-2)	1.188e-2(7.747e-4) (1.035e-2,1.335e-2)
bank	1.087e-2(8.571e-4) (9.282e-3,1.259e-2)	1.086e-2(9.448e-4) (8.996e-3,1.275e-2)	9.652e-3(8.392e-4) (7.965e-3,1.132e-2)	9.908e-3(9.333e-4) (8.023e-3,1.183e-2)
den	1.670e-2 (5.091e-3) (7.462e-3,2.851e-2)	1.410e-2(5.190e-3) (5.160e-3,2.504e-2)	1.948e-2(5.083e-3) (1.087e-2,3.095e-2)	1.865e-2(5.618e-3) (8.682e-3,3.164e-2)
rirs	-7.928(1.701) (-1.104e+1,-4.233)	-8.645(1.735) (-1.224e+1,-5.131)	-7.982(1.511) (-1.090e+1,-5.120)	-8.052(1.806) (-1.188e+1,-4.347)
agen	-1.774e+1(1.801e+1) (-6.177e+1,8.238)	-1.037e+1(1.348e+1) (-4.328e+1,1.219e+1)	-7.459(1.255e+1) (-3.880e+1,1.341e+1)	-9.404(1.337e+1) (-4.036e+1,1.164e+1)
school	-2.783e-1(3.035e-1) (-8.609e-1,3.277e-1)	-2.748e-1(3.020e-1) (-8.655e-1,3.286e-1)	-8.653e-1(2.631e-1) (-1.426,-3.891e-1)	-7.456e-1(2.757e-1) (-1.313,-2.443e-1)
vaagr	1.333(6.189e-1) (1.630e-1,2.556)	5.356e-1(6.331e-1) (-6.978e-1,1.783)	1.050(6.503e-1) (-2.175e-1,2.355)	1.174(6.483e-1) (-5.735e-2,2.433)
fam	-4.129e+1(9.229) (-6.039e+1,-2.412e+1)	-3.265e+1(7.146) (-4.654e+1,-1.904e+1)	-5.223e+1(7.596) (-6.591e+1,-3.788e+1)	-4.886e+1(8.200) (-6.436e+1,-3.578e+1)
inef	-7.605e-1(1.066) (-2.896,1.344)	-5.303e-1(1.011) (-2.527,1.494)	-1.325(1.115) (-3.395,9.217e-1)	-1.266(1.049) (-3.325,7.607e-1)
trust	2.680e+1(1.165e+1) (8.301,5.363e+1)	1.202e+1(7.515) (-1.515,3.166e+1)	3.349e+1(8.819) (1.318e+1,5.008e+1)	2.875e+1(9.178) (1.469e+1,4.687e+1)
σ_ϵ^2	4.076e+2(3.987e+1) (3.331e+2,4.902e+2)	3.969e+2(3.889e+1) (3.251e+2,4.814e+2)	3.548e+2(3.747e+1) (2.821e+2,4.337e+2)	3.697e+2(4.233e+1) (2.958e+2,4.607e+2)
- loglike	2518.5	2501.5	2463.5	2479.5
AIC	506	5031	4955	4993
BIC	5065.18	5031.18	4955.18	4989.19
HQC	5045.51	5011.51	4935.51	4969.33

show that the 95% credible intervals of these posterior means do not include zero, indicating that these estimates are significantly different from zero, but estimates appear quite different. Moreover, the parameters estimated for each of the two spatial panel models with skew-normal and skew-t error terms can be summarized as follows: (i) The 95% equal-tail credible intervals means associated with each parameter fitted values from the symmetric models are wider than that of the corresponding asymmetric models; (ii) for the skew models, the parameters estimated have a smaller standard deviation compared to that for the symmetric models; (iii) the estimate of the variance component of error is smaller in the class of skew models compared to the symmetric ones, which is expected because skew-normal and skew-t models take into account skewness of the data; (iv) the estimated 95% equal-tail credible intervals of δ_ϵ do not include zero. This finding suggests that there is a significantly positive skewness in the data and confirms the fact that the distribution of the original insurance data is skewed. These results indicate that consideration of departures from normality can lead to an improvement in model fit. According to the above findings, skew-t is the favored (best) model.

Table 4: The posterior mean (standard deviation) and equitailed 95% credible intervals for the parameters in the skew-t spatial-temporal models fitted to the Italian non-life insurance data.

par.	homogeneous	heterogeneous	fixed	random
rgdp	0.01216(5.897e-4) (1.085e-2,0.01315)	1.090e-2(6.665e-4) (9.740e-3,0.01237)	1.180e-2(7.757e-4) (1.046e-2,0.01321)	1.179e-2(6.352e-4) (1.050e-2,0.01304)
bank	0.00988(6.161e-4) (8.647e-3,0.01103)	9.916e-3(7.078e-4) (8.340e-3,0.01115)	9.416e-3(6.760e-4) (8.012e-3,0.01067)	9.965e-3(5.393e-4) (8.859e-3,0.01102)
den	0.01541(3.392e-3) (8.595e-3,0.02194)	1.231e-2(3.200e-3) (5.708e-3,0.01846)	1.706e-2(3.554e-3) (1.017e-2,0.02391)	1.451e-2(3.223e-3) (7.833e-3,0.0206)
rirs	-7.034(1.522) (-9.768,-3.906)	-9.835(1.902) (-1.355e+1,-6.491)	-7.572(1.593) (-9.986,-3.383)	-7.577(1.423) (-1.032e+1,-4.819)
agen	-38.07(1.977e+1) (-7.496e+1,-1.522)	-2.093e+1(1.725e+1) (-6.536e+1, 6.566)	-2.004e+1(1.747e+1) (-5.844e+1,4.77)	-3.148e+1(1.969e+1) (-7.172e+1,1.646)
school	-0.5176(2.415e-1) (-9.452e-1,-0.03139)	-4.010e-1(2.533e-1) (-9.241e-1,0.05182)	-7.730e-1(2.864e-1) (-1.392,-0.2414)	-3.615e-1(1.976e-1) (-7.985e-1,-0.02606)
vaagr	0.6778(5.550e-1) (-4.967e-1,1.726)	2.663e-2(5.453e-1) (-1.018,1.138)	4.791e-1(5.950e-1) (-6.809e-1,1.661)	5.487e-1(5.580e-1) (-5.228e-1,1.687)
fam	-46.81(5.183) (-5.667e+1,-36.08)	-2.870e+1(7.510) (-4.441e+1,-14.41)	-5.136e+1(5.045) (-6.219e+1,-41.9)	-4.262e+1(4.768) (-5.065e+1,-32.54)
inef	-3.022(9.671e-1) (-4.708, -0.971)	-3.689(1.08) (-5.598,-1.455)	-3.007(1.094) (-5.147,-0.8216)	-3.435(1.159) (-5.615,-0.993)
trust	18.82(5.752) (8.165,28.43)	1.286e+1(5.426) (4.427,24.180)	2.816e+1(4.893) (1.890e+1,37.03)	1.577e+1(4.967) (6.391,25.16)
σ_ϵ^2	73.04(2.586e+1) (3.445e+1,134.3)	7.536e+1(2.548e+1) (3.662e+1,139.3)	7.284e+1(2.645e+1) (3.074e+1,134.3)	6.371e+1(2.126) (3.031e+1,107.9)
δ_ϵ	38.99(3.273) (3.253e+1,45.34)	3.742e+1(3.005) (3.114e+1,42.65)	3.644e+1(3.26) (2.990e+1,42.49)	3.995e+1(3.012) (3.409e+1,46.1)
- loglike.	2244.5	2259	2244	2202
AIC	4519	4546	4516	4436
BIC	4519.19	4546.18	4516.18	4436.21
HQC	4498.11	4526.51	4496.51	4413.72

Most authors (e.g., Beenstock *et al.*, 1988, Outreville, 1990, Enz, 2000, Esho *et al.* 2004 and Millo and Carmeci, 2010) have commented on the elasticity of insurance consumption with respect to income and wealth. Elasticity estimates are frequently used as one of the basic indicators as they are unit-free, easily interpreted, and comparable across studies. In econometrics models with log-transformed data, elasticity values are estimated as the coefficients of the independent variables when the features of the dependent variable, estimation technique, and model structure are determined. Although there are only minor differences among the coefficient estimates and elasticities.

In a wider economic sense, the elasticity value of insurance consumption to income is an interesting subject for research. According to Millo and Carmeci (2010), model specifications and real GDP per capita account for both income and the general level of economic activity, and real bank deposits per capita, as an instrument for the stock of wealth. A negative value for income elasticity is considered consis-

tent with the hypothesis of insurance as an inferior good (Millo and Carmeci, 2010). In empirical studies, it is discussed whether insurance is a superior or a normal good. If income elasticity is positive and smaller than one, then insurance is considered as a normal good. On the contrary, if elasticity is greater than one, insurance is a superior good.

The results of the different regression coefficients estimates (posterior means) in Tables 1 to 4 are reasonable given the observable insurance determinantes. In this research, the elasticity of insurance consumption with respect to income and wealth, which is important in spatial econometrics applications areas, is computed. The posterior means for skew panel models with unobserved space-specific heterogeneity in the constant terms are largely consistent with each other despite some numerical differences. Concentrating on the results of the *random* model with skew-t distribution in Table 4, the elasticity of insurance consumption with respect to income (1.049) turns out to be statistically positive and greater than one, thus asserting the view of non-life insurance as a superior good. The elasticity of wealth is positive but lower than one (0.440), suggesting that the tendency to insure is actually decreasing with wealth. Inhabitants density proves negative and significant, supporting the claim that it is not a good agent for risk conditions. The effect of interest rates is significantly negatively related with insurance consumption. The density of agencies turns out as a positive exciter, consistently with the view that insurance is a complicated good with a substantial cost of searching for an appropriate contract. Family numerosity plays a weak role with a negative but significant coefficient, and so does schooling; person capital does not seem to apply an impression on the non-life insurance market. The share of agriculture has a significant and positive coefficient. Judicial system inefficiency has a meaningfully negative impact on insurance. This confirms with the argument that bad implementation of property laws negatively affects peoples willingness to insure. For probably similar reasons as in Guiso *et al.* (2004), trust is an important positive determinant of insurance. These findings are important for analyzing the economic development of the non-life insurance market for Italy.

The Markov chain history and density for some parameters under *random* skew-t spatial panel model is presented in Figures 3 and 4, respectively. Note that convergence of the MCMC is confirmed with the following plots.

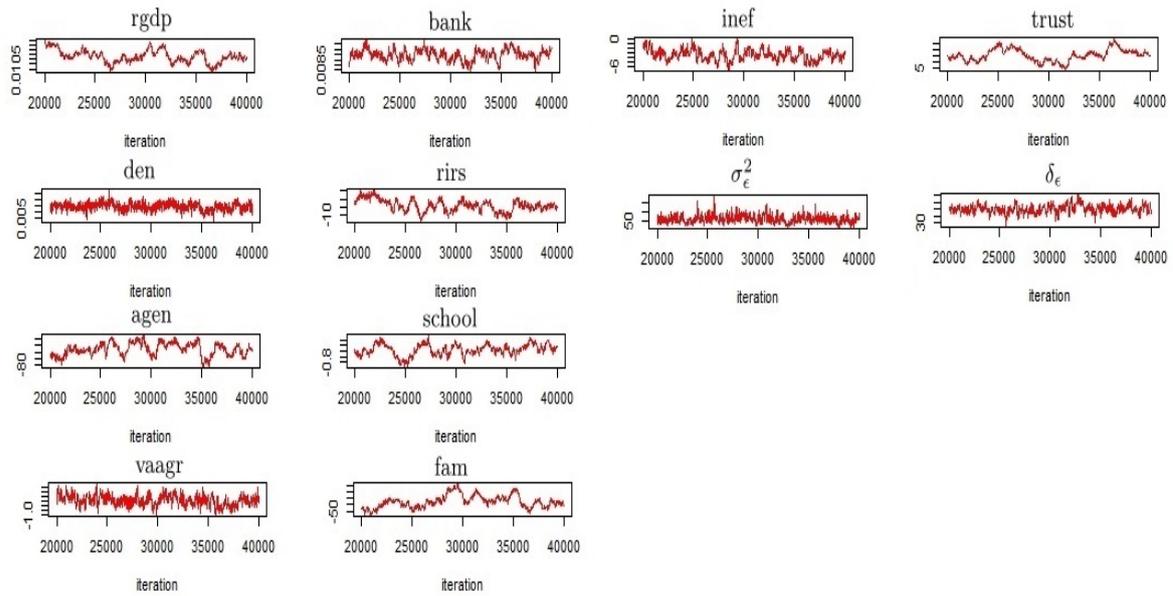


Figure 3: Markov chain history for parameters under *random* model of the Italian non-life insurance data.

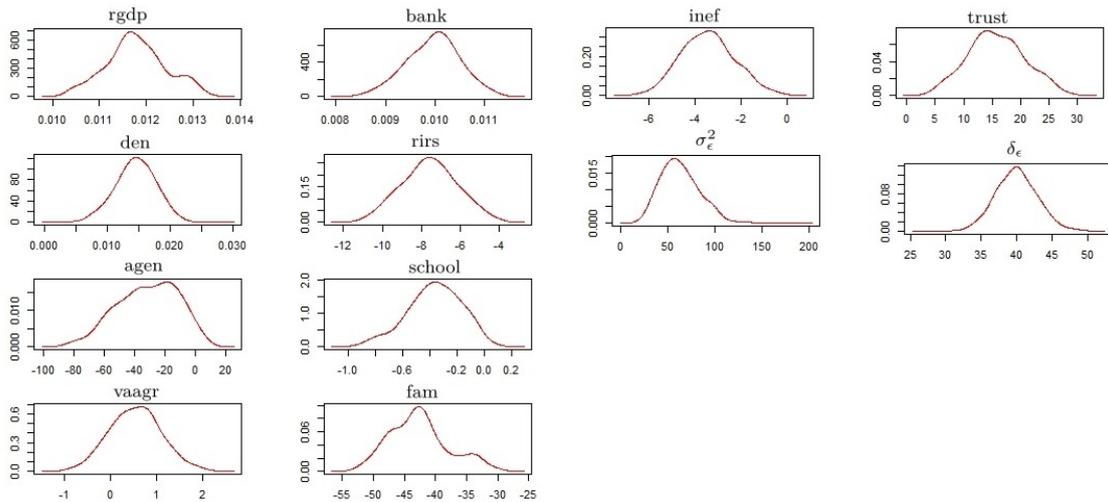


Figure 4: Markov chain density for parameters under *random* model of the Italian non-life insurance data.

6 Conclusion

In previous studies, the application of spatial heterogeneity, or the variation in relationships across spaces, in spatial panel data model rarely examined the behaviour of the error terms. In this paper, we have considered multivariate skew-elliptical distributions for the remainder error components of the spatial panel data model. Specifically, our methodology is suitable for modelling data that exhibit skewness.

It provides a direct approach to model these data without the need to apply transformations. The approach of allowing for skewness in the error terms of the econometrics is new and differs from the procedures used by other studies to calculate other econometrics parameters estimates methods. Parameter estimation for our model was carried out using the Bayesian hierarchical approach. The model fitting results for each of the sub models of our skew-elliptical spatial heterogeneity models were investigated by applying several criteria. To demonstrate the usefulness of our proposed methodology, we conducted a simulation study and applied our models to a real datasets, namely the insurance consumption data sets. The results from these analyses demonstrated that it is very important to take into account the skewness in the data. Adopting a skew-elliptical spatial panel data model for the response models can achieve more reliable results in the estimation of econometrics parameters. To achieve this, we implemented the MCMC sampling scheme using the R software. The non-life insurance consumption data set is analyzed to demonstrate the proposed methodologies and the analysis results are reported. We have also discussed the elasticity of insurance consumption with respect to income and wealth. The use of non-normal remainder error components in the panel models not only provides much improved results, but also avoided the need to perform transformation or to work with the unrealistic normality assumption.

References

- Akaike, H. (1974), A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, **19**, 716-723.
- Arellano-Valle, R.B., Genton, M.G. (2005), On fundamental skew distributions, *Journal of Multivariate Analysis*, **96**, 93-116.
- Arellano-Valle, R.B., Azzalini, A. (2006), On the unification of families of skew-normal distributions. *Scand. J. Stat.*, **33**, 561-574.
- Azzalini, A. (2005), The skew-normal distribution and related multivariate families, *Scand. J. Stat.*, **32**, 159-188.
- Azzalini, A. (2014), The skew-normal and related families.
- Beck, T., Webb I. (2003), Economic, demographic and institutional determinants of life insurance consumption across countries, *World Bank Economic Review*, **17**, 51-88.
- Beenstock, M., Dickinson, G., Khajuria, S. (1988), The relationship between property-liability insurance premiums and income: an international analysis, *The Journal of Risk and Insurance*, **55**, 259-272.
- Browne, M.J., Chung, J., Frees, E.W. (2000), International property-liability insurance consumption, *The Journal of Risk and Insurance*, **67**, 73-90.

- Chesson, P. L. (1985), Coexistence of competitors in spatially and temporally varying environments: a look at the combined effects of different sorts of variability, *Theor. Popul. Biol.*, **28**, 263-287.
- Enz, R. (2000), The s-curve relation between per-capita income and insurance penetration, *Geneva Papers on Risk and Insurance: Issues and Practice*, **25**, 396-406.
- Esho, N.A., Kirievsky, A., Ward, D., Zurbruegg, R. (2004), Law and the determinants of property-casualty insurance, *The Journal of Risk and Insurance*, **71**, 265-283.
- Gelman, A., Hill, J. (2006), Data analysis using regression and multi-level/hierarchical models, *Cambridge university press*.
- Guiso, L., Sapienza, P., Zingales, L. (2004), The role of social capital in financial development, *American Economic Review*, **94**, 526-556.
- Hannan, E.J., Quinn, B.G. (1979), The determination of the order of an autoregression, *Journal of the Royal Statistical Society Series B*, **41**, 190-195.
- Jara, A., Quintana, F., San Martin, E. (2008), Linear mixed models with skew-elliptical distributions: A Bayesian approach, *Computational statistics and data analysis*, **52(11)**, 5033-5045.
- Lee, S.X., Mclachlan, G.J. (2013), On mixtures of skew normal and skew t-distributions, *Advances in Data Analysis and Classification*, **7**, 241-266.
- Lunn, D., Spiegelhalter, D., Thomas, A., Best, N. (2009), The BUGS project: Evolution, critique and future directions, *Statistics in Medicine*, **28**, 3049-3067.
- Millo, G., Carmeci, G. (2010), Non-life insurance consumption in Italy, *A sub-regional panel data analysis*.
- Millo, G., Piras, G. (2012), splm: Spatial panel data models in R, *Journal of Statistical Software*, **47**, 1-38.
- Outreville, J.F. (1990), The economic significance of insurance markets in developing countries, *The Journal of Risk and Insurance*, **18**, 487-498.
- Sahu, S.K., Dey, D.K., Branco, M.D. (2003), A new class of multivariate skew distributions with applications to Bayesian regression models, *Canadian Journal of Statistics*, **31**, 129-150.
- Schwarz, G. (1978), Estimating the dimension of a model, *The Annals of Statistics*, **6**, 461-464.
- Zheng, Zhu, Y., J., Li, D. (2008), Analyzing spatial panel data of cigarette demand: A Bayesian hierarchical modeling approach, *Journal of Data Science*, **6**, 467-489.



Seismic Hazard Analysis of Lifeline Networks Considering Anisotropy of Earthquake Ground-motion Intensity

Alireza Garakaninezhad¹, Morteza Bastami^{2*}

¹Iranian Academic Center for Education, Culture and Research, Kerman, Iran.

²International Institute of Earthquake Engineering and Seismology (IIEES), Tehran, Iran.

Abstract:

Spatial correlation the intra-event residuals of ground motion intensity measures (IMs) has an important effect on seismic hazard analysis of spatially-distributed structures such as lifeline networks. Modeling of correlated domain is strongly affected by the assumption variogram, especially whether or not the isotropic assumption is valid. In the first part of the present study, a nonparametric test is applied to residuals of IMs and then the effect of anisotropic assumption on seismic hazard analysis of lifeline networks is investigated. The anisotropic domain is modeled using two main parameters, namely, anisotropic ratio and anisotropic angle. The results show that anisotropy may increase the estimated intensity measure of ground motion, especially, in rare events.

Keywords: Earthquake hazards, Earthquake ground motions, Spatial analysis, Statistical methods.

Mathematics Subject Classification (2010): 60Gxx, 60Hxx, 60Bxx.

1 Introduction

Quantifying the spatial correlation of ground-motion intensity measures (IMs) is necessary for assessing the seismic risk of lifeline networks (for example; transportation, electricity, gas, telecommunications, water supply) and building portfolios. Ground

*Speaker: m.bastami@iiees.ac.ir

motion prediction equations (GMPEs) estimate the IMs, such as peak ground acceleration (PGA), peak ground velocity (PGV), peak ground displacement (PGD) and spectral accelerations (SAs) for a single site. GMPEs can account for the effects of spatial variation, but to do this, the database and the regression method used are very important.

The partially nonergodic GMPE models developed by [Baker et al. \(2009\)](#) account for spatial variability for distinct regions. [Baker et al. \(2009\)](#) presented a fully nonergodic ground motion model for California records with coefficients that vary continuously on a spatial scale. Most old GMPEs, however, cannot model the spatial correlation of IMs at different sites because of limitations in the regression methods or database used. In other hands, it has been shown that ignoring or overestimating the spatial correlation of IMs can cause the overestimation of frequent losses and underestimation of rare losses ([Bastami, 2007](#); [Baker et al., 2009](#)). Hence, to consider spatial correlation of IMs among sites in a region, it is required to apply models for spatial correlation with these GMPEs.

Several spatial correlation models for IMs have been introduced. [Baker et al. \(2009\)](#) used 1994 Northridge earthquake observations to develop a spatial correlation model for PGA. [Baker et al. \(2009\)](#) computed the spatial correlation of PGV using several earthquakes in Japan and the 1999 Chi-Chi earthquake. [Baker et al. \(2009\)](#) and [Baker et al. \(2009\)](#) developed spatial correlation models based on the 1999 Chi-Chi earthquake and well-recorded earthquakes in California. [Baker et al. \(2009\)](#) used only earthquakes in California. In these studies, the models were proposed using well-recorded individual earthquakes, such as the 1994 Northridge earthquake. In these studies, the correlation range of each earthquake was investigated separately, then a model based on the obtained ranges was proposed.

In other approaches, spatial correlation was investigated by gathering data from a group of earthquakes. [Baker et al. \(2009\)](#) proposed models based on comprehensive databases accumulated in Japan. [Baker et al. \(2009\)](#) and [Baker et al. \(2009\)](#) used the Italian accelerometric archive and the European strong-motion database. [Baker et al. \(2009\)](#) used Vrancea (Romania) intermediate-depth earthquakes. [Baker et al. \(2009\)](#) proposed models for vertical and horizontal components using northern Iran seismic events. [Baker et al. \(2009\)](#) presented a model for vertical component of response spectral accelerations using ten earthquake events. Most of aforementioned studies consider residuals of IMs as an isotropic domain. Recently, [Baker et al. \(2009\)](#) examined the isotropic assumption of spectral acceleration and proposed a model for seismic hazard analysis of spatially-distributed networks considering anisotropy of IMs.

2 Statistical Tests of Isotropy

A domain is isotropic when the dependence between any two observations relies only on the length of the separation distance and not on its direction vector. It is called anisotropy if the relative orientation between two observations affects the semivariogram. Anisotropy can be either geometric or zonal. Geometric anisotropy exists if the ranges of the semivariogram change in different directions but the sill remains the same. Zonal anisotropy occurs when the sill of a semivariogram changes as the direction changes. Anisotropy can be checked using graphical tests such as directional semivariograms and rose diagrams (Cressie 1993).

