

**BAYESIAN SPATIAL DATA ANALYSIS
WITH APPLICATION TO
THE MISSOURI OZARK FOREST
ECOSYSTEM PROJECT**

A Dissertation

Presented to

the Faculty of the Graduate School

University of Missouri-Columbia

In Partial Fulfillment

Of the Requirements for the Degree

Doctor of Philosophy

by

XIAOQIAN SUN

Dr. Zhuoqiong He, Dissertation Supervisor

AUGUST 2006

The undersigned, appointed by the Dean of the Graduate School,
have examined the dissertation entitled.

BAYESIAN SPATIAL DATA ANALYSIS
WITH APPLICATION TO
THE MISSOURI OZARK FOREST ECOSYSTEM PROJECT

presented by XIAOQIAN SUN

A candidate for the degree of Doctor of Philosophy

And hereby certify that in their opinion it is worthy of accep-
tance.

Dr. Zhuoqiong He _____

Dr. Paul Speckman _____

Dr. Dongchu Sun _____

Dr. Christopher Wikle _____

Dr. John Kabrick _____

Acknowledgements

I would like to express my sincere gratitude to my advisor, Dr. Zhuoqiong He for her guidance, encouragement and patience throughout this project. Also, I would like to thank Dr. John Kabrick and Dr. Dongchu Sun for their kind help during preparing this project. This work would never have been done without their inspiration and enthusiasm.

I am also very grateful to Dr. Speckman and Dr. Wikle for serving on my advisory committee. Special thanks go to Judy and Tracy for their help during my stay at Department of Statistics, University of Missouri–Columbia. In addition, I would like to express my appreciation to Mr. B. J. Gorlinsky of Missouri Conservation Department for providing the useful data sets.

Foremost, I wish to acknowledge all the love from my wife Yanhong and my son Bo. It is to them that I dedicate this work.

Contents

Acknowledgements	ii
List of Tables	vi
List of Figures	vii
Abstract	x
1 Introduction	1
2 Bayesian Inference on a Covariance Matrix in a Star-shaped Model	4
2.1 Introduction	4
2.2 MLEs and Unbiased Estimators	9
2.2.1 Sample observations	9
2.2.2 Cholesky decomposition	10
2.2.3 The maximum likelihood estimates	11
2.2.4 Unbiased estimators	14
2.3 The Invariant Haar Measures and Noninformative Priors of Ψ	15
2.3.1 The Invariant Haar measures	15
2.3.2 The Jeffreys prior and a reference prior	16
2.4 Properties of Posteriors of Ψ	17
2.5 Bayesian Estimators of Ω under the Stein Loss	19
2.6 Bayesian Estimators of Ω under the Entropy Loss	23
2.7 Bayesian Estimators of Ω under the Symmetric Loss	25
2.8 Estimating the Covariance Matrix	27
2.9 Simulation Results	28
2.10 Concluding Remarks	34
2.11 Appendix	36

3	An Efficient Simulation Algorithm for Bayesian Spatial Analysis	46
3.1	Introduction	46
3.2	The Model	48
3.2.1	The likelihood	48
3.2.2	The Priors	49
3.2.3	Integrated likelihood	50
3.3	Posterior Simulation	51
3.3.1	Gibbs sampler	51
3.3.2	The generalized Ratio-of-Uniforms method	52
3.3.3	A new simulation algorithm	53
3.4	An Illustrative Example	55
3.5	Concluding Remarks and Discussions	62
4	Bayesian Spatial Reference Analysis with Application to Site Index Prediction	64
4.1	Introduction	64
4.2	The Gaussian Model	67
4.2.1	Data structure and the likelihood	67
4.2.2	The correlation function	71
4.2.3	The prior	72
4.3	Statistical Inference for Parameters	73
4.3.1	Estimating the smoothness parameter	73
4.3.2	Posterior simulation	75
4.4	Model Validation	84
4.5	Spatial Prediction of the Site Index	86
4.5.1	Bayesian prediction at one location	86
4.5.2	Bayesian prediction of spatial block average	90
4.5.3	Bayesian prediction of spatial CDFs	97
4.6	Concluding Remarks and Discussions	97
5	Bayesian Spatial Models with Repeated Measurements — with Application to the Herbaceous Data Analysis	103
5.1	Introduction	103
5.2	Model Development	105
5.2.1	Hierarchical Model	105
5.2.2	The Likelihood	108
5.3	The Posterior Distribution and Sampling Schemes	109

5.3.1	The Full Conditional Distributions.....	109
5.3.2	Shrinkage Slice Sampler.....	130
5.3.3	Simulation Algorithm.....	112
5.4	Prediction.....	113
5.5	Application to the Herbaceous Data Analysis.....	114
5.6	Discussion and Conclusion.....	124
	Bibliography	130
	Vita	142

List of Tables

4.1	Summary of Covariates	70
4.2	Posterior Quantities of $(\boldsymbol{\beta}, \sigma^2, \theta)$	84
5.1	Posterior Quantities of $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$	117

List of Figures

2.1	Risk Comparisons when $(p_1, p_2, p_3) = (2, 1, 1)$, $(n_1, n_2, n_3) = (3, 4, 5)$ and $5 \leq n \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* . . .	29
2.2	Risk Comparisons when $(p_1, p_2, p_3) = (2, 1, 1)$, $(n, n_2, n_3) = (5, 4, 5)$ and $5 \leq n_1 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* . . .	30
2.3	Risk Comparisons when $(p_1, p_2, p_3) = (2, 1, 1)$, $(n, n_1, n_3) = (5, 4, 5)$ and $5 \leq n_2 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* . . .	30
2.4	Risk Comparisons when $(p_1, p_2, p_3) = (2, 2, 1)$, $(n_1, n_2, n_3) = (3, 5, 4)$ and $5 \leq n \leq 20$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* . . .	31
2.5	Risk Comparisons when $(p_1, p_2, p_3) = (2, 2, 1)$, $(n, n_2, n_3) = (6, 5, 4)$ and $3 \leq n_1 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* . . .	31
2.6	Risk Comparisons when $(p_1, p_2, p_3) = (2, 2, 1)$, $(n, n_1, n_3) = (6, 3, 4)$ and $5 \leq n_2 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* . . .	32
2.7	Risk Comparisons when $(p_1, p_2, p_3, p_4, p_5) = (2, 2, 3, 4, 1)$, $(n_1, n_2, n_3, n_4, n_5) = (3, 5, 6, 7, 4)$ and $13 \leq n \leq 20$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^*	32
2.8	Risk Comparisons when $(p_1, p_2, p_3, p_4, p_5) = (2, 2, 3, 4, 1)$, $(n, n_2, n_3, n_4, n_5) = (14, 5, 6, 7, 4)$ and $3 \leq n_1 \leq 10$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^*	33
2.9	Risk Comparisons when $(p_1, p_2, p_3, p_4, p_5) = (2, 2, 3, 4, 1)$, $(n, n_1, n_3, n_4, n_5) = (14, 4, 6, 7, 4)$ and $5 \leq n_2 \leq 11$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^*	33
2.10	Risk Comparisons between $\mathbf{\Omega}_{iB}$ and $\mathbf{\Omega}_{Bi}$ when $(p_1, p_2, p_3) = (2, 2, 3)$, $(n, n_1, n_2, n_3) = (10, 3, 5, 4)$ and $\mathbf{\Sigma} = \text{diag}(\lambda, 1, 1, 1, 1, 1, 1)$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^*	34
3.1	The Reference Prior $\pi^R(\theta)$	57
3.2	The Non-normalized Marginal Posterior Density of θ	57
3.3	Acceptance Area based on the Generalized Ratio-of-Uniforms Method.	59