A directional semivariogram can be obtained by computing semivariogram values from data pairs falling within the prescribed directional bands as well as within a specific bin distance. The azimuth of the direction vector, angular tolerance and bandwidth can be found in Figure 1 (a) and are used to compute the directional semivariogram values. For example, for the region illustrated in Figure 1(b), the semivariogram values at azimuth θ and angular tolerance $\delta \theta$ can be obtained as follows. First, select an arbitrary site A and determine all sites which are in region $[\theta - \delta \theta/2, \theta + \delta \theta/2]$ and categorize them according to the considered bin distances. This procedure should be repeated for all sites. The experimental semivariograms for the considered azimuth can be obtained as described previously. A rose diagram can be obtained by plotting the ranges of the semivariograms in different directions. In the case of geometric anisotropy, the correlation in one direction will be stronger than in the other directions. These ranges in a 2D diagram will fall on the edge of an ellipse. The major and minor axes of the ellipse correspond to the largest and shortest directional semivariogram ranges (Cressie 1993). The problems with graphical methods such as these are that the assessment of the results can be difficult, are subject to personal interpretation and can be misleading because they do not generally contain a measure of uncertainty. Statistical nonparametric test developed using spatial statistics is used in the current study to verify these test results.

2.1 Isotropy of Intra-Event Residuals

The isotropic assumption for each IM is examined using the `sm` package for the R statistical computing environment. The validity of the isotropic assumption is determined using the p-value from the `sm` test. For instance, Figure 2 shows the smoothed variogram of PGA residuals for the Chi-Chi earthquake. The smoothed variogram is a function of distance and angle and the variogram value is color coded. The variogram is calculated by smoothing matrix (S_1) which is obtained based on

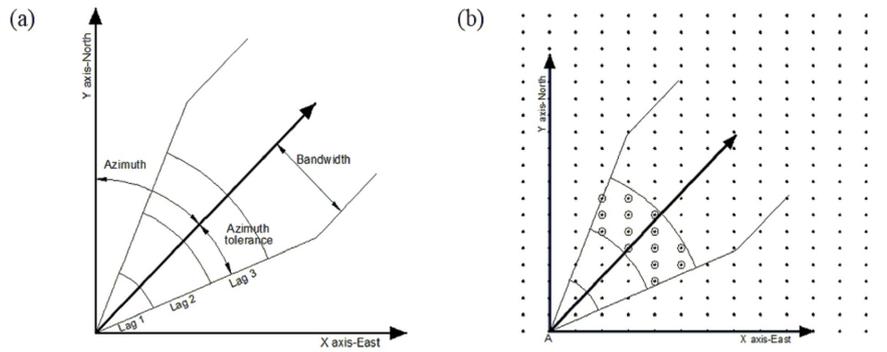


Figure 1: (a) Parameters in directional semivariogram; (b) sampling points of an arbitrary region.

vector h considering distance and angle. The color coded value for each point in this figure indicates variogram value for sites whose distance and orientation determined by the vector between origin and the considered site. For each angle, the variogram is shown by the coded values along the given radius. For instance, the smoothed directional variogram along y axis, which shows north-south direction, is presented as the values in this direction. Although the separated vector has no sign associated with it, it is useful to show the smoothed variogram value over $(0, 2\pi]$ by reflection. The contours show the distance, in unit of standard error, between the variogram obtained as a function of angle and distance and the variogram as a function of only distance. In this example, $p = 0.0125$, which implies that the isotropic assumption for PGA residuals of the Chi-Chi event is not valid. The slower increasing rate near the origin shows the direction of anisotropy, which, in this example, the anisotropy direction is aligned at north-south direction. Figure 3 shows the smoothed variograms for the other IMs. In an anisotropic domain, the anisotropy ratio and angle are required to determine the spatial correlation between sites. These parameters are generally calculated using directional semivariograms estimated for four azimuths (0° , 45° , 90° , and 135°). In this study, the directional semivariograms were estimated by considering a bin separation distance of 4-6 km, an angle tolerance of 22.5° and a bandwidth of 10 km.

3 Seismic Hazard Analysis Considering Anisotropy

The following example illustrates the importance of anisotropy in seismic hazard analysis. Consider a region 30×30 km in size which is divided into 900 square cells. The V_{s30} values follow a normal distribution with a mean value of 380 m/s and a standard deviation of 120 m/s. Assume a fault with reverse mechanism is located in

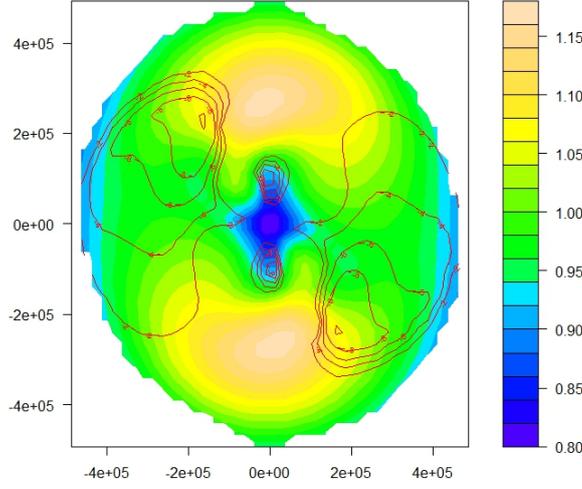


Figure 2: Smoothed variogram for PGA residuals for Chi-Chi earthquake obtained from sm package.

N-S direction near to the region. In this study, we generate an earthquake scenario using the empirical relationship between source dimensions and magnitude proposed by Wells and Coppersmith (1994). Assuming surface rupture length of 40 km, a magnitude 7 earthquake scenario with epicenter near the origin of the region and corresponding annual rate of exceedance of $\lambda = 1/500$ is considered.

The median predicted PGA at each site location is calculated as suggested by Campbell and Bozorgnia (2014). The model proposed by ? is used to obtain an anisotropy ratio and is equal to 2.84 for the PGA. The annual rate of exceedance of PGA is estimated applying the Monte Carlo simulation (MCS) to generate 10000 realizations of anisotropic spatially-correlated PGA values over the region. Given a specified value of PGA (PGA^*) and its exceedance area ratio (AR^*) which is defined as the ratio of the area in which PGA values exceed the considered PGA^* value against the total area of the region, the annual rate of exceedance can be estimated as

$$\lambda = \lambda_m \cdot P(PGA > PGA^* \text{ and } AR > AR^*) \quad (3.1)$$

In this example, the exceedance area ratio is assumed equal to 5. The annual rate of exceedance is computed by MCS for two cases in which the isotropic assumption is valid or is not valid. For purposes of comparison, this procedure is carried out for three ranges of semivariogram at 5, 10 and 20 km. Figure 3 shows two Monte

Carlo realizations of PGA residuals for tow ranges of 5 km and 20 km with and without isotropy assumption. In anisotropy events Figures 3(b) and 3(d) show a more uniform spatial distribution of PGA residuals in the assumed anisotropy direction (N-S). Figure 4 shows the annual exceedance curve for PGA and indicates that the effect of anisotropy for different correlation ranges. It is clear that the hazard level for cases with a lower correlation range will be underestimated because anisotropy has not been considered.

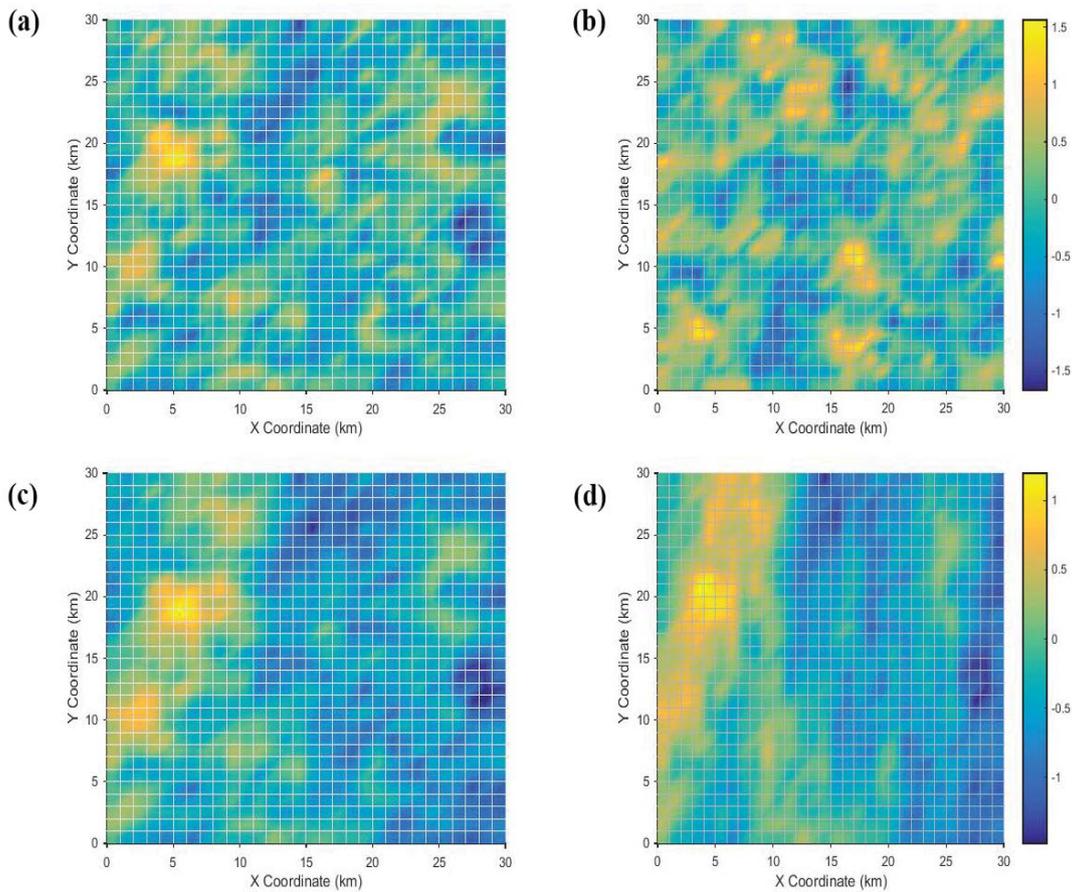


Figure 3: Two realizations of PGA residuals (a) and (b) correlation range of 5 km, with and without isotropy assumption, respectively and (c) and (d) correlation range of 20 km, with and without isotropy assumption, respectively.

4 Conclusion

Seismic hazard analysis of lifeline networks is influenced by spatial correlation among earthquake ground-motion intensity measures. The structure of the residual variogram has strongly effect on correlated domain. However, it is assumed that the

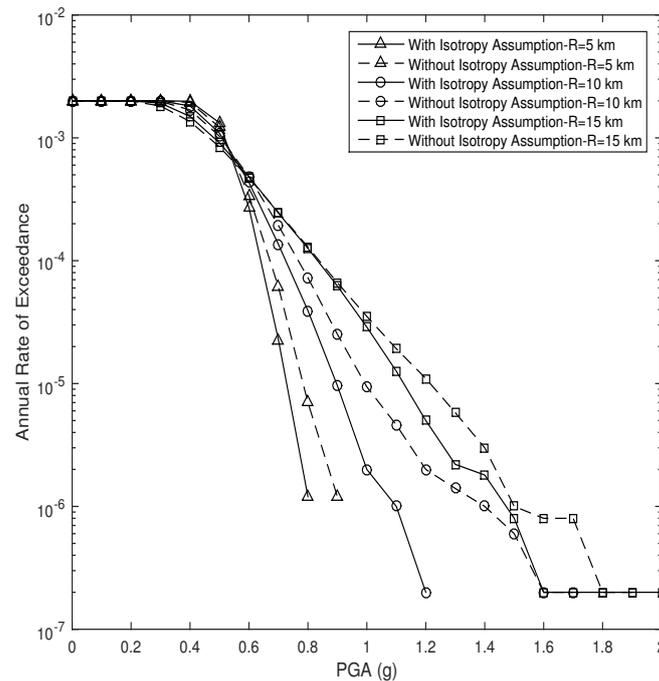


Figure 4: Annual rate of exceedance for PGA for different correlation ranges as 5 km, 10 km and 20 km.

residuals are isotropic and the variogram is independent of directions. In the present study, a nonparametric test is applied to investigate the isotropic assumption. The results showed that this assumption for some IMs is not valid. In addition, the effect of anisotropy was investigated on seismic hazard analysis of lifeline networks. The results depicted that ignoring anisotropy may lead to increase IMs, especially for rare events.

References

- Bastami, M. (2007), Seismic Reliability of Power Supply System Based on Probabilistic Approach, PhD thesis, Kobe university, 2007.
- Bazzurro, P. Luco, N. (2007), Effects of Different Sources of Uncertainty and Correlation on Earthquake-Generated Losses, *Australian Journal of Civil Engineering*, 4, 1-14.
- Boore, D., Gibbs, J., Joyner, W., Tinsley, J. and Ponti, D. (2003), Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of

the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California, *Bulletin of the Seismological Society of America*, **93**, 2737-2751.

Boore, D., Gibbs, J., Joyner, W., Tinsley, J. and Ponti, D. (2003), Estimated Ground Motion From the 1994 Northridge, California, Earthquake at the Site of the Interstate 10 and La Cienega Boulevard Bridge Collapse, West Los Angeles, California, *Bulletin of the Seismological Society of America*, **93**, 2737-2751.

Esposito, S. Iervolino, I. (2011), PGA and PGV Spatial Correlation Models Based on European Multievent Datasets, *Bulletin of the Seismological Society of America*, **101**, 2532-2541.

Esposito, S. Iervolino, I. (2011), Spatial Correlation of Spectral Acceleration in European Data, *Bulletin of the Seismological Society of America*, **102**, 2781-2788.

Garakaninezhad, A. Bastami, M. (2017), A Novel Spatial Correlation Model Based on Anisotropy of Earthquake Ground-Motion Intensity, *Bulletin of the Seismological Society of America*, **107**, 2809-2820.

Garakaninezhad, A. Bastami, M. (2019), Intra-event Spatial Correlation Model for the Vertical Component of Response Spectral Accelerations, *Journal of Seismology*, **107**, 1-15.

Garakaninezhad, A. Bastami, M. (2017), Spatial Correlation for Horizontal and Vertical Components of Acceleration from Northern Iran Seismic Events, *Journal of Seismology*, **21**, 1505-1516.

Gianniotis, N., Kuehn, N. and Scherbaum, F. (2014), Manifold Aligned Ground Motion Prediction Equations for Regional datasets, *Computers and Geosciences*, **69**, 72-77.

Goda, K. and Hong, H. (2008), Spatial Correlation of Peak Ground Motions and Response Spectra, *Bulletin of the Seismological Society of America*, **98**, 354-365.

Goda, K. and Atkinson, G. (2010), Intraevent Spatial Correlation of ground-Motion Parameters using SK-net Data, *Bulletin of the Seismological Society of America*, **100**, 3055-3067.

Hong, H., Zhang, Y. and Goda, K. (2009), Effect of Spatial Correlation on Estimated Ground-Motion Prediction Equations, *Bulletin of the Seismological Society of America*, **99**, 928-934.

Jayarama and Baker (2009), Correlation Model for spatially Distributed Ground-Motion Intensities, *Earthquake Engineering and Structural Dynamics*, **38**, 1687-1708.

Kotha, S., Bindi, D. and Cotton, F. (2016), Partially Non-Ergodic Region Specific GMPE for Europe and Middle-East, *Bulletin of Earthquake Engineering*, **14**, 1245-1263.

Landwehr, N., Kuehn, N., Scheffer, T. and Abrahamson, N. (2016), A Nonergodic GroundMotion Model for California with Spatially Varying Coefficients, *Bulletin of the Seismological Society of America*, **106**, 2574-2583.

Lee and Kiremidjian (2007), Uncertainty and Correlation for Loss Assessment of Spatially Distributed Systems, *Earthquake Spectra*, **23**, 753-770.

Pavel, F. and Vacareanu, R. (2016), Spatial Correlation of Ground Motions from Vrancea (Romania) Intermediate-Depth Earthquakes, *Bulletin of the Seismological Society of America*, **107**, 489-494.

Sedaghati, F. and Pezeshk, S. (2017), Partially Nonergodic Empirical Ground-Motion Models for Predicting Horizontal and Vertical PGV, PGA, and 5 Damped Linear Acceleration Response Spectra Using Data from the Iranian Plateau, *Bulletin of the Seismological Society of America*, **107**, 934-948.

Stafford, P. (2014), Crossed and Nested Mixed-Effects Approaches for Enhanced Model Development and Removal of the Ergodic Assumption in Empirical Ground-Motion Models, *Bulletin of the Seismological Society of America*, **104**, 702-719.

Wang, M. and Takada, T. (2005), Macrospatial Correlation Model of Seismic Ground Motions, *Earthquake spectra*, **31**, 1137-1156.



Clustering Hyperspectral Images Using Spatial Dependencies

Mousa Golalizadeh*, Hamideh Sadat Fatemi Ghomi
Department of Statistics, Tarbiat Modares University, Tehran, Iran.

Abstract:

Images, as the source of data, can be analyzed using some tools from spatial analysis. The key to do this task is Hidden Markov Random Field (HMRF). In this paper, we combine the HMRF with Latent Block Model (LBM) to both summarize large array of data sets in the images and to cluster the hyperspectral images. We show how spatial spectrum information can be invoked to shorten statistical inference on high dimensional images without confining to estimate a large numbers of parameters. This is provided by LBM framework on finding homogeneous blocks and then segmentation map among images. The outcome of such segmentation map is used in the spectral-spatial classification stage. To reduce data volume, a feature selection based on either Kullback-Libler divergence or a principal component transformation are also employed. We illustrate the application of our proposed model on real-life data and show its superiority in compare with standard multivariate tools.

Keywords: Spatial analysis, Image segmentation, Hidden Markov Random Field, Latent block model, Clustering.

Mathematics Subject Classification (2010): 62H86, 62H30.

1 Introduction

The topic of high dimensional data is frequently used among researchers. A common example in this context is the images. One is statistically interested in either cluster or classify images based on various features or spectral properties. There are

*Speaker: golalizadeh@modares.ac.ir

numerous examples in which these two multivariate techniques have been invoked in analyzing hyperspectral images. To name some, we can mention urban planning, precision agriculture, and environmental monitoring. See, e.g. [Aykroyd \(2015\)](#) for more examples.

To cluster the images, one can simply consider the spectral feature while taking the spatial information into account in the learning process. To improve the ultimate clustering accuracies in various applications, [Ghamisi, et al. \(2014\)](#), [Sun, et al. \(2015\)](#), [Golipour, et al. \(2016\)](#) and [Wang, et al. \(2017\)](#), among others, proposed different tricks. One of the simple trick to take the spatial information into account is to perform image segmentation. Recently, [Tarabalka, et al. \(2009\)](#) proposed a method to utilize this along with spectral information by combing together the clustering and classification steps. Note that the clustering technique is applied to the relevant features to find some segmentation on the image. This trick seeks an adaptive spatial partition of pixels and helps to smoothly classify images by, simultaneously, integrating the spectral and spatial information.

It is known, see for example [Cressie \(1993\)](#), that one of the key concept in spatial statistics is the Markov Random Field (MRF). One can utilize it through exploiting spatial information inherited in an image via considering a neighborhood dependency among pixels. It is known that an unobservable random variable, possessing Markovian property, is usually modeled by a Hidden Markov Random Field (HMRF). This technique considers a particular structure of the Hidden Markov Model (HMM); the underlying stochastic process is a MRF rather than a Markov chain. So, the HMRFs can be applied in two-dimensional spatial problems such as image processing. Recently [Ghamisi, et al. \(2014\)](#) used an HMRF model for segmentation of hyperspectral images based on spatial features. In particular, they applied a segmentation method to get non-overlapping homogeneous regions aiming on classification of hyperspectral images. This segmentation map is achieved by a clustering method based on EM algorithm on the first principal component of spectral. But, this method only works on a single spectrum image.

We utilize the Latent Block Model (LBM), which is a powerful tool to summarize a vast array of data sets. In particular, we are going to, properly, combine LBM with HMRF to benefit the capabilities of both models. This merged tool is called LBMHMRF throughout this paper. The purpose of this method is to seek homogeneous blocks in image and to include more information given by neighborhood dependencies among pixels. It is expected that one can then utilize more spectral information in analyzing high dimensional images. The main purpose of this new model is twofold. To ease the complexity of statistical inference via reduce the

number of parameters to estimate and to shorten the computational cost through terminating some unnecessary tasks.

The remainder of this paper is organized as follows. An overview of clustering hyperspectral images based on the HMRF combined with the LB model is provide in section 2. An application of the proposed method in a real-life example is reported in section 3. The paper ends with some general conclusions.

2 Spatial Clustering of Hyperspectral Images

Let us assume $Y = [y_{ij}]$ for $i = 1, \dots, n$, $j = 1, \dots, N$ represents a matrix, where N is the number of pixels and n is the number of input bands. This matrix is essentially constructed using, say I , bands with n elements and spectrums of, say J , pixels with N elements. It worths to mention that these typical notations are common in image analysis. See, e.g. [Jensen and Lulla \(1987\)](#) for more details. Then, clustering images using the LB model corresponds to partitioning Y into block matrices, say Z and X , respectively constructed with the inputs being the sets of all possible row label z of I into g clusters and column label x of J into L clusters. More details are given below.

The most common procedure to cluster the objects is the model-based clustering ([Vapnik, 1998](#)). To accommodate it into our problem, let us assume that i -th row (input band) of data matrix, say y_i for $i = 1, \dots, n$, is an *i.i.d* sample from a probability distribution function (*pdf*) with the density

$$f(y_i; \theta) = \sum_{l \in L} \pi(l) f_l(y_i; \theta_l) \quad (2.1)$$

where, for $l \in L$, $\pi(l)$, is the probability of those pixels being in l -th cluster, $f_l(\cdot)$ is *pdf* for classes with the label l , and L is the set of all possible labels. Following [Govaert and Nadif \(2003\)](#), *pdf* for y_i given θ , say $f(y_i; \theta)$, can then be written as $f(y_i; \theta) = \sum_{x \in X} w(x) f(y_i|x; \theta)$, where $w(x)$ is the probability of the observation having the label x .

Following [Lomet, et al. \(2012\)](#), the LB is simply a generalization of the mixture models. Particularly, a variable modelled by the LB can be considered as a mixture of block components including the latent row and column classes. Hence, we can write

$$f(y_i; \theta) = \sum_{(z,x) \in Z \times X} p(z, x|\pi) f(y_i|z, x; \theta).$$

Note that, the row labels are assumed to be *i.i.d*, i.e. $p(z) = \prod_i p(z_i) = \prod_{i,k} \pi_k^{z_{ik}}$. Such

assumption also holds for column labels if one intends to follow the LBM.

Let us consider the common mixture model for y , i.e. $f(y; \theta) = \sum_{u \in U} g(u) f(y|u, \theta)$, where U represents possible simultaneous partitions on the set of rows and columns, say the set $I \times J$. If assigning the labels for the row is independent of that for column, then we can write $g(u) = \pi(z)\pi(x)$; so

$$f(y; \theta) = \sum_{(z,x) \in Z \times X} \pi(z)\pi(x) f(y|z, x; \theta). \quad (2.2)$$

where $f(y|z, x; \theta) = \prod_{i,j,k,l} f(y_{ij}; \alpha_{kl})^{z_{ik}x_{jl}}$, with α_{kl} s indicating the relevant parameters in the model. Note that we used the principle of local independency to write conditional *pdf* in (2.2).

Here we can augment our spatial information into modeling rows and columns. For simplicity, let us assume X is a MRF. Then, the Gibbs distribution for four-neighborhoods system can be expressed as

$$\begin{aligned} P(x) &= \frac{1}{A} e^{-U(x)} = \frac{1}{A} e^{-\sum_{c \in C} V_c(x)} = \frac{1}{A} \exp\left(-\sum_{\{i,j\} \in C} V_c(x_i, x_j)\right) = \frac{1}{A} \exp\left(-\sum_{j \in S} \sum_{i \in N_j} V_c(x_i, x_j)\right) \\ &= \frac{1}{A} \prod_{j=1}^N \exp\left(-\sum_{i \in N_j} V(x_i, x_j)\right) = \frac{1}{A} \prod_{j=1}^N \prod_{l=1}^L \left(\exp\left(-\sum_{i \in N_j} V(x_i, l)\right)\right)^{x_{jl}} = \prod_{j,l} \frac{1}{A'} P(l|x_{N_j})^{x_{jl}}, \end{aligned}$$

where $\rho_{jl} = P(l|x_{N_j})$ is the conditional probability of class l for pixel j , with known classes of her neighborhoods and A' is the normalizing constant for the function appeared in the corresponding expression. If the neighborhoods systems is different from that considered above, i.e. four-neighborhoods, one then can apply corresponding approximation of the Gibbs distribution, proposed by [Besag \(1986\)](#).

To follow the same procedure for Z and then plug the components into (2.2), leads to

$$f(y; \theta) = \sum_{(z,x) \in Z \times X} \prod_{i,k} \pi_k^{z_{ik}} \frac{1}{A'} \prod_{j,l} \rho_{jl}^{x_{jl}} \prod_{i,j,k,l} f(y_{ij}; \alpha_{kl})^{z_{ik}x_{jl}}. \quad (2.3)$$

Now assume one is going to make statistical inference on the parameters based on the *pdf* given in (2.3). Apart from common maximum likelihood procedure, she should recall the EM algorithm here because of encountering with the latent labels in this model. However, to employ this algorithm is not tractable for our model due to typical spatial dependency among pixels. To circumvent this problem, [Govaert and Nadif \(2003\)](#) proposed a variational procedure to estimate the parameters and to predict the latent variables appeared in the LB model. We provide a criterion

based on this variational procedure to make the implementation and interpretation of the EM algorithm straightforward.

The variational procedure provides a simple trick to make statistical inference on the parameters. Precisely, rather than directly maximizing the likelihood, one should maximize

$$G(c, t, \theta) = L_C(c, t, \theta) + H(P_z) + H(P_x) = L_C(c, t, \theta) + H(c) + H(t) \quad (2.4)$$

where $c = (c_{ik})$, $t = (t_{jl})$ such that $c_{ik} = P_z(z_{ik} = 1|y, \theta)$, and, similarly, $t_{jl} = P_x(x_{jl} = 1|y, \theta)$ and H is the entropy function. Moreover,

$$L_C(c, t, \theta) = \sum_{i,k} c_{ik} \log(\pi_k) + \sum_{j,l} t_{jl} \log(\rho_{jl}) + \sum_{i,j,k,l} c_{ik} t_{jl} \log f_{kl}(y_{ij}; \alpha_{kl}) - \log(A').$$

The optimization of criterion (2.4) can be done with two alternated steps defined as follows:

- I) Implement a HMRF model on columns. We call it **HMRF-EM** algorithm.
- II) Fit a mixture model on rows. We call it **FM-EM** algorithm.

More details on this and other relevant topics are discussed in [Fatemi, et al. \(2018\)](#)

Note that to achieve the LB model, these two algorithm should be called in turn. Utilizing each algorithm, parameters of blocks (α_{kl}) are updated and the observations are clustered as blocks in both rows and columns. Meanwhile, three necessary steps should be repeated to estimate parameters of the LBMHMRF model. They are as follows:

Step I) Find class label of pixels through MAP estimation derivation.

Step II) Compute the conditional probability $t_{jl}^{(c+1)}$ with $\theta^{(c)}$ being fixed (**E-step**).

Step III) Derive the maximum likelihood estimates of the parameters for each block (**M-step**).

One should note that the **Step II** induces the insertion of the spatial information into building the LBM and so will be effective in clustering the pixels and then constructing the segmentation map.

3 Real Application

One of the popular examples to analysis in the area of image analysis is the Pavia University data set. We are going to implement models proposed in this paper as well as other traditional methods on this example. Following [Ghamisi, et al. \(2014\)](#),

to invoke the HMRF model, we applied it only on the first principle components. Also, we used SVM, as nonlinear classifier, in the classification map stage. To have more comparison option, a multivariate distribution (MultiHMRF) for each class is also considered.

Since there are too many features in the images, we need to first reduce them on the bands using standard statistical methods. As pointed out by [Martinez, et al. \(2007\)](#), we utilize a method based on a hierarchical clustering technique to collect bands together to minimize and maximize the inter-cluster and intra-cluster variances, respectively. To do this, we can consider the criteria based on information measures, such as distances based on mutual information or Kullback-Leibler divergence. Here, we use a symmetric version of later criterion as a dissimilarity measure. Moreover, to select the most effective informative features, we utilize the transformation arising from principal component analysis. Also, we apply methods described in this paper using %5 of labeled data for each class as training samples. Specific sample sizes for both training and test data for each scenario are reported in table 1.

Note that the accuracy measures used in this analysis refer to correct assigning of classes in the clustering procedure. Hence, Overall Accuracy (OA) shows the percentage of accurate clustering throughout the image. Average Accuracy (AA) is defined accordingly and Kappa is well-known clustering criterion. Whenever we combined the SVM with other methods, we add a plus sign between them to show this.

As expected, the more PCs the more improvement is seen in applying any method to analysis the data. However, this might not be the case for our example due to spatial dependency among bands. We are not going in more details on this challenge here. Instead, our aim is to compare the performance of various models applied on our current example. However, we observed (results not shown here) that there is not a monotone change on accuracies when the numbers of PCs are increasing. As an example, SVM+HMRF had no more improvement in the accuracy of classification after adding PCs.

We used Kullback-Leibler divergence to select appropriate bands. Moreover, the numbers of bands to evaluate the performance of the model were set on 5, 7 and 10. We then derived the relevant accuracy measures for each scenario. The results for all combinations are reported in table 1. As seen, there is significant improvement on accuracies when the bands are increased. In particular, moving from five to seven bands turned to gain more information used in modeling and then led to achieve higher accuracy. It indicates that inclusion of spatial information in the model can

be very effective with no extra cost on computational time.

Table 1: Classification accuracies, in percent, gained using the selected bands with specific number of training and test samples while invoking different models.

Class	Sample		SVM	SVM+ HMRF	#Bands=5		#Bands=7		#Bands=10	
	#Train	#Test			SVM+	SVM+	SVM+	SVM+	SVM+	SVM+
					LBMHMRF	MultiHMRF	LBMHMRF	MultiHMRF	LBMHMRF	MultiHMRF
Asphalt	331	6300	93.55	92.58	91.07	98.09	97.59	98.72	96.50	96.04
Meadows	932	17717	97.99	97.57	99.78	99.81	99.84	99.13	99.50	99.71
Gravel	104	1995	76.73	79.91	78.47	86.71	77.56	88.33	79.98	82.61
Trees	153	2911	93.84	78.02	93.9	91.42	87.64	91.92	79.87	91.61
Meta sheets	67	1278	99.31	86.60	90.59	99.77	90.55	99.86	79.55	99.92
Bare soil	251	4778	83.51	94.07	95.18	92.68	95.79	93.2	92.93	94.48
Bitumen	66	1264	82.98	92.97	97.40	99.66	96.70	97.57	98.99	98.21
Bricks	184	3498	88.63	97.70	95.95	88.20	97.89	96.92	97.19	96.67
Shadows	47	900	99.86	99.88	100	99.79	99.93	99.79	99.88	98.71
OA	-	-	93.07	93.69	95.73	96.45	96.49	97.12	95.07	96.78
AA	-	-	90.71	91.03	93.59	95.12	93.72	96.16	91.60	95.33
Kappa	-	-	90.75	91.60	94.30	95.27	95.32	96.17	93.40	95.72
Time(Sec)	-	-	-	-	149.6674	1090.7	187.0494	1436.1	148.4568	1426.7

It might be of interest to evaluate the performance of the proposed model in retrieving the original (real) map of the example, i.e. Pavia university map. We have done this using the LBMHMRF model. Figure 1 is a visual representation of classification maps for our example using two PCs. As seen, our model did manage to provide somehow more clear view of the map than its noisy version.

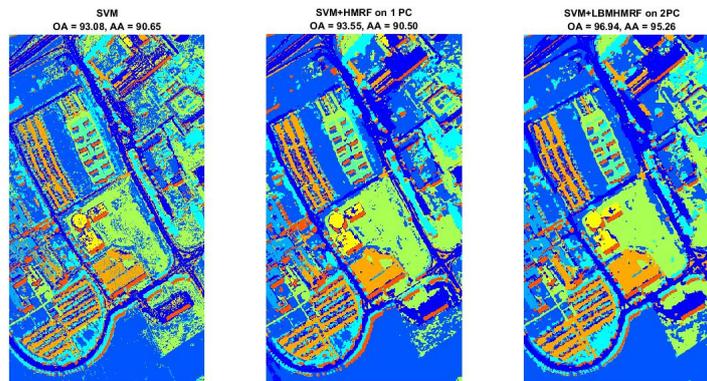


Figure 1: Sample classification maps for paviaU data set.

Conclusion

We proposed a modified version of spectral-spatial clustering and classification for the hyperspectral images. The idea comes from combining the LB with HMRF models to enable us for considering the spatial information inherited in the images. It was shown that, unlike traditional methods, the combined model helps one to

retrieve the image with small numbers of bands. Moreover, reducing the numbers of parameters to estimate and then saving computational time is noticeable while invoking the LBMHMRF.

The implementation of the LBMHMRF has been done with $1 \times L$ blocks. The accuracy might be increased if one use $g \times L$ blocks where $g \leq n$. This can be considered as a possibility for future work.

Acknowledgement

The authors acknowledge constructive comments suggested by Dr. Mansoor Rezghi.

References

- Aykroyd, R. G. (2009), *Industrial Tomography: Systems and Applications*, Chapter Statistical Image Reconstruction, Woodhead Publications, New York.
- Besag, J. (1986), On the Statistical Analysis of Dirty Pictures, *Journal of the Royal Statistical Society, Series B (Methodological)*, **48**, 259302.
- Cressie, N. (1993), *Statistics for Spatial Data*, John Wiley, New York.
- Fatemi, H.S., Golalizadeh, M., and Rezghi, M. (2018), Spectral-spatial Classification of Hyperspectral Images Using Latent Block Model Based on Hidden Markov Random fields, Submitted to *Advances in Data Analysis and Classification*.
- Ghamisi, P., Benediktsson, J. A., and Ulfarsson, M. O. (2014), Spectral-Spatial Classification of Hyperspectral Images Based on Hidden Markov Random Fields, *IEEE Transaction on Geoscience and Remote Sensing*, **52**, 25652574.
- Golipour, M., Ghassemian, H., and Mirzapour, F. (2016), Integrating Hierarchical Segmentation Maps with MRF Prior for Classification of Hyperspectral Images in a Bayesian Framework, *IEEE Transactions on Geoscience and Remote Sensing*, **54**, 805816.
- Govaert, G. and Nadif, M. (2003), Clustering with Block Mixture Models, *Pattern Recognition*, **36**, 463 473.
- Jensen, J. R. and Lulla, K. (1987), *Introductory Digital Image Processing: a Remote Sensing Perspective*, Prentice-Hall, New Jersey.

Lomet, A., Govaert, G., and Grandvalet, Y. (2012), *Design of Artificial Data Tables for Co-clustering Analysis*, Universit de Technologie de Compigne, France.

Martinez, A. U., Pla, F., Sotoca, J. M., and García-Sevilla, P. (2007), Clustering-based Hyperspectral Band Selection Using Information Measures, *IEEE Transactions on Geoscience and Remote Sensing*, **45**, 41584171.

Sun, L., Wu, Z., Liu, J., Xiao, L., and Wei, Z. (2015), Supervised spectralspatial hyperspectral image classification with weighted markov random fields. *IEEE Transactions on Geoscience and Remote Sensing*, **53**, 14901503.

Tarabalka, Y., Benediktsson, J. A., and Chanussot, J. (2009), SpectralSpatial Classification of Hyperspectral Imagery Based on Partitional Clustering Techniques, *IEEE Transactions on Geoscience and Remote Sensing*, **47**, 29732987.

Vapnik, V. (1998), *Statistical Learning Theory*, John Wiley, New York.

Wang, L., Zhang, J., Liu, P., Choo, K.-K. R., and Huang, F. (2017), SpectralSpatial Multi-Feature-Based Deep Learning for Hyperspectral Remote Sensing Image Classification, *Soft Computing*, **21**, 213221.



Utilization of Pearson Correlation to Grab Spatial Correlation in the Absence of Information

Lida Kalhori Nadrabadi*

Statistical Research and Training Center, Tehran, Iran.

Abstract:

In many practical situations, spatial correlation is intrinsically existed and should be intervened in the analysis. Spatial models have been widely used in statistical analysis of epidemiology and health issues when the spatial correlation exist and basic assumptions are satisfied. However, in official statistics issues there is no attention to this crucial aspect. There are different areas that spatial issue should be mentioned in existence of spatial correlation, from sampling scheme to parameter estimation, and neglecting this important aspect leads to misleading results. In practice we do encounter some situations that there exists a sample frame without any information about spatially correlated interested variables. In this paper we are going to review concept of Moran I correlation and investigate that how an existing variable in the sampling frame, which is correlated to the interest variable, can be used to grab the desired spatial structure. Intensive simulation studies with different values of Pearson correlation and distinct sample sizes were done to evaluate the accuracy of the proposed approach as well. Parameter estimation is done using the MLE method. As an application, the data available from Household Income and Expenditure Survey in Tehran city is considered.

Keywords: Spatial correlation, Pearson correlation, Sampling Frame, Household Income and Expenditure Survey.

Mathematics Subject Classification (2010): 91B72, 62DXX.

*Speaker: lidakalhari@yahoo.com

1 Introduction

Spatial models have been widely used in statistical analysis of epidemiology and health issues when the spatial correlation exist and basic assumptions are satisfied. However, in official statistics issues there is no attention to this crucial aspect. There are different areas that spatial issue should be mentioned in existence of spatial correlation, from sampling scheme to parameter estimation. One of the most crucial aspects of sample designs is sampling frame that has significant effects on cost and quality of any survey. It is worth to note that faulty sampling frames are a common source of nonsampling error, particularly under-coverage of important population sub-groups. Ideally, of course, a frame should be current in order for it to fulfil the other two properties of completeness and accuracy. An obsolete frame obviously contains inaccuracies and is likely to be incomplete. Especially in household surveys, the quite important deficiency of a frame is being out of date since it is based on the population census that is several years old. The old census will not accurately reflect new construction or demolition of dwellings, in- or out-migrants in dwelling units, births or deaths. These deficiencies violate the criterion of a probability sample that each member of the target population must have a known chance of selection (Turner, 2003). Another aspect which is not mentioned in construction of sampling frame is geographical information of sampling units, which would be used to find out spatial correlation of sampling units. Spatial correlation provides us a useful tool to capture the existence of global heterogeneity. Spatial autocorrelation is a measure of spatial dependence between values of random variables over geographic locations. So if the geographic information of dwellings is included into the frame, it can be used to assess the spatial heterogeneity and interfered it while selecting the sample. In result, the sample will be distributed in a way to cover dwellings with different characteristics. To a certain extent, samples that are close to each other are more likely to be similar (Dale, 1999) . There are different statistics to evaluate spatial correlation, the most often used and cited one is Morans I (1948). Moran I is a single test statistic that indicates two types of spatial autocorrelation, positive autocorrelation and negative autocorrelation. A positive autocorrelation captures the existence of both high-value clustering and low-value clustering, while a negative autocorrelation captures the existence of high-values next to low-values (Anselin et al., (2000); Haining (1990); Lawson and Denison (2000)). Imagine a situation that we aim to gather information about a variable which is spatially correlated, while the available sampling frame contains no information about the spatially correlated variable of interest. How this deficiency can be encountered if there is no way to correct the frame, for instance it is based on the last census. In this paper we

will aim to show that a correlated variable from the frame can be used to grab the spatial structure of the variable of interest. In Section two, we will talk about the methodology and simulation scheme. Section 3 is devoted to present results of the simulation study. In Section 4, application of the proposed approach on a real data set of HIES is proposed, and the paper is end up by discussions and conclusions.

2 Methodology

Spatial autocorrelation is a measure of spatial dependence between values of a random variable over geographic locations. Morans I is one of the oldest statistics used to examine spatial autocorrelation. This global statistic was first proposed by Moran (1948) and Moran (1950). Later, Cliff (1973) and Cliff and Ord (1981) present a comprehensive work on spatial autocorrelation and suggested a formula to calculate the I as bellow, which is now used in most textbooks and soft wares

$$I = \frac{n}{W} \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2}$$

where n is number of observations, W is the sum of the weights w_{ij} for all pairs in the system, The weight matrix can be specified in many ways, for instance the weight for any two different locations is a constant, all observations within a specified distance have a fixed weight, K nearest neighbours have a fixed weight, and all others are zero, and weight is proportional to inverse distance, inverse distance squared, or inverse distance up to a specified distance. $z_i = x_i - \bar{x}$ where x is the value of the variable at location i, and \bar{x} is the mean value of the variable of interest. Computation of Morans I is achieved by division of the spatial covariation by the total variation. Resulted values are in the range from approximately -1 to 1. Positive sign represents positive spatial autocorrelation, while the converse is true for negative sign, and there is no spatial autocorrelation when Moran I equals to zero. When no statistically significant spatial autocorrelation exists, the pattern of spatial distribution is considered random (Chou , 1997). On the other hand, Pearsons correlation coefficient is another measure of correlation between two variables, but nothing is reflected about interactions based on spatial connection. We are aim to understand whether the measure of Pearson correlation between two variables has any clue about the similarity of spatial correlation structure of the two or not. Assume that based on prior knowledge we know that one variable is following a spatial structure, but we do not have access to the current values of this desired variable. However, we have access to the information about another variable which is correlated to the

variable of interest due to the Pearson correlation coefficient. Can we use the information from the available correlated variable to find out the spatial structure of the other variable? In order to find a solution, we go through a simulation study with different measures of Pearson correlation and different lattice size to check their effects as well. The spatial correlation structure of the response variable is incorporated into the model by adding a random component to the mean model. Suppose $\mathbf{y} = (y(s_1), \dots, y(s_n))^T$ are realizations of random variables $\mathbf{Y} = (Y(s_1), \dots, Y(s_n))^T$ at n distinct locations s_1, \dots, s_n . For simplicity assume y_i and \mathbf{Y}_i denote the $y(s_i)$ and $\mathbf{Y}(s_i)$ respectively, in the following parts. Let $y = (y_1, y_2, \dots, y_n)$ are Gaussian distributed response such that $\mathbf{y} \mid \boldsymbol{\beta}, \sigma^2, \tau \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Also suppose that the random vector $\tau(s) = (\tau(s_1), \dots, \tau(s_n))$ denote a Gaussian Random Field (GRF). For the sake of brevity we will show $\tau(s_i)$ by τ_i , $i = 1, \dots, n$ in the following parts. Therefore, we define the spatial regression model with random effect as bellow

$$\text{Model 1:} \quad \mu_i = \beta x_i + \tau_i, \quad i = 1, \dots, n,$$

where $x_i = (x_{i0}, \dots, x_{i(p-1)})$ and $\beta = (\beta_0, \beta_1, \dots, \beta_{(p-1)})^T$ are vectors of non-stochastic regressors and regression parameters respectively. To have both positive and negative regression coefficients, the considered values for the model parameters are $\beta = (\beta_0, \beta_1, \beta_2) = (1, -2, 3)$ and the covariates are simulated from the uniform distribution $U(0, 1)$. Let the spatial effects τ_i be realizations of a zero mean Gaussian random field with covariance matrix $\boldsymbol{\Sigma}_\tau$, and components of the matrix define as a function of the distance between geographical sites (Wang and Wall, 2003). Assume s_i denotes the geographical coordination of the i th sample unit. There are different possibilities for the covariance matrix, assume that $\sigma_{kl}^2 = \gamma \exp(-|s_k - s_\ell|/\phi)$, which is the exponential covariogram, is the spatial variance, ϕ is the range and $|s_k - s_\ell|$ is the distance between geographical coordinates of sample units k and ℓ . The likelihood function for Model 1 is as follow

$$L(\beta, \phi, \gamma \mid \mathbf{y}) = \int \prod_{i=1}^n f(y_i \mid x, \beta, \phi, \gamma) f(x \mid \phi, \gamma) dx$$

For generating values of the random effect τ_i , as noted before τ_i s are from a zero mean Gaussian random field, that is $\tau = (\tau_1, \dots, \tau_n) \sim N(0, \boldsymbol{\Sigma}_\tau)$ and the ij th component of $\boldsymbol{\Sigma}_\tau$ is $(\boldsymbol{\Sigma}_\tau)_{ij} = \gamma \exp(-|s_i - s_j|/\phi)$. Required parameters are set as $\gamma=0.5$, $\phi=5$. Data are generated on coordinates of 10×10 , 20×20 and 30×30 squares lattice. After generating the $y = (y_1, y_2, \dots, y_n)$ for each value of $n=100, 400, 900$ respectively, we imply the Cholesky decomposition to generate another set of data which in correlated to the first generated data set. For each sample size different values of

Pearson correlation are considered in order to evaluate the effect of correlation value as well. Results of simulation study are presented in the next section.

3 Simulation Study

As shown in Model 1, mean of the response variable arise from two distinct elements. The First one is a fixed structure and the second element is a random component, which reflecting the stochastic spatial process. Therefore, in the process of simulation study, estimation of spatial parameters was done after detrending the response variable. In the first step a regression model is fitted to the generated response variable to obtain the residuals. In the second step, the exponential variogram is fitted on the residuals to estimate the spatial parameters. In our simulation study we generate N=100 data sets from the proposed structure for n=100,400, 900. We computed the relative bias (RelBias) and the square root of MSE (RMSE) for each parameter over the 100 simulated samples. They are defined as

$$\text{RelBias}(\theta) = \frac{1}{100} \sum_{i=1}^{100} \left(\frac{\hat{\theta}_i}{\theta_i} - 1 \right), \quad \text{RMSE}(\theta) = \left\{ \frac{1}{100} \sum_{i=1}^{100} (\hat{\theta}_i - \theta_i)^2 \right\}^{\frac{1}{2}}$$

Where $\theta = (\gamma, \phi)$ and $\hat{\theta}_i$ is the estimate of θ_i for the i th sample. The results of parameter estimation are presented in Table 1. Considering the estimated values of

Table 1: Summary Results of Model1 Based on 100 Simulated data sets

Pearson Corr.	n	Detrend Response				Detrend Correlated Variable		
		Real	Est.	R.Bias	Rmse	Est.	R.Bias	Rmse
0.55	100	0.5	4.46	8.91	18.28	22.13	43.26	99.21
		5	66.95	12.39	260.34	167.87	35.57	601.99
	400	0.5	2.4	3.8	7.12	10.18	19.36	85.18
		5	65.63	12.13	202.90	107.25	20.45	856.16
	900	0.5	0.55	0.09	0.3	0.83	0.66	1.755
		5	5.72	0.14	4.1	9.27	0.85	14.13
0.65	100	0.5	4.66	7.32	17.18	12.64	24.28	45.13
		5	67.19	12.44	250.7	171.24	33.25	596.65
	400	0.5	2.38	3.76	5.16	2.20	3.41	14.46
		5	52.35	9.47	114.29	27.80	4.56	195.53
	900	0.5	0.56	0.12	0.25	0.57	0.35	1.04
		5	6.14	0.23	3.62	6.89	0.55	8.43
0.85	100	0.5	3.21	5.41	7.06	11.87	22.74	39.43
		5	17.91	2.58	46.82	183.97	35.79	571.47
	400	0.5	2.16	3.33	5	1.16	1.32	3.72
		5	39.63	6.93	105.75	14.1	1.82	49.79
	900	0.5	0.51	0.03	0.24	0.23	-0.18	0.3
		5	5.56	0.11	3.21	5.47	0.29	4.46

the parameters for various sample sizes and correlations, it is obvious that for sample

sizes of lower than 900, nor the estimated value from Detrend Response neither the Detrend Correlated Variable are not efficient due to the high values of R.Bias and Rmse. But when the sample size is increasing to 900, parameter estimation for both sources are satisfactory. In result we can conclude that if the sample size is big enough, which contains valuable amount of information, we can implement a correlated variable to capture the spatial structure of the data even in cases when the value of Pearson correlation coefficient is around 0.5, which is a moderate value.