3.4	Histogram, Density Estimate for the Posterior of θ based on Simulation.	60
3.5	Histogram, Density Estimate for the Posterior of σ^2 based on Simulation.	61
3.6	Posterior Densities of $\beta_i, i = 0, 1, \dots, 5$	62
4.1	173 Sampled Black Oaks in Sites One and Two of the MOFEP Study: Locations for Modelling Marked as “.” and Those for Model Validation Marked as “×”.	68
4.2	The marginal density $m(\mathbf{z} \nu)$ in terms of ν	75
4.3	The Marginal Prior $\pi(\theta)$ with $\nu = 0.13$	77
4.4	The Plot of $L^I(\theta; \mathbf{z})\pi(\theta)$ with $\nu = 0.13$	77
4.5	The Acceptance Area with the Ratio-of-Uniforms Method	80
4.6	Histogram of θ for the Model in § 4.2.	82
4.7	Histogram of σ^2 for the Model in § 4.2.	82
4.8	Histograms of $\beta_0, \beta_1, \beta_2, \beta_3$ for the Model in § 4.2.	83
4.9	Histogram of z_0 Sampling from the Predictive Distribution at (665614.875, 4115457.5): Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep)	88
4.10	Histograms of the Predictive Distribution at Four Different Locations: (a) Location (666194.875, 4116397.5), Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep to Very Deep); (b) Location (666124.875, 4114517.5), Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep to Very Deep); (c) Location (664474.875, 4115407.5), Covariate Information (Aspect Class–Protected, Land Type Association–OZ9e and Soil Depth–Shallow to Deep); (d) Location (665564.875, 4115337.5), Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep to Very Deep)	89
4.11	The Map of the Aspect Class at Site One: the Red Represents the Area with Protected and the Blue Represents the Area with Exposed	91
4.12	The Map of the Land Type Association at Site One: the Red Represents Current River Oak-Pine Woodland/Forest Hills (OZ9b) and the Blue Represents Current River Oak Forest Breaks (OZ9e).	92
4.13	The Map of the Soil Depth at Site One: the Red Represents the Soil Depth Varying from Deep to Very Deep and the Blue Represents the Soil Depth Varying from Shallow to Deep.	93
4.14	The Prediction Map of the Site Index at Site One	94
4.15	The Map of Standard Deviation of the Site Index Prediction at Site One	95

5.1	MOFEP Vegetation Plot Design.	106
5.2	Locations of subplots in site one (the square area is for prediction) . .	116
5.3	Histograms of β	118
5.4	Histograms of $(\sigma^2, \tau^2, \theta, \rho)$	118
5.5	The Covariate Information in the Prediction Domain: (a) Aspect Class (1 = Protected, 0 = Exposed), (b) Soil Depth (1 = Deep to Very Deep, 0 = Shallow to Moderate Deep).	119
5.6	The Maps of Prediction and Prediction Error: (a) Prediction, (b) Stan- dard Deviation of the Prediction.	119
5.7	The Maps of Prediction Difference between Four Quadrats and Two Quadrats: (a) Prediction, (b) Standard Deviation of the Prediction .	120
5.8	The Relative Difference of the Predictions between Four Quadrats and Two Quadrats.	120
5.9	The Comparison of Posterior Densities: the Solid Line with the Case of $d = 100$, $a_1 = a_2 = a_3 = 2$ and $b_1 = 0.30, b_2 = 1.00, b_3 = 0.2$ and the Dashed Line with the Case of $d = 100$, $a_1 = a_2 = a_3 = 2$ and $b_1 = 0.48, b_2 = 2.50, b_3 = 0.25$	122

Abstract

This project consists of three distinct but related parts. The first part studies the problem of estimating the covariance matrix Σ and the precision matrix Ω (the inverse of the covariance matrix) in a star-shaped model with missing data. By considering a type of Cholesky decomposition of the precision matrix $\Omega = \Psi'\Psi$, where Ψ is a lower triangular matrix with positive diagonal elements, we introduce a special group \mathcal{G} , which is a subgroup of the group consisting of all lower triangular matrices, and then develop the invariant Haar measures on \mathcal{G} , the reference prior, and the Jeffreys prior of Ψ . We also introduce a class of priors of Ψ that includes all the priors described above. The posterior properties are discussed and the closed forms of Bayesian estimators are derived under any of the Stein loss, the entropy loss, and the symmetric loss. Consequently, the maximum likelihood estimate of the covariance matrix is inadmissible under any of the above three loss functions. Some simulation results are given for illustration.

The second part deals with Bayesian analysis for spatially correlated data. We first propose an efficient algorithm for Bayesian spatial analysis via the generalized Ratio-of-Uniforms method, which works for a general class of priors including the reference prior. The main advantage of our algorithm over other MCMC algorithms is that it

generates independent samples from the resulting posterior distribution. A detailed example based on simulation is provided for illustration. We then present a Bayesian spatial methodology for analyzing the site index data from the Missouri Ozark Forest Ecosystem Project (MOFEP). Based on ecological background and availability, we choose three variables, aspect class, land type association and soil depth as covariates. To allow great flexibility of the smoothness of the random field, we pick up the Matérn family as the correlation function. Because there is no previous knowledge of the parameters in the model, we choose the reference prior as an appropriate prior. An efficient algorithm based on the generalized Ratio-of-Uniforms method is applied for the posterior simulation. Our results show that aspect class and soil depth are both significant while land type association is less significant. Model validation is briefly discussed. In addition, a prediction map of site index in site one of MOFEP is created.

In the third part, we establish a new spatial model taking into account several close measurements as repeated measurements in one location and apply it to the analysis of the total vegetation coverage data in site one of MOFEP. An MCMC algorithm based on the shrinkage slice sampler is developed. Our results show that the soil depth covariate is an important factor while the aspect class is less important when modelling the total vegetation coverage. We also show that the strong spatial effect does exist in the data discussed and the measurements in four quadrats of a subplot are not strongly correlated but are not independent. In addition, prediction of the total vegetation coverage at unmeasured locations is established. Possible generalizations are briefly discussed.

Chapter 1

Introduction

The Missouri Ozark Forest Ecosystem Project (MOFEP) is an on-going, centuries-long experiment that is designed to monitor and assess the short and long-term effects of common management practices on Ozark ecosystems (see Brookshire and Shifley (1997), Shifley and Brookshire (2000) and Shifley and Kabrick (2002)). The MOFEP will provide a comprehensive evaluation of the impacts of operational management practices on a wide array of ecosystem attributes. This project mainly focuses on the analysis of the site index data and it is organized as follows.

In Chapter 2, we study the problem of estimating the covariance matrix $\mathbf{\Sigma}$ and the precision matrix $\mathbf{\Omega}$ (the inverse of the covariance matrix) in the star-shaped model with missing data. By considering a type of Cholesky decomposition of the precision matrix $\mathbf{\Omega} = \mathbf{\Psi}'\mathbf{\Psi}$, where $\mathbf{\Psi}$ is a lower triangular matrix with positive diagonal elements, we get the MLEs of the covariance matrix and precision matrix and prove that both of them are biased. Based on the MLEs, unbiased estimators of the covariance matrix and precision matrix are obtained. A special group \mathcal{G} , which is a subgroup of

the group consisting all lower triangular matrices, is introduced. The invariant Haar measures on \mathcal{G} , the reference prior, and the Jeffreys prior of Ψ also are discussed. We also introduce a class of priors of Ψ , which includes all the priors described above. The posterior properties are discussed and the closed forms of Bayesian estimators are derived under any of the Stein loss, the entropy loss, and the symmetric loss. Consequently, the MLE of the covariance matrix (precision matrix) is inadmissible under any of the above three loss functions. Some simulation results are given for illustration. Combining a star-shaped model with a Bayesian spatial model discussed in Chapters 3 and 4, one may consider to establish a new one, a Bayesian multivariate spatial model with conditional independence structure, which may be applied to the study of the MOFEP in the future.