4 Application

As an application we consider the information available from Household Income and Expenditure Survey (HIES) of Iran. The HIES aims to provide estimates of the average income and expenditure for urban and rural households at provincial and country levels. The HIES has a three stage cluster sampling method which conducts annually with a 0-3 rotating panel design and its target population includes all private and collective settled households in urban and rural areas. There is comprehensive questionnaire which contain a large amount of information about different groups of expenditure and income. Among the favourite variables of HIES data, household income, total expenditure and non-food expenditure have spatial correlation, due to the significant Moran I value. Obviously we have to design the sampling scheme in a way to capture the spatial structure of main variables. The obstacle we have to encounter with is, in the available frame there is no access to the information about income to use for construction of a sample design. Considering the approach presented in this paper, finding a variable which is correlated to these mentioned variables can be solve this problem. After examining candidate variables whose information was available, rent of housing unit is selected as a suitable variable. In this application study we focus on income, but the method is the same for other main variables.

Since income is one of the main output of HIES, it is necessary to produce the sampling design by interfering its spatial structure. Fortunately, we have access to the rent of housing unit from other resources. In conclusion, in order to evaluate the relationship between income and rent, for the sake of grabbing spatial structure of income using another variable, we choose rent. The Pearson correlation coefficient between income and rent is around 0.4. Figure 1, represent the bubble chart of income and rent variables. Since too many bubbles can make the chart hard to read and bubble charts have a limited data size capacity, some points are chosen randomly to make the bubble charts. The sizes of the bubbles in Figure 1 (a) and

(b) are determined by the values of income and rent respectively. Although the two bubble chart are not exactly the same, but it is clear that they are following the same pattern.

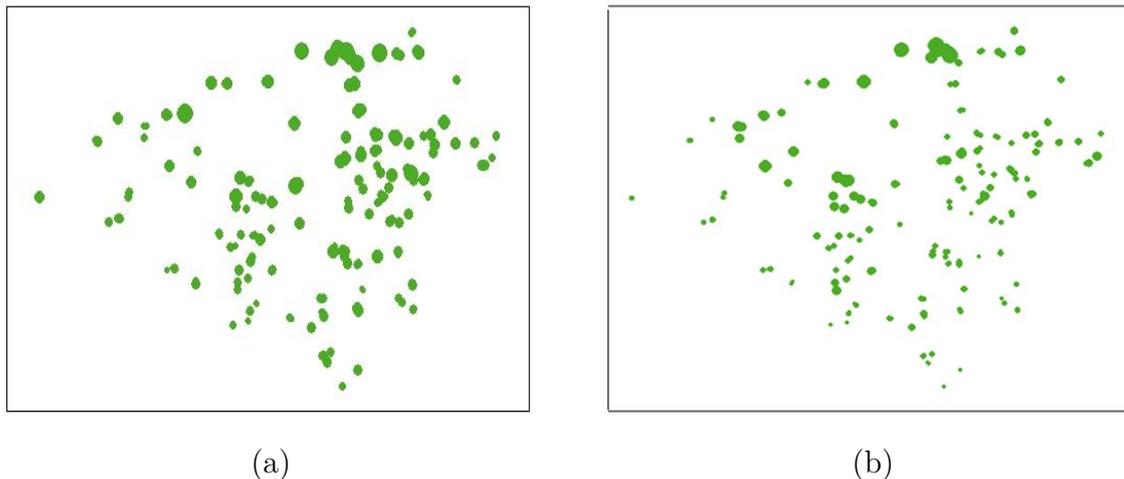


Figure 1: Bubble plot of (a) Income and (b) Rent

The sample households are distributed over geographical areas known as blocks, each block contains some sampling units but their size are not the same. Geographical information about households are available in block level. Therefore, in order to evaluate the spatial structure of income we calculate the average values of household income in block level. The Moran I value showed the significant spatial correlation. We fit different kinds of known variograms, such as Gaussian, Spherical, Matern, Exponential and Power, to the income data using `gstat` and `sp` packages in R. Parameters are estimated by use of MLE method. The exponential variogram is the best fitted one and sum of squared errors of the fitted model (`SSErr`) is equal to 0.1. The fitted variogram is shown in Figure 2.

The nugget effect, sill and range are 0.06, 0.23 and 22.6 respectively. The Pearson correlation coefficient between income and rent is around 0.4, we try to capture the spatial structure of rent by fitting different kinds of variograms. In preliminary study we found that Area of Housing Unit is an effective variable on the rent value. Hence in the first step a regression model is fitted to remove the covariate effect. Then variogram fitting was done on the residuals of the regression model. At last the exponential variogram is the best fitted one as for the income variable. The `SSErr`, nugget effect, sill and range are 0.12, 0.17, 0.4 and 13.75 respectively. So in the absence of information about income we can implement the rent information to capture the spatial structure. Although the estimated parameters values are not

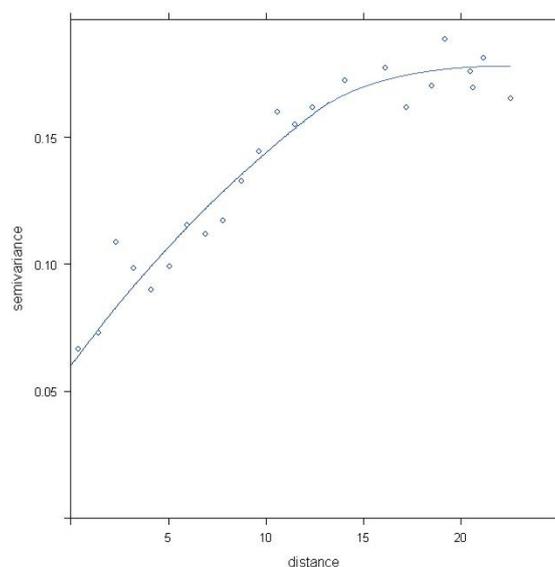


Figure 2: Variogram of Income.

exactly the same for income and rent variogram, but the spatial covariance structure of both of them are follow the exponential pattern. So in the absence of information about income we can implement the rent information to capture the spatial structure, which is not exactly the same as spatial structure of the income but reflect it in a reasonable manner.

5 Discussion and Conclusion

When spatial dependence is existed, finding ways to distinguish its structure and take it to account is essential. Morans I statistic provide information about value and statistical significance of the existence value. In practice we counter some situations where we need to gather information about a spatially correlated variable when there is no information available in the frame. Our intensive simulations for different values of sample sizes and correlations showed that a correlated variable to the spatial interested variable can be used to capture the spatial structure. It is worth to note that even moderate values of Pearson correlation in enough to capture the spatial structure in a reasonable manner. it should be emphasized that Pearson correlation in not the only criteria for decision making and it should be used with care. In conclusion, we recommend to assess and implement correlated variable to grab the spatial structure when required information about the variable of interest is not available. This can be an applicable tool for construction of sampling schemes when there is lack of information in the sampling frame. Obviously achieving to

the spatial structure of variable of interest is not the final step and we have to implement it in designing the sampling scheme. There are different approaches to do spatial sampling and the best one will be selected by experts based on the goals and available requirements.

References

- Anselin, L., Cohen, J., Cook, D., Gorr, W., Tita, G., (2000), Spatial Analyses of Crime, *Criminal Justice*, **4**, 213-262.
- Chou, Y.H. (1997), *Exploring spatial analysis in geographic information systems*. Santa Fe: Onward Press.
- Cliff, A.D., Ord, J.K., (1973), Testing for spatial autocorrelation among regression residuals, *Geographical Anal*, **4**, 267-284.
- Cliff, A.D., Ord, J.K., (1981), *Spatial Processes: Models and Applications*, Pion, London.
- Dale, M.T. (1999), *Spatial Pattern Analysis in Plant Ecology*, Cambridge University Press, Cambridge, UK.
- Haining, R., (1990), *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press, Cambridge, UK.
- Lawson, A.B., Denison, D.G.T., (2000), *Spatial Clustering Modeling*, CRC Press, New York.
- Moran, P.A.P., (1948), The Interpretation of Statistical Maps, *Journal of the Royal Statistical Society. Series B*, **10**, 243-251.
- Moran, P.A.P., (1950), A Test for the Serial Independence of Residuals, *Biometrika*, **37**, 178-181.
- Turner, A. G. (2003), Sampling Frames and Master Samples, *United Nations Secretariat Statistics Division*.
- Wang, F., and Wall, M. M. (2003), Generalized Common Spatial Factor Model, *Biostatistics*, **4**, 569-582.



Approximate Composite Marginal Likelihood Inference in SGLM Models with Skew Normal Latent Variables

Omid Karimi*, Fatemeh Hosseini

Department of Statistics, Semnan University, Semnan, Iran.

Abstract:

Non-Gaussian spatial responses are usually modeled using spatial generalized linear mixed model. In this model, it is a standard assumption that the latent spatial variables have normal distribution, but it is unclear whether the Gaussian assumption holds. The first purpose of this paper is to use a multivariate closed skew normal distribution for the latent variables, which is greater class than the class of multivariate normal. Since the likelihood function of this model cannot usually be given in a closed form, thus maximum likelihood approach is very challenging. The second purpose of this paper is to propose a new approximate pairwise maximum likelihood method for the inference and spatial prediction of the spatial generalized linear mixed models with skew normal latent variables. This approximate inference method is fast and deterministic, using no sampling based strategies. The performance of the proposed model and method are illustrated through a simulation study.

Keywords: Spatial Generalized Linear Mixed Model, Latent Variable, Pairwise Likelihood, Approximate Inference.

Mathematics Subject Classification (2010): 91B72, 62J12.

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). Spatial correlation of the data is usually modelled by latent variables. The

*Speaker: omid.karimi@semnan.ac.ir

most common assumption for the latent random field is to use a Gaussian random field. Inference of the model parameters and spatial prediction has been studied intensely for this Gaussian assumption, see e.g. [Breslow and Clayton \(1993\)](#), [Zhang \(2002\)](#), [Varin *et al.* \(2005\)](#), [Baghishani *et al.* \(2012\)](#), [Baghishani *et al.* \(2011\)](#) [Hosseini \(2016\)](#). Erroneous normal assumptions have influence on the estimation of the model parameters and the accuracy of spatial prediction.

In this paper we will consider a Closed Skew Normal (CSN) distribution for the spatial latent variable. This distribution is fully parametric, and contain several closed form solutions, facilitating efficient inference of the model parameters and prediction of the latent process. [Hosseini and Karimi \(2019\)](#) proposed an approximate likelihood approach for SGLMM with Normal latent variables. For the resulting SGLMM with CSN latent variable, we introduce the new approximate likelihood inference methods similar to [Hosseini and Karimi \(2019\)](#). The main contribution of the current work is to use a CSN approximation for the conditional distribution of the latent variables, and to show that this allows us to perform fast approximate inference and prediction, see e.g. [Hosseini *et al.* \(2011\)](#), [Hosseini and Mohammadzadeh \(2012\)](#) and [Hosseini and Karimi \(2019\)](#). The inference step is done in a manner similar to the Laplace approximation ([Tierney and Kadane \(1986\)](#), [Rue and Martino \(2007\)](#), [Rue *et al.* \(2009\)](#) and [Eidsvik *et al.* \(2009\)](#)), but based on the CSN approximation instead of the usual normal approximation, [Hosseini *et al.* \(2011\)](#) and [Hosseini and Karimi \(2019\)](#).

This paper is organized as follows: In Section 2 the closed skew normal ([Dominguez-Molina *et al.* \(2003\)](#)) and the Spatial GLMM with Closed Skew Normal Latent Variables ([Hosseini and Karimi \(2019\)](#)) are presented. The proposed method are described in Section 3. Section 4 shows results on a simulation study. Closing remarks are given in Section 5.

2 SGLMM with Closed Skew Normal Latent Variables

In this section we define the CSN distributions and present the SGLMM with CSN as the distribution for the spatial latent variables.

2.1 Closed Skew Normal Distribution

The CSN distribution is a class of statistical distribution, which includes the SN and normal distributions as a special cases. In order to the CSN distribution extends the

SN by allowing more flexibility on the skewness directions. 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 \mathbf{x} follows a multivariate CSN distribution, with parameters $\boldsymbol{\mu}$, Σ , D , $\boldsymbol{\nu}$, and Δ , if its density function is given by

$$f_{n,q}(\mathbf{x}|\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta) = [\Phi_q(\mathbf{0}; \boldsymbol{\nu}, \Delta + D\Sigma D')^{-1} \phi_n(\mathbf{x}; \boldsymbol{\mu}, \Sigma) \Phi_q(D(\mathbf{x} - \boldsymbol{\mu}); \boldsymbol{\nu}, \Delta)] \quad (2.1)$$

In short we denote this distribution by $CSN_{n,q}(\boldsymbol{\mu}, \Sigma, D, \boldsymbol{\nu}, \Delta)$, where the length n vector $\boldsymbol{\mu}$ is a location parameter, the positive definite $n \times n$ matrix Σ 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 Δ . For $q = 1$, $\boldsymbol{\nu} = 0$, $\Delta = 1$ and $D = \boldsymbol{\lambda}^\top \Sigma^{-\frac{1}{2}}$, the CSN density reduces to that of the skew normal distribution. When D is a zero matrix, the density in equation (2.1) reduces to the density of a multivariate normal distribution. The first moment of the CSN distribution is

$$E(X) = \boldsymbol{\mu} + \Sigma D^\top \boldsymbol{\psi}, \quad (2.2)$$

where $\boldsymbol{\psi} = \Phi_q^*(\mathbf{r}, \boldsymbol{\nu}, \Delta + D\Sigma D^\top) / \Phi_q(\mathbf{r}, \boldsymbol{\nu}, \Delta + D\Sigma D^\top)|_{\mathbf{r}=\mathbf{0}}$, and for any positive definite matrix Ω , $\Phi_q^*(\mathbf{r}; \boldsymbol{\nu}, \Omega) = [\nabla_r \Phi_q(\mathbf{r}; \boldsymbol{\nu}, \Omega)]^\top$ with gradient operator $\nabla_r = (\partial/\partial r_1, \dots, \partial/\partial r_q)^\top$. The variance of X is

$$V(X) = \Sigma + \Sigma D^\top \boldsymbol{\xi} D \Sigma - \Sigma D^\top \boldsymbol{\psi} \boldsymbol{\psi}^\top D \Sigma$$

where $\boldsymbol{\xi} = \Phi_q^{**}(\mathbf{r}, \boldsymbol{\nu}, \Delta + D\Sigma D^\top) / \Phi_q(\mathbf{r}, \boldsymbol{\nu}, \Delta + D\Sigma D^\top)|_{\mathbf{r}=\mathbf{0}}$ and $\Phi_q^{**}(\mathbf{r}; \boldsymbol{\nu}, \Omega) = [\nabla_r \nabla_r^\top \Phi_q(\mathbf{r}; \boldsymbol{\nu}, \Omega)]^\top$. The CSN distribution has some desirable properties, similar to those of the normal distribution. For instance, the CSN distribution is closed under marginalization and linear conditioning, see [Dominguez-Molina *et al.* \(2003\)](#). The CSN distributions has some favorable properties similar to those of the multivariate normal distribution. It is for example closed under marginalization, partition \mathbf{x} into $\mathbf{x} = (\mathbf{x}_1^\top, \mathbf{x}_2^\top)^\top$, where the dimension of \mathbf{x}_1 is k and the dimension of \mathbf{x}_2 is $n - k$. Marginally,

$$\mathbf{x}_1 \sim CSN_{k,q}(\boldsymbol{\mu}_1, \Sigma_{11}, D^*, \boldsymbol{\nu}, \Delta^*), \quad (2.3)$$

where $D^* = D_1 + D_2 \Sigma_{21} \Sigma_{11}^{-1}$, $\Delta^* = \Delta + D_2 \Sigma_{22 \cdot 1} D_2'$, $\Sigma_{22 \cdot 1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$, with $\boldsymbol{\mu}_1$, Σ_{11} , Σ_{22} , Σ_{12} , Σ_{21} , D_1 and D_2 obtained by appropriate 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 $D = (D_1, D_2)$. The CSN distribution is closed under conditioning; with the previous partitioning, the conditional distribution of \mathbf{x}_2 given

\mathbf{x}_1 is

$$(\mathbf{x}_2|\mathbf{x}_1) \sim CSN_{n-k,q}(\boldsymbol{\mu}_2 + \Sigma_{21}\Sigma_{11}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Sigma_{22.1}, D_2, \boldsymbol{\nu} - D^*(\mathbf{x}_1 - \boldsymbol{\mu}_1), \Delta). \quad (2.4)$$

Karimi *et al.* (2010) and Karimi and Mohammadzadeh (2011) define a type of closed skew Gaussian random field as following:

Let $W(s) = \mu(s) + E_1(s)$, $s \in R^2$, where $\mu(s)$ is a real valued function and $E_1(s)$ is a spatial Gaussian random field with covariance function

$$\Sigma_{\theta}(s_1, s_2) = \text{Cov}(E_1(s_1), E_1(s_2)), \quad (2.5)$$

where $\boldsymbol{\theta}$ is the vector of covariance parameters. Let the random vector $\mathbf{E}_2 = (E_{21}, \dots, E_{2q})^\top$, $q \geq 1$ be independent of $\mathbf{E}_1 = (E_1(s_1), \dots, E_1(s_n))^\top$ and distributed as $N_q(0, \Delta)$, where Δ is a positive definite matrix. Define $V = (V_1, \dots, V_q)^\top$ with j th element given by

$$V_j = -\nu_j + \sum_{i=1}^n d_j(s_i)E_1(s_i) + E_{2j}, \quad j = 1, \dots, q, \quad (2.6)$$

where $\nu_j \in R$ and each $d_j(\cdot)$ is a real valued function. According to Proposition 3.1 in Karimi and Mohammadzadeh (2011), $X(s) = [W(s)|V \geq 0]$ is a discrete form of the CSG random field. Then according to properties of random field, if $\mathbf{x}(s) = (x(s_1), \dots, x(s_n))$ be a realization of the stationary random field $X(s)$, $\mathbf{x}(s)$ has a CSN distribution as $CSN_{n,q}(\boldsymbol{\mu}, \Sigma_{\theta}, D, \Delta, \boldsymbol{\nu})$, where the elements of matrix D and vector $\boldsymbol{\nu}$ are ν_j and d_j in (2.6).

2.2 Model

Let $\mathbf{x} = (x_1, \dots, x_n)'$ with density $CSN_{n,1}(H\boldsymbol{\beta}, \Sigma_{\theta}, \boldsymbol{\lambda}'\Sigma_{\theta}^{-\frac{1}{2}}, \mathbf{0}, I)$ be a realization of a CSG random field at n sites $\{s_1, \dots, s_n\}$ in a domain $\chi \subseteq R^n$. Here, the location parameter $H\boldsymbol{\beta}$ consists of $n \times (p+1)$ matrix H of covariates, and $p+1$ regression parameters $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)'$. I is a $n \times n$ identity matrix. The spatial interaction matrix Σ_{θ} is a positive definite $n \times n$ matrix, with two dimensional parameter $\boldsymbol{\theta} = (\sigma, \phi)$ indicative of the scale and spatial 'correlation' length, respectively. In this paper we use an isotropic exponential 'correlation' structure for the entries in this matrix. This entails that $\Sigma_{\theta}(i, j) = \sigma^2 \exp(-\|s_i - s_j\|/\phi)$, where $\|s_i - s_j\|$ is the Euclidean distance between sites s_i and s_j . We assume, the skewness parameter $\boldsymbol{\lambda} = \lambda_0 \mathbf{1}$ is a length n vector, with one free parameter λ_0 for maintaining a parsimonious model. We define $\boldsymbol{\eta} = (\boldsymbol{\beta}', \boldsymbol{\theta}, \lambda_0)$. From equation (2.1) we have

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

Let $\mathbf{y}' = (y_1, \dots, y_k)$ represent the discrete spatial response variables at the observation sites $\{s_1, \dots, s_k\}$. We assume that the measurements are conditionally independent with likelihood $f(\mathbf{y}|\mathbf{x})$ of an exponential family (McCullagh and Nelder

(1989)). For each observation site s_i , $i = 1, \dots, k$, this can be written $f(y_i|x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}$, where $b(x_i)$ is the cumulant function. 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 a known link function. To summarize, the model has the following components:

$$\begin{aligned} f(\mathbf{y}, \mathbf{x}, \boldsymbol{\eta}) &= f(\mathbf{y}|\mathbf{x})f(\mathbf{x}|\boldsymbol{\eta}) \\ &= \frac{2}{(2\pi)^{n/2}|\Sigma_\theta|^{1/2}} \exp\left(\sum_{i=1}^k [y_i x_i - b(x_i) + c(y_i)] - \frac{1}{2}(\mathbf{x} - H\boldsymbol{\beta})'\Sigma_\theta^{-1}(\mathbf{x} - H\boldsymbol{\beta})\right) \\ &\quad \times \Phi(\boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}(\mathbf{x} - H\boldsymbol{\beta})). \end{aligned} \quad (2.7)$$

3 Approximate Pairwise Likelihood Inference

In this section an approximate algorithm is proposed to obtain estimates of model parameters based on the EM algorithm and pairwise likelihood approach when the latent variables have a CSN distribution. Pairwise likelihood is a special case of a more generalized class of pseudo likelihoods called composite likelihood by the term pairwise likelihood is used by instead. Let $\mathbf{x} = (x_1, \dots, x_n)^\top$ with density $N_n(H\boldsymbol{\beta}, \Sigma_\theta)$ and $f(y_i|x_i) = \exp(y_i x_i - b(x_i) + c(y_i))$, the likelihood function is

$$L(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}) = \int \left\{ \prod_{i=1}^k f(y_i|x_i) \right\} f(\mathbf{x}|\boldsymbol{\beta}, \boldsymbol{\theta}) d\mathbf{x}, \quad (3.1)$$

which cannot usually be given in a closed form, and that is because the integration dimension in (3.1) is equal to the number of latent variables, and consequently it is intractable to find the MLE by directly maximizing L . Pairwise likelihood is the product of the bivariate likelihood as follows

$$\begin{aligned} PL(\boldsymbol{\beta}, \boldsymbol{\theta}|\mathbf{y}) &= \prod_{(\ell, \ell') \in \aleph} L(\boldsymbol{\beta}, \boldsymbol{\theta}|y_\ell, y_{\ell'}) \\ &= \prod_{(\ell, \ell') \in \aleph} \int \int f(y_\ell|x_\ell) f(y_{\ell'}|x_{\ell'}) f(x_\ell, x_{\ell'}|\boldsymbol{\beta}, \boldsymbol{\theta}) dx_\ell dx_{\ell'}, \end{aligned}$$

where, \aleph is a subset of pairwise neighbors. A moving neighborhood method is used by excluding far apart pairs that have little spatial correlation.