In Chapter 3, we propose an efficient algorithm for Bayesian spatial analysis via the generalized Ratio-of-Uniforms method. The algorithm works for a general class of priors including the reference prior developed by Berger et al. (2001). The main advantage of our algorithm over other MCMC algorithms is that it generates independent samples from the resulting posterior distribution. A detailed example based on simulation is provided for illustration.

In Chapter 4, we develop a Bayesian spatial model for analyzing the site index data from the MOFEP. Based on ecological background and availability, we choose three variables, aspect class, land type association and soil depth as covariates. To allow great flexibility of the smoothness of the random field, we choose the Matérn family as the correlation function. We choose the reference prior as an appropriate prior because there is no previous knowledge of the parameters in the model. The algorithm based on the generalized Ratio-of-Uniforms method, proposed in Chapter 3, is applied for

the posterior simulation. One advantage of the algorithm is that this simulation method can generate independent samples from the required posterior distribution, which is much more efficient for both statistical inference of the parameters and prediction of the site indexes at unsampled locations. Our results show that aspect class and soil depth are both significant in the model while land type association is less significant. The model validation is briefly discussed. In addition, our simulation method allows easy realization for computing quantities from the posterior predictive distributions.

In Chapter 5, we establish a new spatial model taking into account several close measurements as repeated measurements in one location and apply it to the analysis of the total vegetation coverage data in site one of the MOFEP. An MCMC algorithm based on the shrinkage slice sampler is developed. Our results show that the soil depth covariate is an important factor while the aspect class is less important when modelling the total vegetation coverage. In addition, the strong spatial effect does exist in the data discussed and the measurements in four quadrats of a subplot are not strongly correlated but are not independent. Prediction of the total vegetation coverage at unmeasured locations is developed. Finally, possible generalizations are briefly discussed.

Chapter 2

Bayesian Inference on a Covariance Matrix in a Star-shaped Model

2.1 Introduction

The multivariate normal distribution plays a key role in multivariate statistical analysis. There is a large literature on estimating the covariance matrix and precision matrix in the saturated multivariate normal population, where no additional restriction other than being positive definite is required. See, for example, Haff (1980), Sinha and Ghosh (1987), Krishnamoorthy and Gupta (1989), Yang and Berger (1994), Sun (1998), Sun and Pang (2000), Zhou et al. (2001) and others. However, as the number of variables p in a multivariate distribution increases, the number of parameters $p(p + 1)/2$ to be estimated increases fast. Unless the number of observations, n , is very large, estimation is often inefficient, and models with many parameters are, in general, difficult to interpret. In many practical situations, there will be some man-

ifest inter-relationships among several variables. One important case is that several group variables that are conditionally independent, given other remaining variables. For the multivariate normal distribution, this will correspond to some zeros among the entries of the precision matrix. See Dempster (1972), Whittaker (1990), or Lauritzen (1996). Bayesian model selection of detecting zeros in precision matrix can be found in Wong et al. (2002).

Assume that $\mathbf{X} \sim N_p(\mathbf{0}, \boldsymbol{\Sigma})$. The vector \mathbf{X} is partitioned into k groups, that is, $\mathbf{X} = (\mathbf{X}'_1, \mathbf{X}'_2, \dots, \mathbf{X}'_k)'$, where \mathbf{X}_i is p_i -dimensional, and $\sum_{i=1}^k p_i = p$. We assume that given \mathbf{X}_1 , the other subvectors $\mathbf{X}_2, \dots, \mathbf{X}_k$ are mutually conditionally independent. From Whittaker (1990) and Lauritzen (1996), the precision matrix $\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1}$ has the following special structure:

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} & \boldsymbol{\Omega}_{13} & \cdots & \boldsymbol{\Omega}_{1k} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} & 0 & \cdots & 0 \\ \boldsymbol{\Omega}_{31} & 0 & \boldsymbol{\Omega}_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Omega}_{k1} & 0 & 0 & \cdots & \boldsymbol{\Omega}_{kk} \end{pmatrix}. \quad (2.1)$$

In fact, we can easily show that (2.1) is equivalent to

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} & \cdots & \boldsymbol{\Sigma}_{1k} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{13} & \cdots & \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1k} \\ \boldsymbol{\Sigma}_{31} & \boldsymbol{\Sigma}_{31}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{33} & \cdots & \boldsymbol{\Sigma}_{31}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{k1} & \boldsymbol{\Sigma}_{k1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{k1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{13} & \cdots & \boldsymbol{\Sigma}_{kk} \end{pmatrix}. \quad (2.2)$$

The case of $k = 3$ is considered in detail by Whittaker (1990) and is called a ‘‘butterfly model.’’ For general k , we called the model a *star-shaped model* in Sun and Sun

(2005a) because the graphical shape of the relationships among the variables described by Whittaker (1990) or Lauritzen (1996) is like a star.

The above model is very popular in some areas, especially in economics. For example, let X_1 be the federal interest rate, which is a global variable, and X_2, \dots, X_{51} be the house price in each state, which are local variables. Then X_2, \dots, X_{51} are conditionally independent given X_1 because each house price $X_i, i = 2, \dots, k$ will normally depend on its local situation if the federal interest rate is fixed.

The above star-shaped model is a special case of the lattice conditional independence model introduced by Andersson and Perlman (1993). Although star-shaped models or general graphical models have been used widely, as far as we know, few theoretic results are obtained on estimating the covariance matrix and the precision matrix in lattice conditional independence models. Andersson and Perlman (1993) gave the form of the maximum likelihood estimator (MLE) of the covariance matrix Σ . Konno (2001) considered the estimation of the covariance matrix under the Stein loss

$$L_1(\hat{\Sigma}, \Sigma) = tr(\hat{\Sigma}\Sigma^{-1}) - \log |\hat{\Sigma}\Sigma^{-1}| - p \quad (2.3)$$

and proved that the MLE of Σ is inadmissible. In fact, the Stein loss for estimating the covariance matrix is equivalent to the following loss for estimating the precision matrix $\Omega = \Sigma^{-1}$,

$$L_1^*(\hat{\Omega}, \Omega) = tr(\hat{\Omega}^{-1}\Omega) - \log |\hat{\Omega}^{-1}\Omega| - p. \quad (2.4)$$

Of course, the Stein loss is related to the commonly used entropy loss. See Robert

(1994). Let $f(\mathbf{x} | \boldsymbol{\Sigma})$ be the density of \mathbf{X} under $\boldsymbol{\Sigma}$. The entropy loss is,

$$\begin{aligned} L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) &= 2 \int \log \left\{ \frac{f(\mathbf{X} | \boldsymbol{\Sigma})}{f(\mathbf{X} | \hat{\boldsymbol{\Sigma}})} \right\} f(\mathbf{X} | \boldsymbol{\Sigma}) d\mathbf{X} \\ &= \text{tr}(\hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}) - \log |\hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}| - p. \end{aligned} \quad (2.5)$$

The Stein loss is obtained from the entropy loss by switching the role of two arguments, $\hat{\boldsymbol{\Omega}}$ and $\boldsymbol{\Omega}$. The loss function L_2 is typical entropy loss and has been studied by many authors such as Sinha and Ghosh (1987), Krishnamoorthy and Gupta (1989), and others.