Theorem 3.1. (*Hosseini et al. (2011)*) Let $\mathbf{x}|\boldsymbol{\eta} \sim CSN_{n,1}(H\boldsymbol{\beta}, \Sigma_\theta, \boldsymbol{\lambda}'\Sigma_\theta^{-\frac{1}{2}}, \mathbf{0}, I)$, $f(y_i|x_i) = \exp\{y_i x_i - b(x_i)\}$. By linearizing the likelihood part of (2.7) at a fixed value of \mathbf{x} , then

$$(\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}) \approx CSN_{n,1}(\hat{\boldsymbol{\mu}}_{\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}}, \hat{\Sigma}_{\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}}, \hat{D}_{\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}}, \hat{\boldsymbol{\nu}}_{\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}}, 1), \quad (3.2)$$

where $\hat{\boldsymbol{\mu}}_{\mathbf{x}|\mathbf{y}, \boldsymbol{\eta}} = H\boldsymbol{\beta} + \Sigma_\theta A'R^{-1}(z(\mathbf{y}, \mathbf{x}^{obs}) - AH\boldsymbol{\beta})$, and $z_i(y_i, x_i) = [y_i - b'(x_i) + x_i b''(x_i)]/b''(x_i)$, $i = 1, \dots, k$, is a linearization of the exponential family likelihood part of equation (2.7) at a fixed value of \mathbf{x} . Moreover, $R = A\Sigma_\theta A' + P$ and $P = P(\mathbf{x})$ is a diagonal matrix with entries element $P(i, i) = 1/b''(x_i)$, $i = 1, \dots, k$. Finally,

$\hat{\Sigma}_{x|y,\eta} = \Sigma_\theta - \Sigma_\theta A' R^{-1} A \Sigma_\theta$, $\hat{D}_{x|y,\eta} = \boldsymbol{\lambda}' \Sigma_\theta^{-\frac{1}{2}}$, and $\hat{\boldsymbol{\nu}}_{x|y,\eta} = \boldsymbol{\lambda}' \Sigma_\theta^{-\frac{1}{2}} (H\boldsymbol{\beta} - \boldsymbol{\mu}_{x|y,\eta})$. See Appendix Hosseini et al(2011) for further explanation.

For fitting a CSN approximation, we choose a starting value $\boldsymbol{x}^{(0)}$. For instance set $\boldsymbol{x}^{(0)} = E(\boldsymbol{x}|\boldsymbol{\eta})$. Set $m = 0$. Then, calculate

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

and let

$$\begin{aligned} \boldsymbol{x}^{(m+1)} &= \hat{E}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta}) = \hat{\boldsymbol{\mu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}) + \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}) D'_{x|y,\eta} \hat{\boldsymbol{\psi}}, \\ \hat{\boldsymbol{\psi}} &= \psi(0, \hat{\Sigma}_{x|y,\eta}(\boldsymbol{x}^{(m)}), D_{x|y,\eta}, \hat{\boldsymbol{\nu}}_{x|y,\eta}(\boldsymbol{x}^{(m)}), 1). \end{aligned}$$

Finally we set $m = m + 1$ and update $\hat{f}(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\eta})$ using the theorem above. Convergence is obtained after a few iterations. The CSN-APEM algorithm is as follows:

- 1) Choose a starting value $\boldsymbol{\eta}^{(0)}$, such that $PL(\boldsymbol{\eta}^{(0)}|\boldsymbol{y}) > 0$ and set $m = 0$.
- 2) Approximation step: choose a starting value $\boldsymbol{x}^{(0)}$. For instance set $\boldsymbol{x}^{(0)}$ to be the mode of $f(\boldsymbol{x}_\ell, x_{\ell'}|\boldsymbol{\eta}^{(m)})$. Set $d = 0$.
- (a) Calculate $\hat{f}(\boldsymbol{x}_\ell, x_{\ell'}|\boldsymbol{y}_\ell, y_{\ell'}, \boldsymbol{\eta}^{(m)})$ from Theorem 3.1.
- (b) Let $\boldsymbol{x}^{(d)}$ to be mode of $\hat{f}(\boldsymbol{x}_\ell, x_{\ell'}|\boldsymbol{y}_\ell, y_{\ell'}, \boldsymbol{\eta}^{(m)})$. Set $d = d + 1$. Go to (a) until convergence is reached.
- 3) Expectation step: evaluate

$$\begin{aligned} Q(\boldsymbol{\eta}|\boldsymbol{\eta}^{(m)}) &= \sum_{(\ell, \ell') \in \mathbb{N}} E(\log\{f(x_\ell, x_{\ell'}, y_\ell, y_{\ell'}|\boldsymbol{\eta})\} | y_\ell, y_{\ell'}, \boldsymbol{\eta}^{(m)}) \\ &= \sum_{(\ell, \ell') \in \mathbb{N}} \int \int \log\{f(x_\ell, x_{\ell'}, y_\ell, y_{\ell'}|\boldsymbol{\eta})\} \\ &\quad \times \hat{f}(x_\ell, x_{\ell'}|y_\ell, y_{\ell'}, \boldsymbol{\eta}^{(m)}) dx_\ell dx_{\ell'}. \end{aligned} \quad (3.3)$$

- 4) Maximization step: choose $\boldsymbol{\eta}^{(m+1)}$ such that $\boldsymbol{\eta}^{(m+1)} = \operatorname{argmax}_\boldsymbol{\eta} \hat{Q}(\boldsymbol{\eta}|\boldsymbol{\eta}^{(m)})$. Set $m = m + 1$ and go to step (2) until convergence is reached.

4 simulation

First, $n = 200$ random locations are generated inside an irregular grid of France region. Then, we fix parameters of the CSN distribution and draw latent variables from $f(\boldsymbol{x}|\boldsymbol{\eta}) = CSN_{200,1}(0, \Sigma_\theta, \lambda_0 \mathbf{1}' \Sigma_\theta^{-\frac{1}{2}}, 0, 1)$. We set parameters $\lambda_0 = 2$ and $\boldsymbol{\theta}^\top = (\sigma^2 = 1, \phi = 3)$. Conditional on the latent variables, binomial responses $y_j, j = 1, \dots, 200$ were generated according to $y_j \sim \text{Bin}(u_j, p_j)$, where $u_j = 5, 50, 100, 150, p_j = \frac{\exp(x_j)}{1 + \exp(x_j)}$. One realization of simulated data is shown in Figure 1. For each observation, a neighborhood of radius 3 is used to construct pairs using 12 neighbors randomly sampled without replacement. We used two Spatial GLM models, with normal spatial latent variables and CSN. To obtain ML estimates of the parameters, we run CSN-APEM algorithm for CSN model and APEM

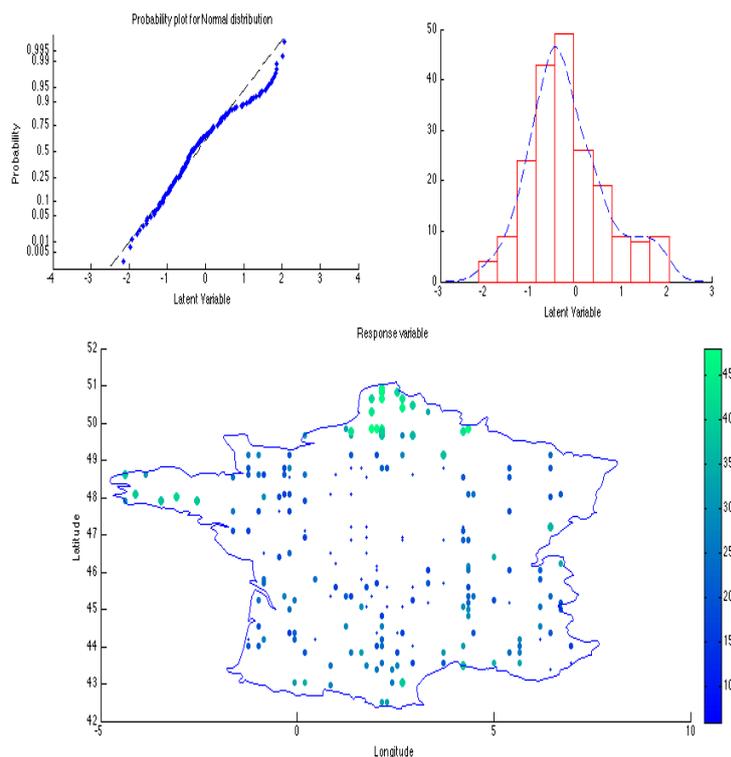


Figure 1: Realization from the spatial Binomial model; Normal QQ plot (top-left); histogram (top-right); map of the simulated latent variables (below).

algorithm in [Hosseini and Karimi \(2019\)](#) for normal model. Keeping the spatial design fixed, the above data generation scheme is carried out for 1000 data set. The constructed results after convergence of the algorithms are summarized in [Table 1](#). The simulation results in [Table 1](#) show that for all parameters the averages of the MSE, estimates and standard deviations are smaller for CSN than for normal model.

Conclusion

In this article we used the closed skew normal distribution to model spatial latent variables and proposed an approximate pairwise maximum likelihood method. A new algorithm was introduced to obtain maximum likelihood estimates of the model parameters. We extended the approach of [Hosseini and Karimi \(2019\)](#) to the spatial generalized linear mixed models with the CSN latent variables. In the simulation study, we used the APEM algorithm ([Hosseini and Karimi \(2019\)](#)) and the proposed algorithm to obtain the maximum likelihood estimates of the normal and the closed skew normal models, respectively. Estimation accuracy of two models were compared and showed wrong normal assumption for latent variables can be

Table 1: Simulation results from 1000 data set: Averages of Estimate, sd and MSE of the estimates

Par.	R. val.	unit	CSN			Normal		
			Ave. Es.	Ave. Std.	Ave. MSE	Ave. Es.	Ave. Std.	Ave. MSE
σ^2	1	5	1.1029	0.0416	0.0816	1.5941	0.4460	0.3098
		50	1.2319	0.0414	0.1103	1.4117	0.0472	0.1895
		100	1.2218	0.0296	0.1011	1.3674	0.0364	0.1219
		150	1.1763	0.0264	0.0858	1.2554	0.0317	0.1139
ϕ	3	5	5.5718	0.2913	1.7036	6.3225	0.4189	2.2812
		50	4.0832	0.2094	1.4427	4.5631	0.2194	1.9961
		100	3.4844	0.1495	1.1756	3.7021	0.1703	1.9306
		150	3.1967	0.1337	0.9180	3.5149	0.1619	1.4249
λ_0	2	5	2.8584	0.2287	0.9688	-	-	-
		50	2.7938	0.2241	0.8619	-	-	-
		100	2.5645	0.2165	0.8178	-	-	-
		150	2.2918	0.2276	0.9417	-	-	-

effect on accuracy of parameters and predictions.

References

- Baghishani, H., Rue, H. & Mohammadzadeh, M., (2012), On a Hybrid Data Cloning Method and its Application in Generalized Linear Mixed Models, *Statistics and Computing*, **22**, 613-597.
- Baghishani, H. & Rue, H. & Mohammadzadeh, M., (2011), A Data Cloning Algorithm for Computing Maximum Likelihood Estimates in Spatial Generalized Linear Mixed Models, *Computational Statistics and Data Analysis*, **55**, 1748-1759.
- Breslow, N. E. & Clayton, D. G. (1993). Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, **88**, 9-25.
- 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.
- 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 Skew Normal 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.

Hosseini, F., Eidsvik, J., & Mohammadzadeh, M., Approximate Bayesian Inference in Spatial GLMM with Skew Normal Latent Variables, *Computational Statistics and Data Analysis*, **55**1791-1806.

Hosseini, F. & Mohammadzadeh, M. Bayesian Prediction for Spatial GLMMs with Closed Skew Normal Latent Variables, *Australian & New Zealand Journal of Statistics*, **54**, 43-62, (2012).

Hosseini, F. A New Algorithm for Estimating the Parameters of the Spatial Generalized Linear Mixed Models, *Environmental and Ecological Statistics*, **23**, 205-217, (2016).

Hosseini, F., Karimi, O. (2019). Approximate Likelihood Inference in Spatial Generalized Linear Mixed Models with Closed Skew Normal latent Variables *Communication in Statistics- Simulation and Computation* Doi:10.1080/03610918.2018.1476700

Hosseini, F., Karimi, O. (2019). Approximate Composite Marginal Likelihood Inference in spatial generalized linear mixed models, *Journal of Applied Statistics* **46:3**, 542-558.

Karimi, O., Omre, H. & Mohammadzadeh, M., (2010). Bayesian Closed-skew Gaussian Inversion of Seismic AVO Data for Elastic Material Properties, *Geophysics*, **75**, R1-R11 (2010).

Karimi, O., Mohammadzadeh, M., (2011). Bayesian Spatial Prediction for Discrete Closed Skew Gaussian Random Field, *Mathematical Geosciences* **43**, 565-582.

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.* **71**, 319-392.

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.



Analysis of Spatial Data with Non-Ignorable Missingness

Mohsen Mohammadzadeh*, Samira Zahmatkesh

Department of Statistics, Tarbiat Modares University, Tehran, Iran.

Abstract:

Many researchers are dealing with spatially dependent data in various sciences such as meteorology, ecology, geology and epidemiology in which there is often a notable amount of missing values. Because spatial data are often collected in non-laboratory environments, some of the factors affecting measurement, such as environmental and atmospheric conditions, sample units locations, or the time of collecting observations, make missing data inevitable. For spatial data, due to dependency between observations, missing values that are located at the spatial or temporal neighbourhoods of the observations can include useful information that the retrieval of this lost data can increase the accuracy of data analysis. In this paper, the joint modeling of the spatial measurement process and the missing process is proposed using the shared parameter model technique. To model data and missingness process, spatial generalized linear mixed model is used, and to make inference, Bayesian approach and integrated nested Laplace approximation is used. Next a computationally effective approach is given by stochastic partial differential equation which consists in performing computation using a Gaussian Random Field thus allowing us to adopt the integrated nested Laplace approximation approach. Then, the presented models, are evaluated and numerically compared in a simulation study, and their application in real data example is showed.

Keywords: Missing Values, Spatial Data, Non-Ignorable, Joint Model, INLA, SPDE.

Mathematics Subject Classification (2010): 62H11, 62N01, 62N86.

*Speaker: mohsen.m@modares.ac.ir

1 Introduction

In this paper we develop methods for dealing with missing data in a spatial response variable when estimating model parameters. Missing outcome data is a problem in a number of applications. Definitions for the missing outcome data mechanisms were originally introduced by Rubin (1976). Missing data are classified into three types: missing completely at random (MCAR), missing at random (MAR) and missing not at random (MNAR). Let m_i be an indicator variable that is 1 if the outcome y_i is observed and 0 otherwise, and let \mathbf{x}_i be a vector of covariates observed for each individual i . Then the data is said to be MCAR if $Pr(m_i|y_i; \mathbf{x}_i) = Pr(m_i)$. In fact MCAR occurs when missing process does not depend on observed and unobserved data. MAR is a less strong assumption in which missing process depends only on the observed data, i.e. $Pr(m_i|y_i; \mathbf{x}_i) = Pr(m_i|\mathbf{x}_i)$. Here the missing data mechanism does not depend on the outcome conditionally on the observed covariates \mathbf{x}_i . If interest lies in the regression of y_i on \mathbf{x}_i , the estimated parameters in a complete case analysis are unbiased under MAR, since we adjust for \mathbf{x}_i . Under MCAR or MAR the missingness mechanism is said to be ignorable and many standard statistical techniques are valid (Molenberghs et al., 2015). Estimating model parameters based on the complete cases will yield unbiased estimates under MAR. Likelihood based techniques are also valid as long as the distribution of y_i is correctly specified. In these two cases missing data is said to be ignorable so that the usual Bayesian and likelihood inferences based on the observed data are valid (Little and Rubin, 2002).

When neither MCAR nor MAR hold the data are MNAR, the missing data mechanism is called non-ignorable, or alternatively the data is called missing not at random (MNAR). In this case the common methods that are effective at ignorable case introduce bias inferences and other methods of inference need to be considered. Therefore a 'statistically principled' approach is to build a joint model and undertake sensitivity analysis, so that it combines information in the observed data with assumptions about the missing value mechanism. One framework for jointly modelling the measurement process and missing process is Shared Parameter Model (SPM). A common random coefficients is introduced in both the missing model and the measurement model (Steinland et al., 2014). Karimi and Mohammadzadeh (2011) proposed a spatial regression model which captures skewness in response variable and also model and predict the missing observations by a Bayesian approach.

In this paper we consider spatial data and assume that they are MNAR, it means that missingness is probably influenced by some latent random fields. Using a spatial generalized linear mixed (SGLM) model (Breslow and Clayton, 1993; Diggle and Tawn, 1998) measurement process and the missing process will be modeled

and in each model, a spatial latent random field describes the association between response and explanatory variables. In order to account for missing values, we define a spatial joint model and employ the SPM approach. So a common spatial latent random field in the model of each process is considered as a shared parameter. For the spatial latent random field, we use a Gaussian prior. Traditionally Markov chain Monte Carlo (MCMC) methods have been used making inference for Bayesian latent Gaussian models. [Rue and Martino \(2007\)](#) introduced an approximated Bayesian approach via Integrated Nested Laplace Approximation (INLA) for latent Gaussian Markov random field (GMRF) models and proved that it is faster than MCMC while the results are equivalent. INLA has successfully been applied in various setting and the approximations are shown to have high accuracy ([Rue and Martino, 2007](#); [Eidsvik et al., 2009](#); [Holand et al., 2010](#)). In order to overcome the computational costs of linear algebra operation required for model fitting, estimations and spatial predictions that arise when facing with large spatial data set we consider an approach which consists in representing a Gaussian Random fields (GRF) that is continuously indexed as a GMRF that is a discretely indexed random process. This approach is provided through stochastic partial differential equations (SPDE) ([Lindgren et al., 2011](#)). So in Bayesian inference for a GMRF it is possible to use INLA algorithm as an alternative to MCMC methods.

In Section 2, the SGLM and joint model formulations and some computational issues and moderating with the SPDE approach are discussed. In Section 3, a simulation study has been implemented in order to show the efficiency of the joint modeling in MNAR case. In Section 4, the surface water temperature data set for lake Vanern in the south of Sweden has been modeled and the results are reported.

2 Generalized Linear Mixed Model

The response variable $y(s)$ at location $s \in D \subseteq R^2$ can be written as

$$y(s) = H\beta + x(s) + \varepsilon(s),$$

where $x(s)$ is a spatial Gaussian random field with covariance matrix $\Sigma_\theta = (C_\theta(s_i, s_j))$, where θ is the vector of spatial dependence parameters, H is an $n \times (p + 1)$ matrix of covariates (design matrix), $\beta = (\beta_1, \dots, \beta_p)$ is the vector of relevant coefficients and $\varepsilon(s)$ is an error term with distribution $N(0, \sigma_\varepsilon^2)$. Let $\eta = (\beta, \theta, \sigma_\varepsilon^2)$ denotes the vector of model parameters and $(y(s_1), \dots, y(s_n))$ and $(x(s_1), \dots, x(s_n))$ are the vectors of response observations and realizations of the random field $x(s)$ at n locations $\{s_1, \dots, s_n\}$, respectively. For simplicity we denote these vectors with $\mathbf{y} = (y_1, \dots, y_n) = (y(s_1), \dots, y(s_n))$ and $\mathbf{x} = (x_1, \dots, x_n) = (x(s_1), \dots, x(s_n))$. Un-

der the conditional independence assumption of $y(s)$ on the latent random field $x(s)$, $\pi(\mathbf{y}|\mathbf{x})$ is in a class of exponential family with the general form $\pi(y_i|x_i) = \exp\{y_i x_i - b(x_i) + c(y_i)\}$, where $b(\cdot)$ and $c(\cdot)$ are known functions. Bayesian hierarchical models requires a likelihood model for response \mathbf{y} and determining prior distributions for the spatial latent random field \mathbf{x} and the hyper parameters. Then marginal posterior distributions of the model parameters $\eta = (\beta, \theta, \sigma_\varepsilon^2)$ and \mathbf{x} are desired. By assuming independence priors, the joint posterior distribution is given by

$$\pi(\mathbf{x}, \beta, \theta, \sigma_\varepsilon^2 | \mathbf{y}) \propto \pi(\mathbf{y} | \mathbf{x}, \beta, \theta, \sigma_\varepsilon^2) \pi(\mathbf{x} | \theta) \pi(\beta) \pi(\theta) \pi(\sigma_\varepsilon^2). \quad (2.1)$$

Calculation of this posterior distribution using MCMC methods, is time-consuming, so an approximate Bayesian method using INLA has been suggested.

3 Joint Modelling Formulation

Let $\mathbf{m} = (m_1, \dots, m_n)$ be a vector of missing data indicators defined by

$$m_i = \begin{cases} 0 & y_i \text{ is observed} \\ 1 & y_i \text{ is missing} \end{cases}$$

where $m_i | \pi_i \sim Bin(1, \pi_i)$. In presence of non-ignorable missing data, by use of this vector the missing process can be modelled (Little and Rubin, 2002). The probability of missingness for each individual, π_i , can be modelled by use of generalized linear model and logit link function as

$$\text{logit}(\pi_i) = \alpha_0 + \alpha_1 x_i.$$

Note that other latent variables and covariates could be added to the model. Then assuming independence of observations \mathbf{y} and missing process \mathbf{m} conditioned on \mathbf{x} , the joint density will be factorized as

$$\pi(\mathbf{y}, \mathbf{m} | \mathbf{x}, \eta, \phi) = \pi(\mathbf{y} | \mathbf{x}, \eta) \pi(\mathbf{m} | \mathbf{x}, \phi), \quad (3.1)$$

here \mathbf{x} is the common part of two models and $\phi = (\alpha_0, \alpha_1)$. In presence of missing data, the association between \mathbf{y} and \mathbf{m} (measurement process and missing process) will be included in the spatial latent random field \mathbf{x} . In the modeling of π_i , parameter α_1 denotes the severity of dependence between missingness and \mathbf{x} , indeed it describes the association between \mathbf{y} and \mathbf{m} . In order to complete Bayesian modeling, it is needed to specify the priors for the hyper-parameters α_0 and α_1 . Now we have an SGLM model with two responses \mathbf{y} and \mathbf{m} . The efficiency of this model will discuss in the next Section by a simulation study. We use INLA for a joint model formulation as introduced in Steinsland et al. (2014) to avoid time-consuming calculations of full conditional and marginal posterior distributions. Its Bayesian inference can be

carried out within a fraction of the computation time of MCMC algorithms. Here we use INLA along with the SPDE-approach to spatial modeling introduced by Lindgren et al. (2011). There are two main advantages to the SPDE approach. Fast computations are enabled through a Markovian approximation and sparse matrix techniques, and flexible spatial models can also be constructed from its formulation. It completely bypasses the step of explicitly specifying a covariance function, which for many complex models can be a difficult task. Markov approximation of the spatial field enables efficient matrix operations due to the sparse structures of the precision matrix. The link between GRF and GMRF is the SPDE and a finite element representation of its elements. This approach addresses the challenges like inverting a dense $n \times n$ matrix with large n with finite element approximation (FEM), which is a numerical approximation for the solution of partial differential equations (Brenner and Scott, 2008).

4 Simulation Study

A simulation study is carried out to explore the importance of considering the missing process in the modeling. In this simulation study we sample from a SGLM model on the form

$$y(s) = \beta_0 + \beta_1 u + x(s) + \varepsilon; \quad s \in \mathcal{R}^2, \quad (4.1)$$

where $x(s)$ is a stationary Matérn GRF with parameters κ and τ , β_0 is a fixed effect and $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ is white noise. The parameter β_1 is a linear fixed effect of a covariate u , which could be e.g. latitude or distance from the coast. This gives $y(s) \sim N(\eta(s), \sigma_\varepsilon^2)$ with $\eta(s) = \beta_0 + \beta_1 u + x(s)$. We simulate from $y(s)$ at n locations, s_1, \dots, s_n . Further, missing values are created according to a Binomial model such that the probability that an observation at location s_i is missing, is given by $\pi_i = P(m_i = 1 | \phi)$ with

$$\text{logit}(\pi_i) = \alpha_1 x(s_i), \quad (4.2)$$

where ϕ is the parameter vector of the missing process, in this case $\phi = (\kappa, \tau, \alpha_1)$. For a Bayesian approach the following priors are used for the parameters:

$$\begin{aligned} \beta_0 &\sim N(0, \infty) & \beta_1 &\sim N(0, 0.001^{-1}) \\ \alpha_1 &\sim N(0, 0.001^{-1}) & \log(\sigma_\varepsilon^{-2}) &\sim \text{logGamma}(1, 5 \times 10^{-5}). \end{aligned}$$

These priors are simply the default priors in the R package `R-INLA`. For the Matérn parameters κ and τ , the prior suggested in Fuglstad et al. (2015) is used. This is a joint prior to the spatial range r and the marginal variance σ^2 . In order to use this prior, the Matérn GRF is reparametrized by r and σ^2 , and two priors are specified

by

$$P(r < r_0) = p_r, \quad P(\sigma > \sigma_0) = p_\sigma, \quad (4.3)$$

where r_0 , σ_0 , p_r and p_σ are quantiles and probabilities determined by the user. A set of 300 simulated datasets is created from the suggested SGLM model. For each dataset, the data are simulated at $n = 800$ locations shown in Figure 1, and data were removed according to the model (4). To compare the joint model and the MAR model, we compare the posterior estimates of the model parameters in terms of bias and coverage of a 95 % prediction interval. The posterior mean is used as the point estimate. In particular, the estimates of the parameters β_0 and σ^2 is of interest. The intercept β_0 provides information about the mean value of the process under study, while σ^2 provides information about the spatial variability. The mean and the variance are important signatures of a spatial process, and we are interested in exploring if the MAR model and the joint model are able to capture these properties properly. Parameter estimation was performed for various amount of missingness by considering three different values of α_1 in the simulations: $\alpha_1 \in (0, 0.3, 0.8)$. Note that $\alpha_1 = 0$ corresponds to the MAR case and the probability of being missed is the same for each position in the spatial domain. Further, we also consider three different values of the marginal standard deviation, $\sigma \in (2, 4, 6, 7)$. The remaining parameters were set to $\beta_0 = 15$, $\beta_1 = -0.08$, $\sigma_\varepsilon^2 = 1$ and $r = 200$. For r and σ , the informative prior in (5) is used with $r_0 = 50$, $\sigma_0 = 10$ and $p_r = p_\sigma = 0.1$.

To compare the prediction accuracy of the joint model and the MAR model, $n = 300$ new datasets were made with one set of parameters, $(\beta_0, \beta_1, \alpha_1, \sigma, \sigma_\varepsilon^2, r) = (15, -0.01, 0.8, 7, 200)$. Out of $n = 800$ locations, $n_t = 100$ random locations were drawn for each dataset and left out as a testing set. The remaining observations were used to predict the values at these locations, and the predictive performance of each dataset was evaluated by the Root-mean-square error given by

$$\text{RMSE} = \left[\frac{1}{n_t} \sum_{i=1}^{n_t} (y(s_i) - \hat{y}(s_i))^2 \right]^{\frac{1}{2}},$$

where $y(s_i)$ is the simulated value at location s_i and $\hat{y}(s_i)$ is the corresponding prediction, here given by the posterior mean.

Figure 2 shows the average bias in the parameter estimates for β_0 and σ^2 for the joint model and the MAR model for $\alpha_1 \in (0, 0.3, 0.8)$ and $\sigma \in (2, 4, 6, 7)$. When $\alpha_1 = 0$ the biases are approximately equal for joint and the MAR models. This result is as expected as $\alpha_1 = 0$ means that there is no association between the missing process and the spatial process under study. When $\alpha_1 = 0.3$, the biases are slightly larger when we apply the MAR model compared to the joint model. This is the case for all values of σ^2 . Also, note that both σ^2 and β_0 are underestimated when

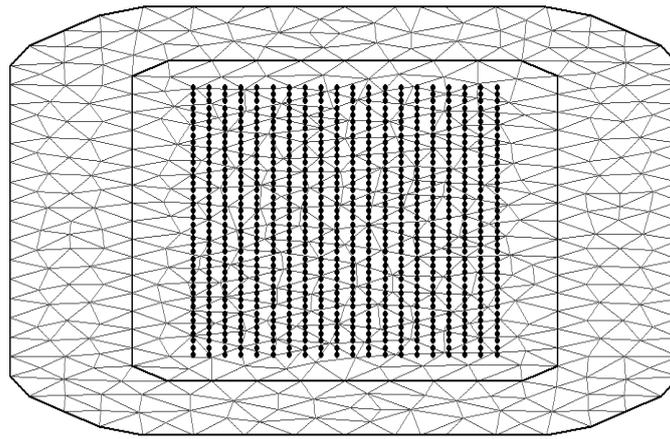


Figure 1: Mesh for the simulated data, Points denote the $n=800$ simulated locations projected on the mesh

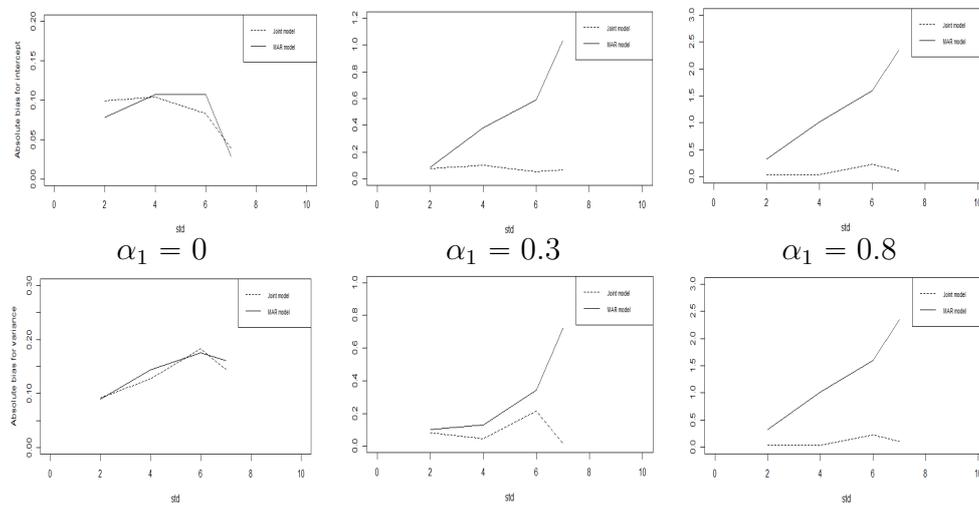


Figure 2: The variations of averaged bias in 300 iterations for β_0 (top) and σ^2 (below)

Table 1: Estimation for Joint and MAR models for simulated data with MNAR

Par	True	Model									
		Joint						MAR			
		Mean	Sd	0.25	0.5	0.975	Mean	Sd	0.025	0.5	0.975
β_0	15	15.240	1.812	11.687	15.214	18.948	13.403	1.851	9.735	13.395	17.119
β_1	-0.01	-0.009	0.0006	-0.011	-0.009	-0.008	-0.010	0.003	-0.016	-0.010	-0.004
α_1	0.8	0.828	0.076	0.683	0.826	0.982	-	-	-	-	-
$\frac{1}{\sigma^2}$	1	1.068	0.145	0.811	1.059	1.378	1.012	0.145	0.755	1.003	1.325
σ^2	6	6.364	0.869	4.934	6.263	8.332	4.598	0.490	3.752	4.555	5.674
r	200	215.950	37.092	156.947	210.975	301.733	155.459	25.683	113.570	152.383	214.025

assuming MAR. This is reasonable when $\alpha_1 > 0$. The largest values of the response have a larger probability of being missing. The MAR model is not able to capture the peaks in $y(s)$ and gives a posterior estimate of the variance σ^2 that is smaller than the actual value used for simulation. The intercept β_0 is also underestimated as the largest values tend to be missing. The biases are considerably larger for MAR compared to the joint model. Again, the MAR model systematically underestimates the true parameter values. For α_1 we also include the mean posterior estimates of all model parameters. These are presented in Table 1. The results show that the joint model is able to identify all parameters, while the MAR underestimates the intercept β_0 , the marginal variance σ^2 and the range r . In addition to estimating the model parameters, the two models are compared in terms of predictive performance. The MAR model gave an average RMSE=46.22 when the response at 100 locations was predicted for 300 simulated data sets. The joint model gave an average RMSE=22.56. Thus, the joint model outperforms the MAR model both in terms of parameter estimation and predictive performance.

The results from the simulation study demonstrate that it is important to consider the missing process in the modeling to avoid biases for the estimated model parameters. Also, predictions of the response at missing locations can be improved by taking the missing process into account. Furthermore, assuming MNAR for a process that actually is MAR does not affect the results negatively as the joint model and the MAR model perform equally well for $\alpha_1 = 0$.

5 Analyzing Missing Values of LSWT Data

In this section, we consider data from the ARC-Lake. ARC-Lake is a European space agency (ESA) funded project (www.geos.ed.ac.uk/arclake), that aim to used satellite observations to derive observations of LSWT for major lakes globally. Here we consider measurements for lake Vanern, the largest lake in Sweden. Monthly measurements of LSWT are available on a grid of 20×35 grid cells covering the lake, from 1995 to 2012. The data give mean LSWT temperature for the grid cells at the time of observation. Let $(y(s_1), \dots, y(s_n))$ be the vector of observations at spatial locations $\{s_1, \dots, s_n\}$ in a region $D \subseteq R^2$. An index for the day is suppressed as days will be analyzed separately. In this paper we study LSWT for June 1, 2008, September 1, 2008, November 1, 2008, and December 1, 2008. For a considerable amount of the lake pixels, LSWT is missing. The number of missing pixels for the four days we study are 21%, 43%, 25% and 27%, respectively. The observed SWLT for the four days together with the indication of missing data are given in Figure

Table 2: Estimates of Joint and MAR models for the first day of four different months

Month	par	Model									
		Joint					MAR				
		Mean	Sd	0.025	0.5	0.975	Mean	Sd	0.025	0.5	0.975
Jun.	β_0	18.555	0.139	18.297	18.549	18.844	17.189	0.534	16.150	17.171	18.344
	α_1	1.706	0.223	1.270	1.705	2.149	—	—	—	—	—
	$\frac{1}{\sigma_\varepsilon^2}$	3.599	0.687	2.434	3.536	5.130	2.873	0.467	2.061	2.837	3.892
	σ	1.955	0.318	1.427	1.921	2.673	1.442	0.265	1.026	1.404	2.064
	r	28.752	6.692	18.402	27.770	44.487	31.341	9.403	18.265	29.495	54.626
Sept.	β_0	14.317	0.074	14.175	14.316	14.468	13.935	0.896	11.945	13.945	15.862
	α_1	1.746	0.229	1.297	1.745	2.198	—	—	—	—	—
	$\frac{1}{\sigma_\varepsilon^2}$	69.770	12.023	49.306	68.696	96.341	70.080	11.516	50.103	69.173	95.311
	σ	0.703	0.184	0.438	0.798	1.712	0.751	0.292	0.394	0.679	1.504
	r	70.701	22.193	40.303	66.215	125.765	93.415	44.527	43.033	81.708	210.243
Nov.	β_0	6.152	0.339	5.357	6.200	6.687	7.181	3.396	-0.513	7.279	14.210
	α_1	-0.679	0.123	-0.924	-0.678	-0.438	—	—	—	—	—
	$\frac{1}{\sigma_\varepsilon^2}$	103.695	26.191	62.016	100.376	164.364	87.488	17.923	57.180	85.841	127.400
	σ	1.663	0.480	0.999	1.567	2.855	1.915	0.902	0.916	1.672	4.288
	r	107.086	34.502	61.198	99.698	193.851	138.537	72.914	61.466	118.264	331.152
Dec.	β_0	2.495	0.279	2.510	1.494	3.003	3.185	1.126	0.544	3.285	5.106
	α_1	-0.670	0.069	-0.798	-0.674	-0.524	—	—	—	—	—
	$\frac{1}{\sigma_\varepsilon^2}$	17182.127	0.1772	866.854	11747.693	64632.838	13.651	0.1584	363.60	8350.608	55948.08
	σ	4.641	1.178	3.108	4.354	7.680	3.080	0.6797	2.130	2.943	4.760
	r	50.476	14.030	32.590	46.970	86.883	35.111	8.963	22.933	33.203	57.531

3. From this visual inspection one can find that there is spatial dependency in both LSWT and missing process (Hook et al., 2012; Maccallum and Merchant, 2012).

The joint model is utilized to make inference about LSWT at lake Vanern for the first day of June, September, November and December 2008. As for the simulation study, the joint model is compared to a model where we assume that the data are MAR. In order to assess predictive performance of the joint model and the MAR model prediction has been done to reconstruct LSWT both for observed and missing locations for the first day of December 2008 and also cross validation is implemented in the prediction of observed values

In order to investigate the association between the response $y(s)$ and the missing process m at lake Vanern, we assume that the true LSWT at location s_i is given by

$$\eta(s_i) = \beta_0 + x(s_i),$$

where β_0 is an intercept and $x(s_i)$ is a Matern Gaussian Random Field with zero mean, range r and marginal variance σ^2 . The true LSWT is observed with normally distributed noise, $y(s_i) \sim N(\eta(s_i), \sigma_\varepsilon^2)$, where σ_ε^2 is the variance of the noise. Furthermore, we consider a binary missing process $m|\pi \sim Bin(1, \pi)$ and the probability that data at location s_i is missing has been modeled as $\text{logit}(\pi_i) = \alpha_1 x(s_i)$. We use

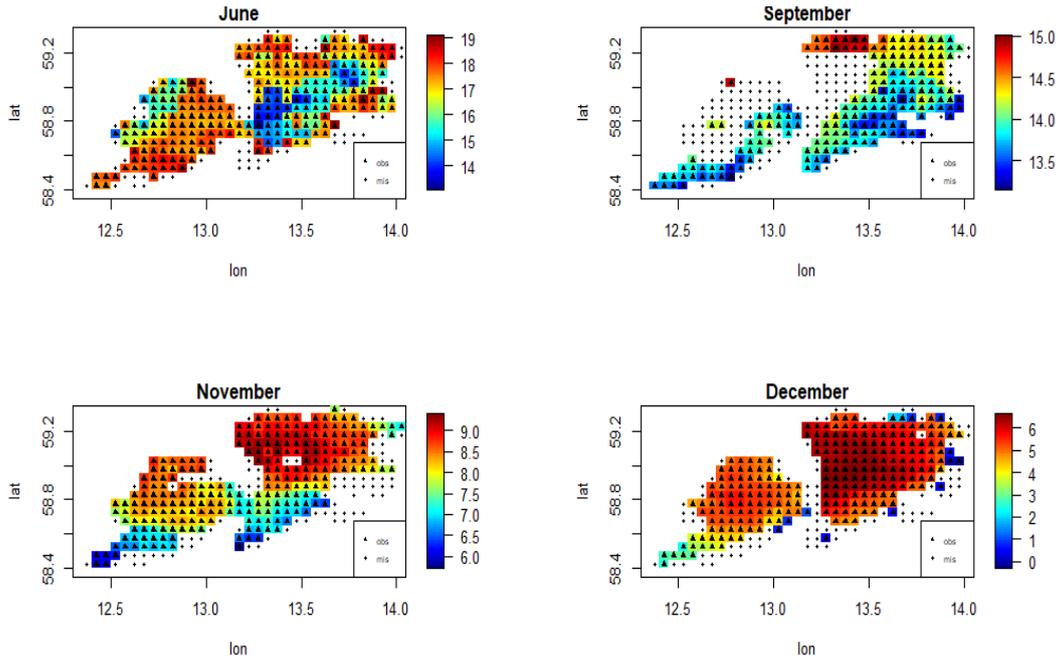


Figure 3: Image of LSWT for four different months 2008 and location of observed and missing values.

the same priors here as for the simulation study.

Relevant posterior estimates are reported for both the joint and the MAR model in Table 2. For real data, it is in particular interesting to study the parameter α_1 which indicates whether there is an association between the response and the missing process. Table 2 shows that for June and September $\alpha_1 = 1.70$ and $\alpha_1 = 1.75$ respectively, while for November and December $\alpha_1 = -0.68$ and $\alpha_1 = -0.67$. These are relatively high values indicating a clear association between the missing process and the LSWT. The α_1 is positive for the two warm months, indicating that high values of LSWT have a larger probability of being missing during summer. Conversely, for the two winter months, α_1 is negative, and small values of LSWT tend to be missing. It is not clear what kind of factors that lead to missingness in our dataset. During the summer it could be that the air is unclear when the LSWT is higher because of evaporation. For cooler months, missingness could be a result of some phenomena like storms, ice cover or clouds.

The parameter β_0 is also interesting to study as it gives the mean LSWT for the Vanern lake. Thus, β_0 gives an interpretable indication of the difference between the joint model and the MAR model. Table 2 shows that the differences in the mean LSWT between the joint model and the MAR model are 1.35, 0.382, -1.029 and -0.69 for June, September, November, and December respectively. These are relatively large differences, showing that the chosen model has a large impact on the

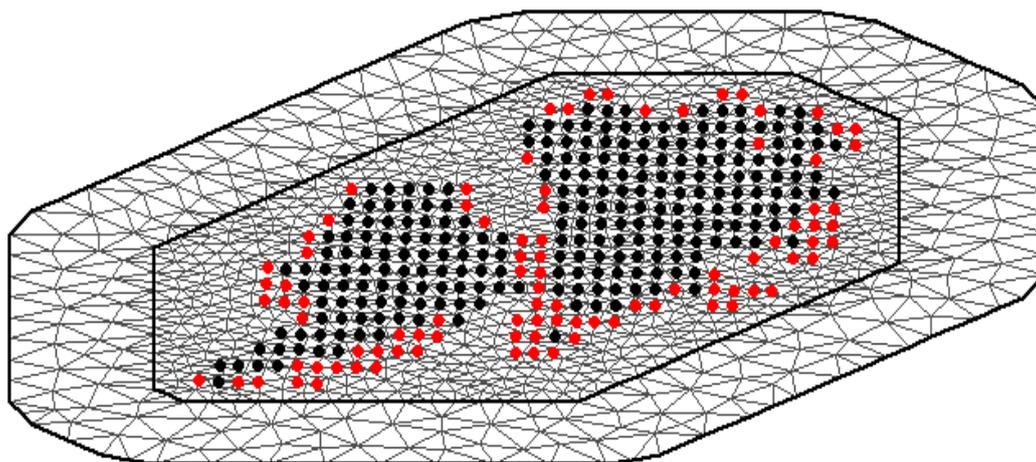


Figure 4: The location of observation (black points) and missing (red points) at Vanern

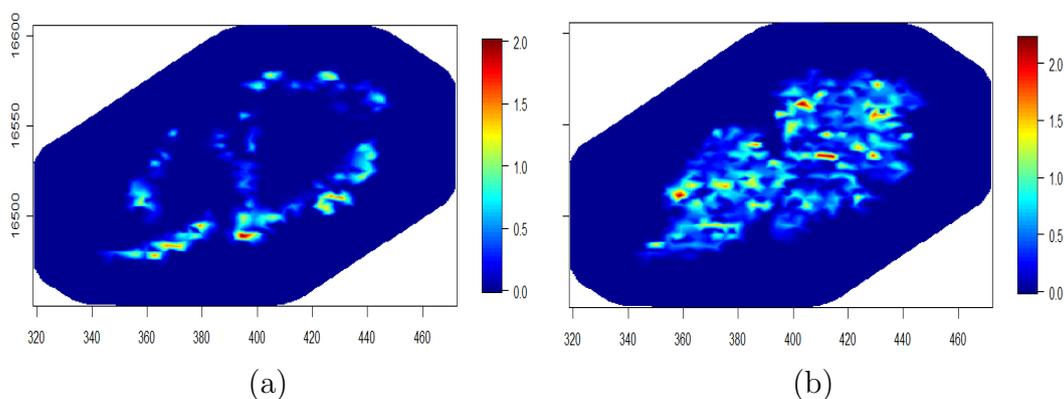


Figure 5: Posterior standard deviation in prediction of response y for a- Joint and b- MAR models

results of the analysis. The difference between the MAR and the joint model is also large for the other model parameters.

The joint model and the MAR model were also utilized for predictions in order to reconstruct LSWT at all locations. In Figure 4 locations of observed and missing values inside the mesh for the first day of December 2008, are displayed. The posterior standard deviations in the prediction of response for two models are shown in Figure 5. We note that the posterior standard deviation of the predictions from the joint model, in general, is lower than the standard deviation of the predictions from the MAR model. Also, leave-one-out cross validation in the prediction of just observed values is implemented and mean square prediction error (MSPE) is

computed for each model as

$$\text{MSPE} = \frac{1}{n_{obs}} \sum_{i=1}^{n_{obs}} (y_i - \hat{y}_{-i})^2,$$

where n_{obs} is the number of observed values, y_i is the i th observed value and \hat{y}_{-i} is the predicted value computed after deleting the i th observed value. The MSPEs for joint and MAR model is equal to 0.645 and 1.762, respectively. These values approve the better performance of the joint model too.

Conclusion

In this paper, the joint model idea for spatial data was implemented and through simulation studies, it was revealed that joint modeling of measurement and missing processes work well in both ignorable and non-ignorable missingness. Simulation studies approved the claim that if we apply the information include in missing data when the missingness is not at random via joint modeling idea the results are more reliable.

References

- Brenner, S. and Scott, R. (2008), *The Mathematical Theory of Finite Element Methods*, Springer Science Business Media.
- Breslow, N. E. and Clayton, D. G. (1993), Approximate Inference in Generalized Linear Mixed Models, *Journal of the American Statistical Association*, **88**, 9-25.
- Diggle, P., Tawn, J. A. and Moyeed, R. A. (1998), Model-Based Geostatistic. *Journal of the Royal Statistical Society*, **47**, 299-350.
- Eidsvik, J., Rue, H. and Martino, S. (2009), Approximate Bayesian Inference in Spatial Generalized Linear Mixed Models, *Scandinavian Journal of Statistics*, **36**, 1-22.
- Fuglstad, G., Simpson, D., Lingren, F. and Rue, H. (2015), Constructing Priors that Penalize the Complexity of Gaussian Random Fields, *Journal of American Statistical Association*, **6**, 1-8.
- Holand, A., Steinsland, I., Martino, S. and Jensen, H. (2010), Animal Models and Integrated Nested Laplas Approximations, Preprint Statistics.
- Hook, S., Wilson, R. C., Maccallum, S. and Merchant, C. J. (2012), Lake Surface Temperature, *Journal of American Meteorological Society*, **93**, S18-S19.
- Karimi, O. and Mohammadzadeh, M., Bayesian spatial regression models with closed skew normal correlated errors and missing observations, *Statistical Papers*, **53**, 205-218 (2010).

Lindgren, F., Rue, H. and Lindstr, O. (2011), An Explicit Link Between Gaussian Fields and Gaussian Markov Random Fields: The Stochastic Partial Differential Equation Approach, *Journal of the Royal Statistical Society*, **73**, 423-498.

Little, R. J. A. and Rubin, D. B. (2002), *Statistical Analysis with Missing Data*, Wiley, New York.

Maccallum, S. N. and Merchant, C. J. (2012), Surface Water Temperature Observations of Large Lakes by Optimal Estimation, *Journal of Canadian Aeronautics and Space*, **38**, 25-45.

Molenberghs, G., Fitzmaurice, G., Kenward, M. G., Tsiatis, A. and Verbeke, G. (2015). Handbook of missing data methodology, CRC Press.

Rubin, D. (1976), Inference and Missing Data, *Biometrika*, **63**, 581-590.

Rue, H. and Martino, S. (2007), Approximated Bayesian Inference for Hierarchical Gaussian Markov Field Models, *Journal of Statistical Planning and Inference*, **10**, 3177-3192.

Steinsland, I., Thorrud Larsen, C., Roulin, A. and Jensen, H. (2014), Quantitative Genetic Modelling And Inference in the Presence Non-Ignorable Missing Data, *International Journal of Organic Evolution*, **68**, 1735-1747.



Modeling of Spatially Clustered Survival HIV/AIDS Data in the Presence of Competing Risks Setting: A Bayesian Approach

Somayeh Momenyan¹, Amir Kavousi^{2*}

¹Department of Biostatistics, Shahid Beheshti University of Medical Sciences, Tehran,
Iran.