Note that because neither L_1 nor L_2 is symmetric, we could consider a symmetric version by adding the Stein loss and entropy loss:

$$L_3(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = L_1(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) + L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}(\hat{\boldsymbol{\Sigma}} \boldsymbol{\Sigma}^{-1}) + \text{tr}(\hat{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Sigma}) - 2p \quad (2.6)$$

The symmetric loss L_3 was introduced by Kubokawa and Konno (1990) and Gupta and Ofori-Nyarko (1995). It can be seen as estimating the covariance matrix and the precision matrix simultaneously.

For estimating the precision matrix $\boldsymbol{\Omega}$, the entropy loss and the symmetric loss will be

$$L_2^*(\hat{\boldsymbol{\Omega}}, \boldsymbol{\Omega}) = L_2(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}(\hat{\boldsymbol{\Omega}} \boldsymbol{\Omega}^{-1}) - \log |\hat{\boldsymbol{\Omega}} \boldsymbol{\Omega}^{-1}| - p \quad (2.7)$$

and

$$L_3^*(\hat{\boldsymbol{\Omega}}, \boldsymbol{\Omega}) = L_3(\hat{\boldsymbol{\Sigma}}, \boldsymbol{\Sigma}) = \text{tr}(\hat{\boldsymbol{\Omega}} \boldsymbol{\Omega}^{-1}) + \text{tr}(\hat{\boldsymbol{\Omega}}^{-1} \boldsymbol{\Omega}) - 2p. \quad (2.8)$$

For convenience, we will still name L_1^* , L_2^* , L_3^* as the Stein loss, the entropy loss, and the symmetric loss for estimating the precision matrix $\boldsymbol{\Omega}$.

Sun and Sun (2005a) considered the problem of estimating the precision matrix under the entropy loss L_2^* and the symmetric loss L_3^* in the star-shaped case with complete observations. They obtained the closed forms of Bayesian estimators with respect to a class of priors of Ψ . Consequently, the MLE of the precision matrix is inadmissible under either the entropy loss L_2^* or the symmetric loss L_3^* .

Considering that missing data problems occur frequently in practice and their analysis can be challenging, we will study the problem of estimating the covariance matrix and the precision matrix in the star-shaped model with missing data. For estimating the covariance matrix without restriction, Anderson (1957) listed several general cases, where the MLEs of the parameters can be obtained in closed form. Among these cases, the monotone missing-data pattern is most important. One also can see the related references by Little and Rubin (1987), Konno (1995), Liu (1999), Domonici et al. (2000) and so on. However, because there are some restrictions on the covariance matrix in our model, we will see a lot of differences.

In this chapter, we will consider estimation of the covariance matrix and the precision matrix in the star-shaped model with missing data, which generalizes some results in Sun and Sun (2005a). In § 2.2, we first introduce the sample observations. By introducing a type of Cholesky decomposition of the precision matrix $\Omega = \Psi'\Psi$, where Ψ is a lower triangular matrix with positive diagonal elements, the MLEs of the covariance matrix and the precision matrix are obtained, and it is proved that both of them are not unbiased. Based on the MLEs, unbiased estimates of the covariance matrix and the precision matrix are given. Considering that the parameter Ψ plays an important role in estimating the covariance matrix and the precision matrix, a special group \mathcal{G} , which is related to the decomposition, is introduced in § 2.3. The invariant

Haar measures of this group are given, and we also prove that the Jeffreys prior of Ψ matrix is exactly the same as the right invariant Haar measure on the group \mathcal{G} . A reference prior is obtained by using the algorithm in Berger and Bernardo (1992). In § 2.4, we introduce a class of priors of Ψ , which includes all priors such as the left and right Haar measure and reference priors as special cases. Some properties on the posterior under such class of priors are discussed. In § 2.5, the closed form of Bayesian estimators with respect to this class of priors is obtained under the Stein loss. Consequently, the best equivariant estimates under the group \mathcal{G} is derived, and thus it shows that the MLE of Ω is inadmissible under the Stein loss. Results on the entropy loss and symmetric loss are shown in § 2.6 and § 2.7. The results on estimating covariance matrix are given in § 2.8. Some simulation results are given in § 2.9. Finally, we give some concluding remarks.

2.2 MLEs and Unbiased Estimators

2.2.1 Sample observations

Now suppose that we have the following observations:

$$\begin{aligned} \mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_n & \text{ from } \mathbf{X} \sim N_p(\mathbf{0}, \Sigma); \\ \mathbf{Z}_{11}, \mathbf{Z}_{12}, \dots, \mathbf{Z}_{1n_1} & \text{ from } \mathbf{X}_1 \sim N_{p_1}(\mathbf{0}, \Sigma_{11}); \\ \mathbf{Z}_{i1}, \mathbf{Z}_{i2}, \dots, \mathbf{Z}_{in_i} & \text{ from } \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_i \end{pmatrix} \sim N_{p_1+p_i} \left(\mathbf{0}, \begin{pmatrix} \Sigma_{11} & \Sigma_{1i} \\ \Sigma_{i1} & \Sigma_{ii} \end{pmatrix} \right), \quad i = 2, \dots, k. \end{aligned} \quad (2.9)$$

All \mathbf{Y}_i s and \mathbf{Z}_{ij} s are independent. Let $\mathbf{S} = \sum_{i=1}^n \mathbf{Y}_i \mathbf{Y}_i'$ and $\mathbf{V}_i = \sum_{j=1}^{n_i} \mathbf{Z}_{ij} \mathbf{Z}_{ij}'$, $i = 1, \dots, k$. Then $\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k$ are mutually independent and are sufficient statistics of

Σ or Ω . Now write $\mathbf{S} = (\mathbf{S}_{ij})$, where \mathbf{S}_{ij} is a $p_i \times p_j$ submatrix and

$$\mathbf{V}_i = \begin{pmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} \end{pmatrix}, \quad i = 2, \dots, k, \quad (2.10)$$

where \mathbf{V}_{i11} is a $p_1 \times p_1$ submatrix. Also let $\mathbf{V}_1 = \mathbf{V}_{111}$ for convenience. Assume that $n > p$, $n_1 > p_1$ and $n_i > p_1 + p_i, i = 2, \dots, k$. Then, $\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k$ are all positive definite with probability one and

$$\begin{aligned} \mathbf{S} &\sim W_p(n, \Sigma), \quad \mathbf{V}_1 \sim W_{p_1}(n_1, \Sigma_{11}), \\ \mathbf{V}_i &\sim W_{p_1+p_i} \left(n_i, \begin{pmatrix} \Sigma_{11} & \Sigma_{1i} \\ \Sigma_{i1} & \Sigma_{ii} \end{pmatrix} \right), \quad i = 2, \dots, k. \end{aligned} \quad (2.11)$$

We will estimate Σ and Ω based on the sufficient statistics $\mathbf{S}, \mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_k$.

2.2.2 Cholesky decomposition

Usually, it is difficult to get appropriate estimators of the covariance matrix or the precision matrix with some restrictions. For example, if you want to estimate Ω in (2.1) directly, you have to estimate $\Omega_{11}, \Omega_{22}, \dots, \Omega_{kk}, \Omega_{12}, \dots, \Omega_{1k}$ first. However, it will not guarantee that the estimate of Ω you obtained in this way is positive definite. Now we will introduce the following Cholesky decomposition method to get MLEs of Σ and Ω in the star-shaped model. This method will guarantee that the estimate of Ω obtained is positive definite. In addition, we will see that this decomposition is still useful in getting Bayesian estimates of Ω or Σ .