²Department of Epidemiology, Shahid Beheshti University of Medical Sciences, Tehran,
Iran.

Abstract:

In some applications, the clustered survival data are arranged spatially such as clinical centers or geographical regions. Competing risks in survival data concern a situation where there is more than one cause of failure, but only the occurrence of the first one is observable. In this paper, we considered a Bayesian hierarchical survival model in the setting of competing risks for the spatially clustered HIV/AIDS data. In this model, a Weibull parametric distribution with the spatial random effects in the form of the county-failure type-level was used. A multivariate intrinsic conditional autoregressive distribution was employed to model the areal spatial random effects. We illustrated the gains of our model through the simulation studies and application to the HIV/AIDS data.

Keywords: Survival data, Clustered data, Spatial HIV/AIDS.

Mathematics Subject Classification (2010): 62H11, 62N01.

1 Introduction

In biomedical studies it is common to have time to event data so that the event of interest is usually death, giving rise to the survival analysis. In the survival analysis,

*Speaker: kavvousi-am@yahoo.com

in many situations, there are some potential risk factors that are immeasurable or unobservable. For solving this problem, Vaupel et al. introduced a model with the univariate random effect that corresponds to a health status of the stratum [Vaupel, Manton and Stallard \(1979\)](#). But, in some situations, the survival data are dependent such as when a sample of individuals was grouped into clusters such as clinical centers, geographic regions, and so on. If individuals come from different regions, there will be a spatial correlation between survival data because the geographically closer regions usually are the same or similar in terms of the environmental and social factors. Ignoring this spatial correlation results in biased estimates and misleading inferences. Moreover, the spatial survival analysis by mapping the spatial distribution could identify some of the geographical inequalities that exist in survival. In the latter paper, Banerjee et al. also compared the geostatistical and areal approaches and showed that the geostatistical frailty model is time-consuming and it produces results that differ little from the areal frailty model. The survival model for capturing spatiotemporal variation in the survival data was investigated by Banerjee and Carlin [Banerjee, Wall and Carlin \(2003\)](#) and Hanson et al [Hanson, Jara and Zhao \(2012\)](#). Pan et al. proposed a spatial Bayesian semiparametric model to analyze interval-censored survival data [Pan, Cai. et al. \(2014\)](#). More recently, Cramb et al. proposed a spatial flexible parametric relative survival model [Cramb, Mengersen, Lambert and et \(2016\)](#), and Zhou and Hanson applied a spatial semi-parametric survival model to arbitrarily censored survival data [Zhou and Hanson \(2018\)](#). In the survival data, there is also a situation where the time from the starting point to an interesting event may not be observable, because of the incidence of another, so-called competing event. Thus, in competing risks data, there is more than one cause of failure, but only the occurrence of the first one is observable. To the best of our knowledge, no survival model for the spatially correlated survival data in the presence of competing risks setting has been applied. Thus, the new aspect of this paper is the extension of survival model from the single failure type to the competing risks in the spatially clustered HIV/AIDS data.

2 The HIV/AIDS Data

The data were from a retrospective cohort study, which was conducted in Hamadan Province, the central-western part of Iran, from 1997 to 2011. All 585 HIV-positive people were included in this study. The explanatory variables included were as follows: age, sex, marital status, method of transmission, co-infection with tuberculosis and patients county of residence. Also, date of HIV diagnosis, date of progression

to AIDS and date of death were collected. The main outcome in this study was the time interval, between HIV diagnosis and AIDS progression, so the event of interest was AIDS progression. However, some of the HIV infected patients die before AIDS progression. Thus, death before AIDS was considered as competing risk because it prevents AIDS progression.

3 Model Formulation

In this section, we suppose n individuals under study are from K counties. The number of individuals in a sample in each county is n_k where $k = 1, \dots, K$ and $\sum_{k=1}^K n_k = n$. Also, let $C_{ik} = (T_{ik}, \delta_{ik})$ be the competing risks data for the i -th individual living in the k -th county, $i = 1, \dots, n_k$, where T_{ik} denotes the time to an event which may be right censored time. In the setting of competing risks, each individual could experience one of the G possible failure types during follow-up or could be right censored. Hence, failure type indicator or δ_{ik} takes value from $\{0, 1, \dots, G\}$.

3.1 Survival Model with Random Effects

To estimate hazards for each failure type, we postulate the cause-specific proportional hazards model:

$$h_{ik}^g(t | X_{ik}) = \lim_{dt \rightarrow 0} \frac{P(t \leq T_{ik} < t + dt, \delta_{ik} = g | T_{ik} \geq t, X_{ik})}{dt} = h_0(t)^g \exp(X_{ik}^T \beta^g),$$

where $h_0(t)^g$ is an unspecified baseline risk function and β^g is a $p \times 1$ vector of regression parameters associated to a $p \times 1$ vector of observed explanatory covariates X_{ik} this model, to estimate the hazard of each failure type, other competing risks are assumed as censored in addition to those who are censored from loss to follow-up. But, sometimes those who experience competing risks events are censored informatively. To avoid biased results, we include the random effects $V_{ik} = (V_{ik}^1, \dots, V_{ik}^G)^T$, to account for the association between times of different failure types.

$$h_{ik}^g(t | X_{ik}, V_{ik}^g) = h_0(t)^g \exp(X_{ik}^T \beta^g + V_{ik}^g), \quad (3.1)$$

On the other hand, since subjects coming from different counties, we consider a survival model with the spatial random effects. Let W_K^g , denotes the spatial effects of latent risk factors for the g -th type of failure, nested within the k -th county. We introduce again, these random effects through 3.1, as

$$h_{ik}^g(t | X_{ik}, V_{ik}^g, W_K^g) = h_0(t)^g \exp(X_{ik}^T \beta^g + V_{ik}^g + W_K^g),$$

Alternatively, for the baseline hazard function, we assume a cause-specific Weibull function.

3.2 Spatial Random Effects

In this paper, we use the multivariate intrinsic conditionally autoregressive (MCAR) distribution for the spatial random effects in the HIV/AIDS data. Let In this paper, we use the MCAR distribution for the spatial random effects in the HIV/AIDS data. Let $W_k = (W_k^1, \dots, W_k^G)^T$ being a G -dimensional vector of the spatial random effects collection for the G possible failure types within the k -th county. Following Mardia (1988) Mardia (1988), the joint distributions for W take the form

$$W \sim N(0, (\Sigma \otimes \Lambda)^{-1}),$$

where $\Sigma = (D - \alpha B_w)$ is a $K \times K$ matrix and therefor $\Sigma \otimes \Lambda$ is a $KG \times KG$ matrix.

3.3 Likelihood

We assume for $V_{ik}^g; i = 1, \dots, n_k, k = 1, \dots, K, g = 1, \dots, G$ a multivariate normal distribution, that is, $V_{ik} \sim N(0, \Lambda)$ where Λ is a $G \times G$ variance-covariance matrix. Thus, in our model, we are using the same Λ to model the MCAR and failure type random effects distributions. Let $(C_{ik}, X_{ik}; i = 1, \dots, n_k, k = 1, \dots, K)$ represents the observed data; therefore, the likelihood function for this model is as follows

$$\prod_{k=1}^K \prod_{i=1}^{n_k} \prod_{g=1}^G [[\rho^g t^{\rho^g - 1} \exp(X_{ik}^T \beta^g + V_{ik}^g + W_k^g)]^{I(\delta_{ik}=g)} \\ \times \exp[- \int_0^{t_{ik}} \rho^g t^{\rho^g - 1} \exp(X_{ik}^T \beta^g + V_{ik}^g + W_k^g) dt]]$$

3.4 Bayesian Approach

The standard non-informative prior distributions for the parameters were considered as follows: for the regression coefficients, β^g , and the Weibull shape parameters, $\rho^g; g = 1, \dots, G$ a multivariate normal, $N(0, \Sigma_\beta)$, and a gamma, $\zeta(a_1, b_1)$, priors were taken respectively. Concerning the spatial random effects, we used the MCAR prior distribution, $W \sim \text{MCAR}(1, \Lambda)$. For the variance-covariance matrix, Λ , an inverse Wishart, $jW(R, k)$, was used.

4 Simulation Study

We generated a total of 100 simulated datasets with three levels of sample size in each county ($n_k = 50, n_k = 100$ and $n_k = 200$) and under three levels of censoring

rate (low, 20%, medium, 40%, and high, 60%). For each dataset, the neighborhood structure was based on the nine counties in Hamadan, Iran. We used a continuous covariate (X_1) and a binary categorical covariate (X_2). For each dataset, we simulated two competing risks, risk 1 and 2, from the cause-specific exponential model. The event time was generated by the overall hazard rate that at each time point is the sum of the cause-specific hazard rates for two types of event, $\lambda_{ik}(t) = \lambda_{ik}^1(t) + \lambda_{ik}^2(t)$. Then the type of event was determined by a Bernoulli experiment with the probability $P_1 = \lambda^1/\lambda$ for the event of type 1 and $P_2 = \lambda^2/\lambda$ for the event of type 2. Finally, for assignment of the event type, a random number R generated from a continuous Uniform distribution on $[0, 1]$ and the event of type 1 was assigned, if $R < P_1$, and the event of type 2, else. The random effects were generated from a multivariate normal with a mean 0 and a 2×2 covariance matrix Λ , that $\Lambda_{11} = 0.1$, $\Lambda_{22} = 0.1$ and $\Lambda_{12} = 0.05$, and then were centered around its mean. Also, the spatial random effects W were generated from the MCAR distribution. The corresponding regression coefficients were set $\beta_1^1 = 1$, $\beta_2^1 = 1$ for risk 1 and $\beta_1^2 = 1$, $\beta_2^2 = 1$ for risk 2. The data above a threshold were right censored which were selected as the $(1 - \alpha)$ -quantile of the sample survival times, such that 100 $\alpha\%$ of the observations were censored.

4.1 Simulation Results

The results of the first simulation study from nine scenarios are reported in Table 1. While estimation of the regression coefficients in the total scenario were close to the truth value 1, but estimation of the covariance matrix Υ had a consistently small bias for the large sample size ($n_k = 200$) and with censoring rate 20% and 40%. Also, all parameters had the coverage probabilities close to the nominal level of 0.95 under all the scenarios. The MSE criterion for all parameters was close together and as sample size increases, the estimation accuracy of parameters increases.

5 Analysis of the HIV/AIDS Data

We analyzed the HIV/AIDS data using proposed model. The summary measures of all parameters of the fourth model were presented in Table 2. We also mapped the summaries of our results. Figure 1 shows 2 maps that represent the posterior spatial relative risk in nine counties of Hamadan Province for the relative risk of AIDS progression and or the relative risk of mortality post-HIV infection. The posterior estimates of county-specific random effects were recorded based on the quintile of their distribution for showing the spatial inequalities on the map. As

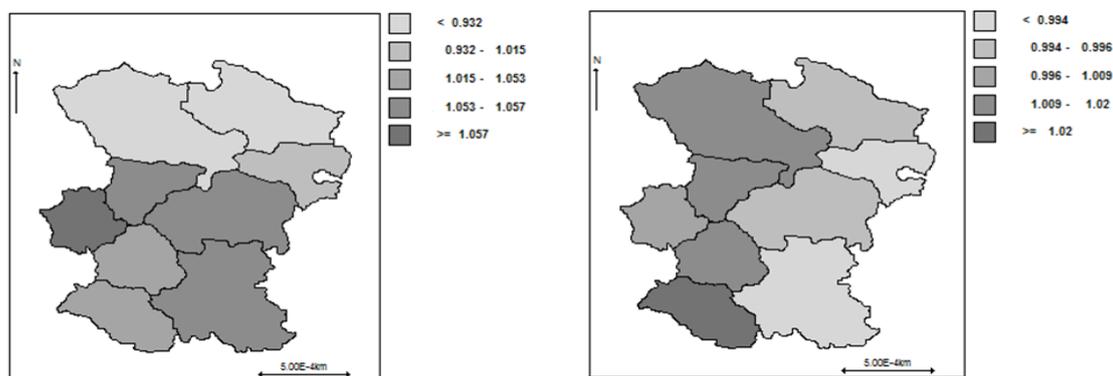


Figure 1: Maps of the spatial relative hazard for ADS progression (left) and mortality post-HIV infection (right) based on the mode.

shown in Figure 1, for the risk of AIDS progression, one cluster of counties was identified with higher risk in the south, southeast, and southwest regions and one cluster with lower risk in the north, northeast, and northwest regions was identified. Also, for the risk of mortality post-HIV infection, the high-risk cluster consists of the northwest, west, and southwest regions and the low-risk cluster consists of the northeast, east, and southeast regions. Furthermore, the values of spatial random effects for the risks of AIDS progression and mortality post-HIV infection were in ranges $(-0.093, 0.048)$ and $(-0.017, 0.028)$, respectively. Such small values of the spatial random effects suggest that regional differences had a small effect.

Table 1: Estimation results for the simulation study

Rate	Par.	$n_k = 50, n = 450$				$n_k = 100, n = 900$				$n_k = 200, n = 1800$			
		Estimate	Bias	MSE	CP	Estimate	Bias	MSE	CP	Estimate	Bias	MSE	CP
20%	$\beta_1^1 = 1$	1.028	0.028	0.018	96	0.995	-0.004	0.006	96	0.995	-0.004	0.003	97
	$\beta_1^2 = 1$	0.997	-0.002	0.018	93	0.992	-0.007	0.007	97	0.985	-0.014	0.005	95
	$\beta_2^1 = 1$	0.962	-0.037	0.022	93	0.938	-0.061	0.015	97	0.983	-0.016	0.005	98
	$\beta_2^2 = 1$	1.019	0.019	0.028	93	0.959	-0.040	0.013	97	0.975	-0.024	0.005	97
	$\Lambda_{11} = 0.1$	0.082	-0.017	0.001	98	0.087	-0.012	0.002	98	0.109	0.009	0.002	100
	$\Lambda_{22} = 0.1$	0.084	-0.015	0.003	97	0.0844	-0.015	0.001	96	0.091	-0.008	0.001	100
	$\Lambda_{12} = 0.05$	0.024	-0.026	0.002	96	0.033	-0.017	0.0009	98	0.055	0.005	0.0007	98
40%	$\beta_1^1 = 1$	1.036	0.036	0.024	95	0.988	-0.011	0.009	96	0.995	-0.004	0.003	96
	$\beta_1^2 = 1$	1.005	0.005	0.018	98	0.991	-0.008	0.009	95	0.988	-0.011	0.004	96
	$\beta_2^1 = 1$	0.977	-0.022	0.028	99	0.932	-0.067	0.021	94	1.045	0.045	0.008	96
	$\beta_2^2 = 1$	1.044	0.044	0.046	93	0.974	-0.025	0.017	98	1.021	0.021	0.009	92
	$\Lambda_{11} = 0.1$	0.082	-0.017	0.001	98	0.090	-0.009	0.002	99	0.102	0.002	0.002	100
	$\Lambda_{22} = 0.1$	0.084	-0.015	0.001	100	0.085	-0.014	0.002	99	0.088	-0.011	0.001	98
	$\Lambda_{12} = 0.05$	0.023	-0.027	0.002	94	0.033	-0.017	0.001	95	0.048	-0.002	0.001	99
60%	$\beta_1^1 = 1$	1.042	0.042	0.030	95	0.985	-0.014	0.013	95	0.993	-0.006	0.005	98
	$\beta_1^2 = 1$	1.002	0.002	0.025	97	0.996	-0.003	0.015	92	0.983	-0.016	0.006	93
	$\beta_2^1 = 1$	0.967	-0.032	0.055	97	0.938	-0.061	0.026	96	1.005	0.005	0.012	96
	$\beta_2^2 = 1$	1.049	0.049	0.071	93	0.969	-0.030	0.030	96	0.995	-0.004	0.015	90
	$\Lambda_{11} = 0.1$	0.081	-0.018	0.002	97	0.098	-0.001	0.002	99	0.087	-0.012	0.002	100
	$\Lambda_{22} = 0.1$	0.088	-0.011	0.002	95	0.086	-0.013	0.002	99	0.073	-0.026	0.001	100
	$\Lambda_{12} = 0.05$	0.023	-0.027	0.002	94	0.026	-0.024	0.001	96	0.030	-0.020	0.001	90

Table 2: Posterior estimation results of model

Variable	Category	Number (%)	AIDS progression		Mortality post-HIV infection	
			Mean (SD)	Hazard ratio (95% credible interval)	Mean (SD)	Hazard ratio (95% credible interval)
Intercept			-3.872(0.35)	0.022 (0.009, 0.039)	-3.336(0.30)	0.037 (0.018, 0.062)
Gender	Male	521 (89.05)		Reference		Reference
	Female	64 (10.94)	0.054 (0.39)	1.14 (0.496, 2.291)	-1.759 (0.94)	0.253 (0.021, 0.844)
Marital Status	Single	268 (45.81)		Reference		Reference
	Married	223 (38.11)	0.052 (0.21)	1.079 (0.689, 1.610)	0.256 (0.20)	1.318 (0.872, 1.914)
	Divorced	75 (12.82)	0.293 (0.27)	1.394 (0.769, 2.275)	0.285 (0.28)	1.384 (0.748, 2.265)
	Widowed	19 (3.24)	0.762 (0.44)	2.357 (0.845, 4.817)	0.028 (0.64)	1.246 (0.247, 3.139)
Age	1-24	70 (11.96)		Reference		Reference
	25-44	457 (78.11)	0.329 (0.34)	1.478 (0.741, 2.894)	0.049 (0.29)	1.099 (0.599, 1.912)
	45-74	57 (9.74)	0.824 (0.44)	2.517 (0.967, 5.554)	0.775 (0.37)	2.331 (1.050, 4.555)
Tuberculosis	No	564 (96.41)		Reference		Reference
	Yes	21 (3.58)	1.705 (0.28)	5.723 (3.077, 9.338)	-0.789 (0.81)	0.601 (0.074, 1.696)
Transmission	Injection	475 (81.19)		Reference		Reference
	Sexual	72 (12.30)	0.256 (0.36)	1.382 (0.614, 2.58)	-0.584 (0.56)	0.645 (0.167, 1.496)
	Mother	9 (1.53)	2.037(0.60)	9.214 (2.296, 25.06)	ND*	ND
	Injection use/sexual	26 (4.44)	0.637 (0.36)	2.022 (0.886, 3.74)	-1.121 (0.80)	0.428 (0.052, 1.217)
$\rho^1 = \rho^2$			1.35 (0.065)			
Λ_{11}			0.029 (0.035)			
Λ_{22}			0.024 (0.019)			
$Corr = \frac{\Lambda_{12}}{\sqrt{\Lambda_{11}}\sqrt{\Lambda_{22}}}$			0.041 (0.038)			

*ND: No Data

6 Discussion

In this paper, we proposed a model for the spatially clustered survival HIV/AIDS data in the presence of competing risks setting. The data were from Hamadan Province, Iran, from 1997 to 2011. For these data, we used a parametric proportional hazard model in a Bayesian setting. In the field of HIV/AIDS disease, understanding of the geographic variation of the risks of AIDS progression and mortality post-HIV infection provides greater opportunity to identify high-burden areas, since they reect both diagnostic and patient management. From the results of the present study, the low-risk cluster in risk of AIDS progression contained counties with lowest population density and the high-risk cluster in risk of AIDS progression consists of some counties with the highest rate of population density. Also, the low-risk cluster was in the remote areas and with much distance from the most populous counties in Hamadan Province. It is worth mentioning that by fitting our proposed model in the HIV/AIDS data the spatial random effects were small values, but the length of 95% credible interval for the covariates was decreased compared to the other models (not shown). Also, our results showed a small positive correlation between the hazards of AIDS progression and mortality post-HIV infection.

References

- Banerjee, S., Wall, MM. and Carlin, BP. (2003), Frailty Modeling for Spatially Correlated Survival Data, with Application to Infant Mortality in Minnesota, *Biostatistics*, **4**, 123-142.
- Cramb, SM., Mengersen, KL., Lambert, PC. and et, al. (2016), A Flexible Parametric Approach to Examining Spatial Variation in Relative Survival, *Statistics in Medicine*, **35**, 5448-5463.
- Hanson, TE., Jara, A. and Zhao, L. (2012), A Bayesian Semiparametric Temporally-Stratified Proportional Hazards Model with Spatial Frailties, *Bayesian Analysis*, **7**, 147-188.
- Mardia, K. (1988), Multi-Dimensional Multivariate Gaussian Markov Random Fields with Application to Image Processing, *Journal of Multivariate Analysis*, **24**, 265-284.
- Pan, C., Cai, B., Cai, L. and et, al. (2014), Bayesian Semiparametric Model for Spatially Correlated Interval-Censored Survival Data, *Computational Statistics a Data Analysis*, **74**, 198-208.
- Vaupel, JW., Manton, KG. and Stallard, E. (1979), The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality, *Demography*, **16**, 439-454.
- Zhou, H., Hanson, T. (2018), A Unified Framework for Fitting Bayesian Semiparametric Models to Arbitrarily Censored Survival Data, *Journal of the American Statistical Association*, **113**, 5448-5463.



On Spatial Skew-Gaussian Process and Its Applications in Survival Analysis

Kiomars Motarjem*, Mohsen Mohammadzadeh
Department of Statistics, Tarbiat Modares University, Tehran, Iran.

Abstract:

Considering a Gaussian random field for the spatial random effects in survival analysis sometimes may not correspond to reality. In this paper, by considering a Spatial Skew Gaussian process for random effects we propose a new class of spatial survival models. In a simulation study, we have shown that the deviation from Gaussian assumption about the random effects have an undesirable effect on the estimation of model parameters, whereas the use of spatial skew Gaussian random effects provides more accurate models.

Mathematics Subject Classification (2010): 62H11, 62N01, 62N86.

1 Introduction

In survival models, it is usually assumed that the failure times of the subjects are independent. However, in many cases and applications, this assumption is not realistic, and failure times are spatially correlated. Random effects are usually latent components of survival data, that can be determined by recognizing the spatial correlation and considered through a spatial survival model. Most spatial survival models introduced by researchers are suitable for lattice data in which the spatial correlation exists between the areas containing survival data. But in geostatistical cases, the analysis of survival models involves complicated parameter estimation. [Motarjem et al. \(2017\)](#) introduced a spatial survival model for analyzing geostatistical survival data, where a Gaussian random field is used for considering the spatial random effects. This paper is organized as follows. In Section 2, the multivariate

*Speaker: K.motarjem@modares.ac.ir

skew normal distribution and the SSG random fields are introduced. In Section 3, we first recall the basic framework of spatial statistics and then present the Spatial Skew Gaussian (SSG) processes. In Section 4, Using simulation study the Cox proportional hazards, frailty, and spatial survival model are discussed and a new class of spatial survival model with SSG random effect is proposed. Finally, results are discussed in the last section.