Let

$$\Omega = \Psi' \Psi \quad \text{or} \quad \Sigma = \Delta \Delta', \quad (2.12)$$

where both Ψ and Δ are p by p lower-triangular matrices with positive diagonal entries. Thus $\Psi = \Delta^{-1}$, where Δ is Cholesky decomposition of Σ . From the structure

of $\mathbf{\Omega}$ given by (2.1), it is easy to show that $\mathbf{\Psi}$ has the following block structure:

$$\mathbf{\Psi} = \begin{pmatrix} \mathbf{\Psi}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Psi}_{21} & \mathbf{\Psi}_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{\Psi}_{31} & \mathbf{0} & \mathbf{\Psi}_{33} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Psi}_{k1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Psi}_{kk} \end{pmatrix}, \quad (2.13)$$

and thus

$$\mathbf{\Delta} = \mathbf{\Psi}^{-1} = \begin{pmatrix} \mathbf{\Psi}_{11}^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{\Psi}_{22}^{-1}\mathbf{\Psi}_{21}\mathbf{\Psi}_{11}^{-1} & \mathbf{\Psi}_{22}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{\Psi}_{33}^{-1}\mathbf{\Psi}_{31}\mathbf{\Psi}_{11}^{-1} & \mathbf{0} & \mathbf{\Psi}_{33}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{\Psi}_{kk}^{-1}\mathbf{\Psi}_{k1}\mathbf{\Psi}_{11}^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{\Psi}_{kk}^{-1} \end{pmatrix}, \quad (2.14)$$

with the $\mathbf{\Psi}_{ii}$ being p_i by p_i lower-triangular matrices, $i = 1, \dots, k$. Note that there is no restriction on $\mathbf{\Psi}_{ij}$ ($i \geq j$) except requiring that all diagonal elements of $\mathbf{\Psi}_{ii}$ are positive. This good property enables us to estimate $\mathbf{\Psi}_{ij}$ first; then we can get the estimates of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$ directly from the relationship between $\mathbf{\Sigma}$ (or $\mathbf{\Omega}$) and $\mathbf{\Psi}$. This method will ensure that the estimate of $\mathbf{\Sigma}$ (or $\mathbf{\Omega}$) obtained is positive definite if the estimates of the diagonal elements in each $\mathbf{\Psi}_{ii}$ are positive. Other properties of this decomposition will be discussed in the next section.

2.2.3 The maximum likelihood estimates

Whittaker (1990) gives the expression of MLE of the covariance matrix $\mathbf{\Sigma}$ for $k = 3$ with complete observations. Sun and Sun (2005a) get the corresponding result for general k . We will generalize them to the star-shaped model with missing

observations. Let

$$\begin{aligned}
\mathbf{W}_{11} &= \mathbf{S}_{11} + \sum_{i=1}^k \mathbf{V}_{i11}; \\
\mathbf{W}_{i11} &= \mathbf{S}_{11} + \mathbf{V}_{i11}, \\
\mathbf{W}_{i1} &= \mathbf{W}'_{1i} = \mathbf{S}_{i1} + \mathbf{V}_{i21}, \\
\mathbf{W}_{i22} &= \mathbf{S}_{ii} + \mathbf{V}_{i22}, \\
\mathbf{W}_{ii-1} &= \mathbf{W}_{i22} - \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} \mathbf{W}_{1i}, \quad i = 2, \dots, k,
\end{aligned} \tag{2.15}$$

and let $m_1 = n + \sum_{t=1}^k n_t$, $m_i = n + n_i$, $i = 2, \dots, k$ throughout this chapter. Also, let $\mathbf{W}_{11 \cdot 1} = \mathbf{W}_{11}$ for convenience.

Proposition 2.1 *Based on the incomplete data $(\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k)$ in the star-shaped model, the MLE $\hat{\boldsymbol{\Sigma}}_M$ of $\boldsymbol{\Sigma}$ is given as follows:*

$$\begin{aligned}
\hat{\boldsymbol{\Sigma}}_{11}^M &= \frac{\mathbf{W}_{11}}{m_1}; \\
\hat{\boldsymbol{\Sigma}}_{i1}^M &= (\hat{\boldsymbol{\Sigma}}_{1i}^M)' = \frac{\mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} \mathbf{W}_{11}}{m_1}, \\
\hat{\boldsymbol{\Sigma}}_{ii}^M &= \frac{1}{m_i} \mathbf{W}_{ii-1} + \frac{1}{m_1} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} \mathbf{W}_{11} \mathbf{W}_{i11}^{-1} \mathbf{W}_{1i}, \quad i = 2, \dots, k; \\
\hat{\boldsymbol{\Sigma}}_{ij}^M &= \frac{1}{m_1} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} \mathbf{W}_{11} \mathbf{W}_{j11}^{-1} \mathbf{W}_{1j}, \quad 1 < i < j \leq k.
\end{aligned} \tag{2.16}$$

Proof. The likelihood function $f(\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k \mid \boldsymbol{\Psi})$ is proportional to

$$\begin{aligned}
& |\mathbf{S}|^{\frac{n-p-1}{2}} |\boldsymbol{\Sigma}|^{-\frac{n}{2}} \text{etr} \left\{ -\frac{1}{2} \boldsymbol{\Sigma}^{-1} \mathbf{S} \right\} |\mathbf{V}_1|^{\frac{n_1-p_1-1}{2}} |\boldsymbol{\Sigma}_{11}|^{-\frac{n_1}{2}} \text{etr} \left\{ -\frac{1}{2} \boldsymbol{\Sigma}_{11}^{-1} \mathbf{V}_1 \right\} \\
& \times \prod_{i=2}^k |\mathbf{V}_i|^{\frac{n_i-p_i-1}{2}} \left| \begin{array}{cc} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{1i} \\ \boldsymbol{\Sigma}_{i1} & \boldsymbol{\Sigma}_{ii} \end{array} \right|^{-\frac{n_i}{2}} \text{etr} \left\{ -\frac{1}{2} \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{1i} \\ \boldsymbol{\Sigma}_{i1} & \boldsymbol{\Sigma}_{ii} \end{pmatrix}^{-1} \mathbf{V}_i \right\} \\
\propto & |\boldsymbol{\Psi}|^n |\boldsymbol{\Psi}_{11}|^{n_1} \prod_{i=2}^k \left| \begin{array}{cc} \boldsymbol{\Psi}_{11} & \mathbf{0} \\ \boldsymbol{\Psi}_{i1} & \boldsymbol{\Psi}_{ii} \end{array} \right|^{n_i} \cdot \text{etr} \left\{ -\frac{1}{2} \boldsymbol{\Psi} \mathbf{S} \boldsymbol{\Psi}' \right\} \text{etr} \left\{ -\frac{1}{2} \boldsymbol{\Psi}_{11} \mathbf{V}_1 \boldsymbol{\Psi}'_{11} \right\} \\
& \times \prod_{i=2}^k \text{etr} \left\{ -\frac{1}{2} \begin{pmatrix} \boldsymbol{\Psi}_{11} & \mathbf{0} \\ \boldsymbol{\Psi}_{i1} & \boldsymbol{\Psi}_{ii} \end{pmatrix} \begin{pmatrix} \mathbf{V}_{i11} & \mathbf{V}_{i12} \\ \mathbf{V}_{i21} & \mathbf{V}_{i22} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Psi}_{11} & \mathbf{0} \\ \boldsymbol{\Psi}_{i1} & \boldsymbol{\Psi}_{ii} \end{pmatrix}' \right\}
\end{aligned}$$

$$\begin{aligned}
&= \prod_{i=1}^k |\Psi_{ii}|^{m_i} \cdot \prod_{i=1}^k \text{etr} \left\{ -\frac{1}{2} \Psi_{ii} \mathbf{W}_{ii \cdot 1} \Psi_{ii}' \right\} \\
&\quad \times \prod_{i=2}^k \text{etr} \left\{ -\frac{1}{2} (\Psi_{i1} + \Psi_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}) \mathbf{W}_{i11} (\Psi_{i1} + \Psi_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1})' \right\} \quad (2.17) \\
&\leq \prod_{i=1}^k |\Psi_{ii}|^{m_i} \cdot \prod_{i=1}^k \text{etr} \left\{ -\frac{1}{2} \Psi_{ii} \mathbf{W}_{ii \cdot 1} \Psi_{ii}' \right\}.
\end{aligned}$$

Hence, the MLE $\hat{\Psi}$ of Ψ will be determined by

$$\begin{aligned}
\hat{\Psi}_{ii}' \hat{\Psi}_{ii} &= m_i \mathbf{W}_{ii \cdot 1}^{-1}, & i = 1, \dots, k; \\
\hat{\Psi}_{i1} &= -\hat{\Psi}_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}, & i = 2, \dots, k
\end{aligned} \quad (2.18)$$

and thus by (2.12) and (2.14) the MLE of Σ is obtained as described in the proposition.

□

Under the conditions $n > p$, $n_1 > p_1$ and $n_i > p_1 + p_i$, $i = 2, \dots, k$, the MLE $\hat{\Sigma}_M$ is positive definite with probability one. In addition, by (2.18), the MLE $\hat{\Omega}_M$ of the precision matrix Ω can be straightforwardly obtained as follows:

Proposition 2.2 *Based on the incomplete data $(\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k)$ in the star-shaped model, the MLE $\hat{\Omega}_M$ of Ω is given by*

$$\begin{aligned}
\hat{\Omega}_{11}^M &= m_1 \mathbf{W}_{11}^{-1} + \sum_{i=2}^k m_i \mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii \cdot 1}^{-1} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}; \\
\hat{\Omega}_{1i}^M &= (\hat{\Omega}_{i1}^M)' = -m_i \mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii \cdot 1}^{-1}, \\
\hat{\Omega}_{ii}^M &= m_i \mathbf{W}_{ii \cdot 1}^{-1}, \quad i = 2, \dots, k.
\end{aligned} \quad (2.19)$$

The MLE $\hat{\Omega}_M$ of the precision matrix Ω also can be obtained by the following relationships between Ω and Σ ,

$$\begin{aligned}
\Omega_{11} &= \Sigma_{11}^{-1} + \sum_{i=2}^k \Sigma_{11}^{-1} \Sigma_{1i} \Sigma_{ii \cdot 1}^{-1} \Sigma_{i1} \Sigma_{11}^{-1}; \\
\Omega_{1i} &= -\Sigma_{11}^{-1} \Sigma_{1i} \Sigma_{ii \cdot 1}^{-1}, \\
\Omega_{ii} &= \Sigma_{ii \cdot 1}^{-1}, \quad i = 2, \dots, k,
\end{aligned} \quad (2.20)$$

where

$$\Sigma_{ii-1} = \Sigma_{ii} - \Sigma_{i1}\Sigma_{11}^{-1}\Sigma_{1i}, \quad i = 2, \dots, k.$$

Remark 2.1 For the star-shaped model with missing data, the MLE $\hat{\Sigma}_M$ is no longer a minimal sufficient statistic for Σ , which is different from the case with complete observations in Sun and Sun (2005a). In fact, $\mathbf{W}_{11}, \mathbf{W}_{211}, \dots, \mathbf{W}_{k11}, \mathbf{W}_{21}, \dots, \mathbf{W}_{k1}, \mathbf{W}_{22-1}, \dots, \mathbf{W}_{kk-1}$ are minimal sufficient statistics of Σ , which can be shown by (2.17).

Sun and Sun (2005a) showed that for a star-shaped model with complete observations, the MLE $\hat{\Sigma}_M$ of the covariance matrix Σ is unbiased while $\hat{\Omega}_M$ is biased. However, the following proposition shows that for the missing case, neither $\hat{\Sigma}_M$ nor $\hat{\Omega}_M$ is unbiased.

Proposition 2.3 Consider the star-shaped model with missing data.

- (a) The MLE $\hat{\Sigma}_M$ in (2.16) is not an unbiased estimate of Σ ;
- (b) The MLE $\hat{\Omega}_M$ in (2.19) is not an unbiased estimator of Ω .

Proof. See Appendix 2A. □

2.2.4 Unbiased estimators

Based on $\hat{\Sigma}_M, \hat{\Omega}_M$, we create unbiased estimates of Σ and Ω , respectively.

Proposition 2.4 Consider the star-shaped model with missing data.

- (a) An unbiased estimate $\hat{\Sigma}_U$ of Σ is given by

$$\hat{\Sigma}_{ii}^U = \left[1 - \frac{p_1(m_1 - p_1 - 1)}{m_1(m_i - p_1 - 1)} \right] \frac{\mathbf{W}_{ii-1}}{m_i - p_1} + \frac{\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}}{m_1}, \quad i = 2, \dots, k.$$

and $\hat{\Sigma}_{ij}^U = \hat{\Sigma}_{ij}^M$ for other i, j , where $\hat{\Sigma}_{ij}^M$ is shown by (2.16) in Proposition 2.1.

(b) An unbiased estimate $\hat{\Omega}_U$ of Ω is given by

$$\begin{aligned}\hat{\Omega}_{11}^U &= (m_1 - p_1 - 1) \left(1 - \sum_{i=2}^k \frac{p_i}{m_i - p_1 - 1} \right) \mathbf{W}_{11}^{-1} \\ &\quad + \sum_{i=2}^k (m_i - p_1 - p_i - 1) \mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii.1}^{-1} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}; \\ \hat{\Omega}_{1i}^U &= -(m_i - p_1 - p_i - 1) \mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii.1}^{-1}, \\ \hat{\Omega}_{ii}^U &= (m_i - p_1 - p_i - 1) \mathbf{W}_{ii.1}^{-1}, \quad i = 2, \dots, k.\end{aligned}$$

Proof. See Appendix 2B. □

2.3 The Invariant Haar Measures and Noninformative Priors of Ψ

2.3.1 The Invariant Haar measures

From Sun and Sun (2005a), noninformative priors of Ψ play an important role in getting appropriate estimates for the covariance matrix Σ and the precision matrix Ω for the complete data case. We will discuss noninformative priors of Ψ for incomplete data in this section. Define

$$\mathcal{G} = \left\{ \mathbf{A} \in R^{p \times p} \mid \mathbf{A} \text{ has a structure as (2.13)} \right\}. \quad (2.21)$$

Sun and Sun (2005a) showed that \mathcal{G} is a group with respect to matrix multiplication.