2 Multivariate Skew Normal Distribution

Multivariate skew-normal distributions are based on the normal distribution, but skewness is added to extend the applicability of the normal distribution while trying to keep most of the interesting properties of the Gaussian distribution. In a Gaussian framework, spatial data are analyzed using the skew normal distribution [Kim and Mallick \(2002\)](#), but without a precise definition of skew normal spatial process, will be shown in Section 3, this model leads to a very small amount of skewness and therefore is not very useful in practice. In this work, we use the multivariate closed skew-normal distribution [Dominguez et al. \(2007\)](#), [Gonzalez et al. \(2004\)](#). It stems from the classical skew-normal distribution introduced by Azzalini and its co-authors. A drawback is that notations can become cumbersome. It has the advantages of being more general and having more properties similar to the normal distribution than any other skew-normal distributions

An n -dimensional random vector Y has a multivariate closed skew Normal distribution was introduced by [Gonzalez et al. \(2004\)](#), denoted by $CSN_{n,m}(\mu, \Sigma, D, \gamma, \Delta)$, if it has the density function as

$$\begin{aligned} f(y; \mu, \Sigma, \gamma, \Delta, D) &= C_m \phi_n(Y; \mu, \Sigma) \Phi_m(D^T(Y - \mu); \gamma, \Delta) \\ C_m^{-1} &= \Phi_m(0, \gamma, \Delta + D^T \Sigma D) \end{aligned} \quad (2.1)$$

where $\mu \in R^n$, $\gamma \in R^m$, $\Sigma \in R^{n \times n}$ and $\Delta \in R^{m \times m}$ are both covariance matrices, $D \in R^{n \times m}$ is skewness matrix, $\phi(Y; \mu, \Sigma)$ and $\Phi_n(Y; \mu, \Sigma)$ are the n -dimensional normal pdf and cdf with mean μ and covariance matrix Σ . In the multivariate closed skew Normal distribution with the density function (2.1), if $D = 0$, it reduces to the multivariate normal and for $m = 1$, it is equal to Azzalini's density [Azzalini and Dalla Valle \(1996\)](#). The $CSN_{n,m}(\mu, \Sigma, D, \gamma, \Delta)$ distribution that defined by (2.1) is generated from the following bivariate vector

$$\begin{pmatrix} U \\ Z \end{pmatrix} \sim N_{m+n} \left(\begin{pmatrix} \gamma \\ 0 \end{pmatrix}, \begin{pmatrix} \Delta + D^T \Sigma D & -D^T \Sigma \\ -\Sigma D & \Sigma \end{pmatrix} \right), \quad (2.2)$$

where U and Z are the Gaussian vectors of dimensions m and n , respectively. The

multivariate closed skew Normal distribution is obtained by combining two random bivariate vectors. Given the combination above, it is easy to show that, condition on $U \leq 0$ the random vector $\mu + [Z|U \leq 0]$ is distributed as (2.1). Note that the inequality $U \leq 0$ corresponds to $U_i \leq 0$ for all $i = 1, \dots, m$. Also, it is well known that the conditional random vector $[Z|U]$ is also a Gaussian vector with the distribution

$$[Z|U] \sim N_n(-\Sigma D(\Delta + D^T \Sigma D)^{-1}(U - \nu), \Sigma - \Sigma D(\Delta + D^T \Sigma D)^{-1} D^T \Sigma). \quad (2.3)$$

This property provides a two-step algorithm for simulating a *CSN* vector Z : (i) generate samples of the Gaussian vector $U \stackrel{d}{=} N_m(N, \Delta + D^T \Sigma D)$ such that $U \leq 0$; (ii) generate samples of the Gaussian vector $[Z|U]$ according to (2.3). Note that generating a vector U provided that $U \leq 0$ is not straightforward. In particular direct sequential simulation procedures can not be implemented. MCMC methods must be used instead. In this paper, we choose a Gibbs sampling approach to simulate the vector $[U|U \leq 0]$.

3 Spatial Skew Gaussian Process

Let $Z(s)$ with $s \in R^2$ be a spatial, stationary, zero-mean Gaussian random field with variogram

$$2\gamma_Z(h) = \text{Var}(Z(s+h) - Z(s)); \quad h \in R^2,$$

where $\text{Var}(Z(s)) = \sigma^2$, for more details on the variogram, refer to [Cressie \(1993\)](#). The covariance matrix of $Z = (Z(s_1), \dots, Z(s_n))^T$ is conducted from $C_Z(h) = \sigma^2 - \gamma_Z(h)$ and is denoted by Σ . For linking this structure with the skew Normal distribution (2.2), the *CSN* random field $Y(s)$ is defined as

$$Y(s) \stackrel{d}{=} \mu + [Z(s)|U \leq 0], \quad (3.1)$$

where for each n -dimension vector $Z = (Z(s_1), \dots, Z(s_n))^T$, we have

$$Y \stackrel{d}{=} \mu + [Z|U \leq 0],$$

in which $U \sim N_m(0, \Delta + D^T \Sigma D)$ and Z has the distribution given in (2.2).

In practice, observations from $(Y(s_1), \dots, Y(s_n))^T$, U and Z are not available, but Y can be obtained from two independent random variables U and V as

$$Y = \begin{pmatrix} U \\ V \end{pmatrix} \sim N_{m+n} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \Delta + D^T \Sigma D & 0 \\ 0 & I_n \end{pmatrix} \right),$$

where I_n is an n -dimensional identity matrix. This way, we can rewrite Z as $Z = -FU + G^{\frac{1}{2}}V$, where $F = \Sigma D(\Delta + D^T \Sigma D)^{-1}$ and $G = \Sigma - \Sigma D(\Delta + D^T \Sigma D)^{-1} D^T \Sigma$. Alard and Navy [Alard and Navy \(2007\)](#) by considering (3.1) and the independency

of U and V , it has been shown that

$$Y \stackrel{d}{=} \mu - F[U|U \leq 0] + G^{\frac{1}{2}}V. \quad (3.2)$$

To avoid complexity and heavy computation, a simple case is simulated from a Spatial Skew Gaussian random field. To increase the amount of skewness in the process $Y(\cdot)$ with choosing proper values for m , D and Δ will have a great impact on the volume of these calculations. To generate data from the proposed random field in (3.2), we put $D = \delta A$, where the constant δ shows skewness and A is a non-zero matrix with independent components from δ . In this case, (3.2) can be written as

$$\begin{aligned} Y \stackrel{d}{=} & \mu - \delta \Sigma A (\Delta + \delta^2 A^T \Sigma A)^{-1} [U|U \leq 0] \\ & + (\Sigma - \delta^2 \Sigma A (\Delta + \delta^2 A^T \Sigma A)^{-1} A^T \Sigma)^{\frac{1}{2}} V. \end{aligned} \quad (3.3)$$

It is clear that when $\delta = 0$, Y is independent from U and consequently Y is a Gaussian vector with mean μ and covariance matrix Σ . By setting $A^T \Sigma A = \Sigma_m$,

$$Y \stackrel{d}{=} \mu - \frac{\delta}{1 + \delta^2} \Sigma A \Sigma_m^{-1} [U|U \leq 0] + (I - \frac{\delta^2}{1 + \delta^2} \Sigma A \Sigma_m^{-1} A^T \Sigma)^{\frac{1}{2}} V. \quad (3.4)$$

we consequently opt to set $m = n$ that is called *homotopic case*. For $n = m$ and $A = I_n$, we have

$$\begin{aligned} Y \stackrel{d}{=} & \mu - \frac{\delta}{1 + \delta^2} [U|U \leq 0] + \frac{1}{\sqrt{1 + \delta^2}} \Sigma^{\frac{1}{2}} V \\ Y \stackrel{d}{=} & \mu + \frac{\delta}{1 + \delta^2} [U|U \geq 0] + \frac{1}{\sqrt{1 + \delta^2}} \Sigma^{\frac{1}{2}} V \end{aligned} \quad (3.5)$$

since $[U|U \leq 0] = -[U|U \geq 0]$ The vector Y is the sum of a sample of a stationary process with covariance function $\frac{C(h)}{(1+\delta^2)}$ and a vector U conditional on $[U \geq 0]$. Figure 1 displays the histogram and variogram computed from 100 realizations of model (3.5), with $n = 100$ and an exponential covariance function $C(h) = \exp(-\frac{\|h\|}{a})$, with $a = 0.1$. The experimental variogram is the arithmetic average of $0.5[Y(x_i) - Y(x_j)]^2$ across the realizations. Two series of simulations were performed corresponding to the different amount of skewness. These graphs indicate the model (3.5) generates sufficient skewness if the parameter δ is large enough. When $\delta = 1$, there is a limited amount of skewness; when $\delta = 4$, the histogram is skewed towards the positive values; there are very few negative values. Note that the experimental variograms show in both cases a spatial structure with a range close to the parameter but the variance illustrated by the sill of the variograms decreases as δ increases.

In the following section, after a brief introduction of the survival models, by using (3.5) we introduce a new model for survival analysis with an SSG random field for random effects that offers more efficiency and flexibility compared to existing models.

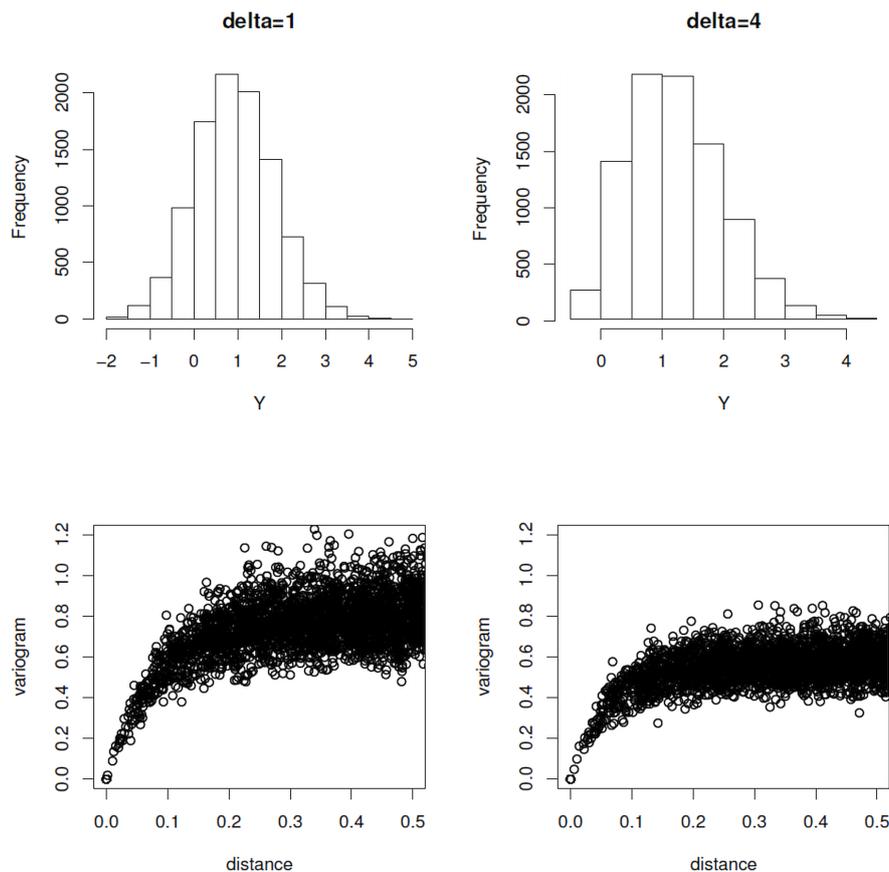


Figure 1: Histogram and experimental variogram of simulated homotopic model.

4 Survival Models

Survival analysis is generally defined as a set of methods for analyzing data where the outcome variable is the time until the occurrence of an event of interest. The event can be death, the occurrence of a disease, marriage, divorce, etc. The time to event or survival time can be measured in days, weeks, years, etc. For example, if the event of interest is a heart attack, then the survival time can be the time in years until a person develops a heart attack. Several models are available to analyze the relationship between a set of predictor variables with the survival time. Methods include parametric, nonparametric and semi-parametric approaches. The most important of these models is the Cox proportional hazards model.

If the values of survival data are dependent on their location, neither cox and frailty models can not provide an accurate analysis of such data. But in some applications, the location of each item can cause an unknown risk factor where each survival time depends locally on the other survival times. Besides, the close observations are more correlated than those farther ones. But in real data, this

spatial effect might not be either measurable or observable. Thus, according to spatial correlation in a survival time data set a class of frailty models is presented that is called spatial survival models. Typically, a random field is added to the model to consider the spatial correlation of these data. So the spatial hazard function [Motarjem et al. \(2017\)](#) is defined as

$$h(t|X, Z) = h_0(t) \exp(\beta'X + Z(s)), \tag{4.1}$$

where $h_0(\cdot)$ is baseline hazard function, β is the regression parameter vector, $Z(\cdot)$ is a Gaussian random field and s denotes the location of spatial survival data in $D \subseteq R^2$. The spatial covariance or covariogram between two different items, at different locations s_i and s_j is defined by

$$Cov(Z(s_i), Z(s_j)) = C(h); \quad h = \|s_i - s_j\|$$

where $\|\cdot\|$ is ordinary Euclidean distance norm.

As we said earlier, due to the existence of skewness in survival data, the Gaussian assumption for random effects may not be realistic. in the following , by considering a Spatial Skew random fields for random effects, a new model for skew spatial random effects is proposed. The spatial survival model with SSG random effects is a generalized version of the model (4.1) in this model the SSG random process $Y(\cdot)$ represents the spatial random effects as follows

$$h(t|X, Y) = h_0(t) \exp(\beta'X + Y(s)). \tag{4.2}$$

where $h_0(\cdot)$ is baseline hazard function, β is the regression parameter vector, and $Y(s)$ is the realization of an SSG random process that generated by (3.5).

For parameters estimation of spatial survival model with skew-Gaussian random effects By generalizing the spatial survival model we can find the rank likelihood function of the survival times as

$$L(\beta, \eta) = \int \cdots \int \prod_{i=1}^n \left\{ \frac{\exp(\beta'X_i + Y_i)}{\sum_{i'} \xi_{i'}(t_i) \exp(\beta'X_{i'} + Y_{i'})} \right\}^{\delta_i} dF(Y_1, \dots, Y_n), \tag{4.3}$$

where $\eta = (a, \sigma^2)$ is the vector of spatial parameters, $Y_i = Y(s_i)$ and

$$\xi_{i'}(t_i) = \begin{cases} 1, & t_{i'} \geq t_i \\ 0, & t_{i'} < t_i \end{cases}$$

is an indicator function. The estimator of the cumulative baseline hazard function $\hat{H}_0(t)$ is completely the same as the Cox model. So we have

$$\hat{H}_0(t_i) = \sum_{\{j; t_j \leq t_i\}} \frac{\delta_j}{\sum_{k \in R(t_k)} e^{\beta'X_k}},$$

where $R(t_k)$ is the set of all subjects at risk. The regression coefficients and spatial parameters should be estimated by maximizing the rank likelihood function, that does not have not a closed form. So, numerical methods such as MCMC algorithm are usually used.

Since Y_1, \dots, Y_n are correlated we are not able to calculate the integrate (4.3) to obtain $L(\beta, \eta)$. The main idea is transforming $dF(Y_1, \dots, Y_n)$ to $\prod_{i=1}^n dF_i(u_i)$ and a solution is using Cholesky decomposition to calculate this term. The Cholesky decomposition of covariance matrix $Q = Cov(Y)$ where $Y = (Y_1, \dots, Y_n)$, results a lower triangular matrix A such that $Q = AA'$. Then by generating $u = (u_1, \dots, u_n)$ from a standard Normal distribution we can generate the random vector, $\nu = (\nu_1, \dots, \nu_n)$, where $\nu = Au$. Since

$$E(\nu) = AE(u) = \mathbf{0},$$

$$Var(\nu) = A Var(u)A' = AA' = Q,$$

The random vector ν is equally distributed with Y . Thus we can rewrite (4.3) as

$$L(\beta, \eta) = \int \dots \int \prod_{i=1}^n \left[\frac{\exp(\beta' X_i + \sum_{k=1}^i A_{ik}(\eta)u_k)}{\sum_{i'} \xi_{i'}(t_i) \exp(\beta' X_{i'} + \sum_{k=1}^j A_{jk}(\eta)u_k)} \right]^{\delta_i} \prod_{i=1}^n d\Phi(u_i), \quad (4.4)$$

where $\Phi(\cdot)$ is the cumulative standard Normal distribution and A_{ik} is an element of matrix A . Now, we generate vector $u = (u_1, \dots, u_n)$ for M times. Then (4.4) can be approximated by

$$\hat{L}(\beta, \eta) = \frac{1}{M} \sum_{b=1}^M \prod_{i=1}^n \left[\frac{\exp(\beta' X_i + \sum_{k=1}^i A_{ik}(\eta)u_k^{(b)})}{\sum_{i'} \xi_{i'}(t_i) \exp(\beta' X_{i'} + \sum_{k=1}^j A_{jk}(\eta)u_k^{(b)})} \right]^{\delta_i}. \quad (4.5)$$

By maximizing (4.5), the estimate of parameters can be achieved. Obviously, increasing M gives more precise estimates.

5 Simulation Study

In this section spatial survival and proposed model for geostatistical survival data are evaluated and numerically compared. For this purpose, a spatial survival data set with SSG random field should be generated. A simulation method for generating survival data was proposed by ?. Here this method is modified to generate spatial survival data by using an SSG random field.

Assume that, the baseline hazard function, $h_0(\cdot)$, in the survival model has a known parametric form. According to the spatial survival model with SSG random effect we have

$$S(t|X) = \exp(-H_0(t) \exp(\beta' X + Y(s))),$$

where $H_0(t) = \int_0^t h_0(u)du$ is the cumulative hazard function, β is the regression parameter vector, $Y(s)$ is a SSG random field and $S(\cdot)$ is the survival function. as we know, the survival function is defined as

$$S(t|X) = 1 - F(t|X)$$

where F is the cumulative distribution function of T . based on spatial survival model with SSG random effect we can generate a spatial survival data set. For instance, for a exponential baseline hazard function with scale parameter $\lambda > 0$, we have

$$T = -\left[\frac{\log U}{\lambda \exp(\beta' X + Y(s))}\right]. \quad (5.1)$$

By generating $Y(\cdot)$ from (3.5) and inserting in (5.1) and specifying the values of λ , β and U , the spatial survival data with SSG random effects can be simulated.

Based on the above, a total of 400 spatial survival data are generated in a regular 20×20 grid for simulating spatial survival data with SSG random effects. To simulate the data, the baseline hazard function with exponential distribution and scale parameter $\lambda = 1$ is used, the vector of explanatory variables is derived from the standard Normal distribution and $\beta = 1$ is the regression coefficient. In addition $Y(s)$ is a realization of SSG random field with the exponential covariogram, $C(h) = \sigma^2 \exp(-\frac{\|h\|}{a})$, where h is the spatial lag, $\sigma^2 = 1$, $a = 1$ and $\delta = 5$, are the variance, range and skewness of the random field, respectively. According to the importance of censoring survival times and different types of censoring, in this study, right censored data is used for 20% and 80% levels. Using the *R* programming environment, datasets of spatial survival data with size $n = 400$ and also 100 and 500 times iteration are generated. Next, the Cox, frailty, spatial survival and proposed models fitted to these data. Note that when $\delta = 0$ the spatial survival data with Gaussian random effects will be produced.

Table 1: Parameter estimation of spatial survival model with Gaussian random effects.

Percentage		M=100			M=500		
Censor	Paramter	Estimate	MAPB	MSE	Estimate	MAPB	MSE
20%	β	0.96	3.21	0.012	0.98	1.98	0.007
	a	0.95	5.32	0.008	0.97	3.21	0.004
	σ^2	1.97	2.13	0.004	1.98	1.19	0.001
80%	β	0.92	7.12	0.016	0.94	5.19	0.009
	a	0.90	9.18	0.015	0.93	6.12	0.001
	σ^2	1.91	7.11	0.011	1.95	4.14	0.009

The mean square error (MSE) and the mean absolute percentage bias (MAPB) is defined by

$$MAPB = \frac{1}{M} \sum_{i=1}^M \left| \frac{\hat{\theta}_i - \theta}{\theta} \right| \times 100$$

are used for evaluation of parameter estimations, where $\theta = (\sigma^2, a, \beta, \delta)$ is the model

Table 2: Parameter estimation of spatial survival model with SSG random effects.

Percentage		M=100			M=500		
Censor	Parameter	Estimate	MAPB	MSE	Estimate	MAPB	MSE
20%	β	0.85	14.38	0.032	0.88	12.78	0.025
	a	0.81	19.64	0.019	0.83	16.92	0.016
	σ^2	1.71	29.16	0.039	1.77	23.16	0.023
80%	β	0.79	21.13	0.043	0.83	17.23	0.031
	a	0.77	24.16	0.039	0.81	19.64	0.024
	σ^2	1.65	35.18	0.055	1.69	29.12	0.042

parameters and M denotes the number of iterations.

The results are reported in Tables 1 through 4.

Table 3: Parameter estimation of proposed model with Gaussian random effects.

Percentage		M=100			M=500		
Censor	Parameter	Estimate	MAPB	MSE	Estimate	MAPB	MSE
20%	β	0.95	3.27	0.014	0.97	2.09	0.008
	a	0.96	5.24	0.007	0.96	3.39	0.006
	σ^2	1.94	2.18	0.007	1.96	1.27	0.003
	δ	0.000	0.000	0.000	0.000	0.000	0.000
80%	β	0.91	7.69	0.021	0.95	5.11	0.007
	a	0.93	9.04	0.012	0.90	6.24	0.003
	σ^2	1.84	8.68	0.019	1.93	4.93	0.015
	δ	0.000	0.000	0.000	0.000	0.000	0.000

As Table 1 shows, fitting spatial survival model to spatial survival data with Gaussian random effects has good results. But the comparison of the results of the Tables 1 and 2 show that the spatial survival model in the presence of SSG random effects does not provide proper results. By considering the results of Tables 3 and 4 and comparing with Table 1 and 2, we find that the proposed model in the presence of SSG random effect gives much better fit to the data than the competing models. However, in the case of Gaussian random effects, it has the same performance as the other model. Nonetheless, the proposed model can be used to analyze spatial survival data. The MAPB criterion in Table 1 through 4, show that by increasing M , the biases of estimates are reduced. The simulation results show that the proposed model is more efficient in the presence of the skewness and has the equal efficiency of the spatial survival model for the non-skewed cases.

Table 4: Parameter estimation of proposed model with SSG random effects.

Percentage		M=100			M=500		
Censor	Parameter	Estimate	MAPB	MSE	Estimate	MAPB	MSE
20%	β	0.94	4.11	0.013	0.96	3.76	0.008
	a	0.93	6.38	0.011	0.95	4.39	0.009
	σ^2	1.92	5.11	0.009	1.94	4.18	0.007
	δ	4.96	2.11	0.013	4.94	1.96	0.005
80%	β	0.90	8.72	0.021	0.92	5.13	0.014
	a	0.89	8.11	0.016	0.91	6.12	0.013
	σ^2	1.89	9.12	0.013	1.91	7.01	0.012
	δ	4.89	6.17	0.018	4.94	5.019	0.011

6 Discussion and Results

In this paper, a method for generating spatial survival data with SSG random effect is introduced and a simulation study is performed for analyzing the efficiency of different survival models in fitting spatially correlated survival times. The simulation study showed that increasing the censoring percentage generally caused imprecise estimate parameters while this inaccuracy in the proposed model is less than other models for estimating regression parameters. The discussion on such properties the isotropy and stability of random fields seem to be a good subject for future studies.

References

- Allard, D. and Naveau. P. (2007), A New Spatial Skew-Normal Random Field Model, *Communications in Statistics* , **36**, 1821-1834.
- Azzalini, A. and Dalla Valle, A. (1996), The Multivariate Skew-normal Distribution, *Biometrika*, **83**, 715-726.
- Bender, R., Augustin, T., and Blettner, M. (2005), Generating Survival times to Simulate Cox Proportional Hazards Models, *Statistics in Medicine*, **24**, 1713-1723.
- Cressie, N. (1993), *Statistics for Spatial Data*, Revised Edition, John Wiley & Sons, New York.
- Dominguez-Molina, J., Gonzalez-Farias, G., and Ramos-Quiroga, R. (2004), Skewnormality in stochastic frontier analysis. In Genton, M., Editor, *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, pages 223-241. Boca-Raton: Chapman & Hall/CRC.
- Gonzalez-Farias, G., Dominguez-Molina, J. and Gupta, A. (2004), The Closed Skewnormal Distribution, In: *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*, Genton, M. G., Ed., Chapman & Hall, Boca Raton

Kim, H. and Mallick, B. (2002), Analyzing spatial data using skew-gaussian processes. In Lawson, A. and Denison, D., Editors, *Spatial Cluster Modelling*. Boca Raton: Chapman & Hall/CRC.

Motarjem, K., Mohammadzadeh, M. and Abyar, A. (2017), Geostatistical Survival Model with Gaussian Random Effect, *Statistical Papers*, DOI: 10.1007/s00362-017-0922-8.