For any $i = 1, 2, \dots, k$, let

$$\Psi_{ii} = \begin{pmatrix} \psi_{i11} & 0 & \cdots & 0 \\ \psi_{i21} & \psi_{i22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi_{ip_i1} & \psi_{ip_i1} & \cdots & \psi_{ip_i p_i} \end{pmatrix}. \quad (2.22)$$

And for $i = 2, \dots, k$, let

$$\Psi_{i1} = \begin{pmatrix} \phi_{i11} & \phi_{i12} & \cdots & \phi_{i1 p_1} \\ \phi_{i21} & \phi_{i22} & \cdots & \phi_{i2 p_1} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{ip_i1} & \phi_{ip_i2} & \cdots & \phi_{ip_i p_1} \end{pmatrix}. \quad (2.23)$$

The left invariant Haar measure and the right invariant Haar measure of \mathcal{G} are given by

$$\nu_{\mathcal{G}}^l(d\Psi) \propto \frac{d\Psi}{\prod_{j=1}^{p_1} \psi_{1jj}^j \cdot \prod_{i=2}^k \prod_{j=1}^{p_i} \psi_{ijj}^{p_i+j}}, \quad (2.24)$$

$$\nu_{\mathcal{G}}^r(d\Psi) \propto \frac{d\Psi}{\prod_{j=1}^{p_1} \psi_{1jj}^{p-j+1} \cdot \prod_{i=2}^k \prod_{j=1}^{p_i} \psi_{ijj}^{p_i-j+1}}, \quad (2.25)$$

respectively. In addition, we can readily verify that $\nu_{\mathcal{G}}^r(d\Psi) = \nu_{\mathcal{G}}^l(d\Delta)$ and $\nu_{\mathcal{G}}^l(d\Psi) = \nu_{\mathcal{G}}^r(d\Delta)$ because $\Delta = \Psi^{-1}$.

2.3.2 The Jeffreys prior and a reference prior

The following proposition gives the Jeffreys prior and one reference prior of Ψ in the star-shaped model with missing data.

Proposition 2.5 *Consider the star-shaped model with missing data.*

(a) *The Jeffreys prior $\pi_J(d\Psi)$ of Ψ is the same as the right invariant Haar measure $\nu_{\mathcal{G}}^r(d\Psi)$ of \mathcal{G} given by (2.25);*

(b) *The reference prior of Ψ for the ordered group $\{\psi_{111}, (\psi_{121}, \psi_{122}), \dots, (\psi_{1p_1}, \dots, \psi_{1p_1 p_1}), (\phi_{211}, \dots, \phi_{21p_1}, \psi_{211}), \dots, (\phi_{k11}, \dots, \phi_{kp_k p_1}, \psi_{kp_k 1}, \dots, \psi_{kp_k p_k})\}$ is given by*

$$\pi_R(d\Psi) \propto \frac{d\Psi}{\prod_{i=1}^k \prod_{j=1}^{p_i} \psi_{ijj}}. \quad (2.26)$$

Proof. See Appendix 2C. □

It is interesting to note that the Jeffreys prior of Ψ in the model with missing data is the same as that in the model with complete data by Sun and Sun (2005a) and the same conclusion holds for the reference prior of Ψ .

Remark 2.2 *Unlike the case with complete data in Sun and Sun (2005a), it seems impossible to get the closed form of equivariant estimators of Σ or Ω with respect to \mathcal{G} in the star-shaped model with missing data.*

2.4 Properties of Posteriors of Ψ

In this section, we consider a class of priors of Ψ

$$p(\Psi) \propto \prod_{i=1}^k \prod_{j=1}^{p_i} \psi_{ijj}^{a_{ij}} \exp(-b_{ij} \psi_{ijj}^2), \quad (2.27)$$

where $a_{ij} \in R$, $b_{ij} \geq 0$, $j = 1, \dots, p_i$, $i = 1, \dots, k$. This class includes the left Haar invariant measure $\nu_{\mathcal{G}}^l(\Psi)$, the right Haar invariant measure $\nu_{\mathcal{G}}^r(\Psi)$ (the Jeffreys prior $\pi_J(\Psi)$), and the reference prior $\pi_R(\Psi)$.

We have the following posterior properties. The proof is straightforward and is omitted.

Theorem 2.1 *For the star-shaped model with missing data, the posterior $p(\Psi \mid \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k)$ under the prior $p(\Psi)$ in (2.27) has the following properties:*

(a) $p(\Psi \mid \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k)$ is proper if and only if $m_i + a_{ij} + 1 > 0$, $j = 1, \dots, p_i$, $i = 1, \dots, k$;

(b) $\Psi_{11}, (\Psi_{21}, \Psi_{22}), \dots, (\Psi_{k1}, \Psi_{kk})$ are mutually independent;

(c) For $i = 2, \dots, k$, conditional posterior distribution of Ψ_{i1} given Ψ_{ii} is

$$\Psi_{i1} \mid \Psi_{ii}, (\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) \sim N_{p_i, p_1}(-\Psi_{ii} \mathbf{W}_{i1} \mathbf{W}_{i1}^{-1}, \mathbf{I}_{p_i} \otimes \mathbf{W}_{i1}^{-1});$$

(d) For $i = 1, \dots, k$, the marginal posterior of Ψ_{ii} is

$$\Psi_{ii} \mid (\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) \sim \exp\left\{-\frac{1}{2} \text{tr}(\Psi_{ii} \mathbf{W}_{ii} \Psi_{ii}')\right\} \prod_{j=1}^{p_i} \psi_{ijj}^{m_i + a_{ij}} \exp(-b_{ij} \psi_{ijj}^2).$$

From Theorem 2.1, each of the posterior densities under the left Haar invariant measure $\nu_G^l(\Psi)$, the right Haar invariant measure $\nu_G^r(\Psi)$ (the Jeffreys prior $\pi_J(\Psi)$), and the reference prior $\pi_R(\Psi)$ will be proper. Specifically, the posterior distribution under the left Haar invariant measure $\nu_G^l(\Psi)$ is related to Wishart distribution as shown below.

Corollary 2.1 *If we take the left invariant Haar measure of the group \mathcal{G} , $\nu_G^l(d\Psi)$ as a prior, then the posterior distribution of Ψ in the star-shaped model with missing data satisfies*

(a) $\Psi_{11} \mathbf{W}_{11} \Psi_{11}' \mid (\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) \sim W_{p_1}(m_1, \mathbf{I}_{p_1});$

(b) For $i = 2, \dots, k$, $\Psi_{ii} \mathbf{W}_{ii} \Psi_{ii}' \mid (\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) \sim W_{p_i}(m_i - p_1, \mathbf{I}_{p_i}).$

2.5 Bayesian Estimators of Ω under the Stein Loss

The following lemma, which was proved in Sun and Sun (2005a), will be useful in finding Bayesian estimators of Ω with respect to the prior $p(\Psi)$.

Lemma 2.1 *Let \mathbf{G} be a scalar positive definite matrix and $\mathbf{C} = (c_{ij})_{m \times m}$ be its Cholesky decomposition. Assume that $\mathbf{Z} = (z_{ij})_{m \times m}$ is a random lower-triangular matrix with positive diagonal elements whose distribution follows*

$$[\mathbf{Z}] \propto \exp\left\{-\frac{1}{2}\text{tr}(\mathbf{Z}\mathbf{G}\mathbf{Z}')\right\} \prod_{i=1}^m z_{ii}^{a_i} \exp(-b_i z_{ii}^2). \quad (2.28)$$

(a) *If $a_i > 0$, $b_i \geq 0$, $i = 1, \dots, m$, then*

$$E(\mathbf{Z}'\mathbf{Z}) = (\mathbf{C}')^{-1} \text{diag}(\delta_1, \dots, \delta_m) \mathbf{C}^{-1}, \quad (2.29)$$

where $\delta_i = (a_i + 1)/(1 + 2b_i c_{ii}^{-2}) + m - i$, $i = 1, \dots, m$.

(b) *If $a_i > 1$, $b_i \geq 0$, $i = 1, \dots, m$, then*

$$E(\mathbf{Z}'\mathbf{Z})^{-1} = \mathbf{C} \text{diag}(\eta_1, \dots, \eta_m) \mathbf{C}', \quad (2.30)$$

where $\eta_1 = u_1$, $\eta_j = u_j \prod_{i=1}^{j-1} (1 + u_i)$, $j = 2, \dots, m$ with $u_i = (1 + 2b_i c_{ii}^{-2})/(a_i - 1)$, $i = 1, \dots, m$.

Now define

$$\mathbf{T} = \begin{pmatrix} \mathbf{T}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{W}_{21} \mathbf{W}_{211}^{-1} \mathbf{T}_{11} & \mathbf{T}_{22} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{W}_{31} \mathbf{W}_{311}^{-1} \mathbf{T}_{11} & \mathbf{0} & \mathbf{T}_{33} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{W}_{k1} \mathbf{W}_{k11}^{-1} \mathbf{T}_{11} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{kk} \end{pmatrix}, \quad (2.31)$$

and then we have

$$\mathbf{R} = \mathbf{T}^{-1} = \begin{pmatrix} \mathbf{T}_{11}^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{T}_{22}^{-1} \mathbf{W}_{21} \mathbf{W}_{211}^{-1} & \mathbf{T}_{22}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ -\mathbf{T}_{33}^{-1} \mathbf{W}_{31} \mathbf{W}_{311}^{-1} & \mathbf{0} & \mathbf{T}_{33}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{T}_{kk}^{-1} \mathbf{W}_{k1} \mathbf{W}_{k11}^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{T}_{kk}^{-1} \end{pmatrix}, \quad (2.32)$$

where \mathbf{T}_{ii} is Cholesky decomposition of \mathbf{W}_{ii-1} , $i = 1, 2, \dots, k$.

Theorem 2.2 *Suppose that $m_i + a_{ij} - 1 > 0$, $j = 1, \dots, p_i$, $i = 1, \dots, k$. Then under the Stein loss L_1^* , the Bayesian estimate of $\boldsymbol{\Omega}$ with respect to the prior $p(\boldsymbol{\Psi})$ in (2.27) is given by*

$$\hat{\boldsymbol{\Omega}}_1 = \mathbf{R}' \mathbf{B}_1 \mathbf{R}, \quad (2.33)$$

where \mathbf{R} is given by (2.32), $\mathbf{B}_1 = \text{diag}(\mathbf{B}_{11}, \mathbf{B}_{12}, \dots, \mathbf{B}_{1k})$, $\mathbf{B}_{11} = \mathbf{D} + \sum_{i=2}^k p_i \mathbf{T}'_{11} \mathbf{W}_{i11}^{-1} \mathbf{T}_{11}$ and $\mathbf{D} = \text{diag}(d_1, \dots, d_{p_1})$ with $d_j = (m_1 + a_{1j} + 1) / (1 + 2b_{1j} t_{1jj}^{-1}) + p_1 - j$, $j = 1, \dots, p_1$; $\mathbf{B}_{1i} = \text{diag}(b_{1i1}, \dots, b_{1ip_i})$ with $b_{1ij} = (m_i + a_{ij} + 1) / (1 + 2b_{ij} t_{ijj}^{-1}) + p_i - j$, $j = 1, \dots, p_i$, $i = 2, \dots, k$. And t_{ijj} is the j th diagonal element of \mathbf{T}_{ii} , $j = 1, \dots, p_i$, $i = 1, \dots, k$.

Proof. See Appendix 2D. □

By the same argument with Corollary 2 in Sun and Sun (2005a), we can get the best \mathcal{G} -equivariant estimator of $\boldsymbol{\Omega}$, which is the same as the Bayesian estimator with respect to the left Haar invariant measure $\nu_{\mathcal{G}}^l(\boldsymbol{\Psi})$ listed below.

Corollary 2.2 *Under the Stein loss L_1^* , the best \mathcal{G} -equivariant estimator of $\boldsymbol{\Omega}$ is the same as Bayesian estimator with respect to the left Haar invariant measure $\nu_{\mathcal{G}}^l(\boldsymbol{\Psi})$ and is given by*

$$\hat{\boldsymbol{\Omega}}_{1B} = \mathbf{R}' \mathbf{B}_{1B} \mathbf{R}, \quad (2.34)$$

where \mathbf{R} is given by (2.32), $\mathbf{B}_{1B} = \text{diag}(\mathbf{B}_{11B}, \mathbf{B}_{12B}, \dots, \mathbf{B}_{1kB})$, $\mathbf{B}_{11B} = \mathbf{D}_1 + \sum_{i=2}^k p_i \mathbf{T}'_{11} \mathbf{W}_{i11}^{-1} \mathbf{T}_{11}$ and $\mathbf{D}_1 = \text{diag}(d_{11}, \dots, d_{1p_1})$ with $d_{1j} = m_1 + p_1 - 2j + 1, j = 1, \dots, p_1$; $\mathbf{B}_{iB} = \text{diag}(b_{i1B}, \dots, b_{ip_iB})$ with $b_{ijB} = m_i - p_1 + p_i - 2j + 1, j = 1, \dots, p_i, i = 2, \dots, k$.

Remark 2.3 *It is well-known that the group of lower-triangular matrices is solvable and thus its subgroup \mathcal{G} is also solvable (see Bondar and Milnes (1981) for a survey). By Kiefer (1957), the best \mathcal{G} -equivariant estimator $\hat{\boldsymbol{\Omega}}_{1B}$ is also minimax with respect to the Stein loss L_1^* .*

In addition, we obtain Bayesian estimators of $\boldsymbol{\Omega}$ with respect to the right Haar invariant measure $\nu_{\mathcal{G}}^r(\boldsymbol{\Psi})$ (the Jeffreys prior $\pi_J(\boldsymbol{\Psi})$) and the reference prior $\pi_R(\boldsymbol{\Psi})$ in the following:

Corollary 2.3 *Consider the Stein loss L_1^* . (a) the Bayesian estimator $\hat{\boldsymbol{\Omega}}_{1J}$ of $\boldsymbol{\Omega}$ with respect to the Jeffreys prior $\pi_J(\boldsymbol{\Psi})$ is*

$$\hat{\boldsymbol{\Omega}}_{1J} = \mathbf{R}' \mathbf{B}_{1J} \mathbf{R}, \quad (2.35)$$

where

$$\begin{aligned} \mathbf{B}_{1J} &= \text{diag}(\mathbf{B}_{11J}, \mathbf{B}_{12J}, \dots, \mathbf{B}_{1kJ}), \\ \mathbf{B}_{11J} &= (m_1 - p + p_1) \mathbf{I}_{p_1} + \sum_{j=2}^k p_j \mathbf{T}'_{11} \mathbf{W}_{j11}^{-1} \mathbf{T}_{11}; \quad \mathbf{B}_{1iJ} = m_i \mathbf{I}_{p_i}, \quad i = 2, \dots, k. \end{aligned} \quad (2.36)$$

(b) The Bayesian estimator $\hat{\boldsymbol{\Omega}}_{1R}$ with respect to the reference prior $\pi_R(\boldsymbol{\Psi})$ is

$$\hat{\boldsymbol{\Omega}}_{1R} = \mathbf{R}' \mathbf{B}_{1R} \mathbf{R}, \quad (2.37)$$

where

$$\mathbf{B}_{1R} = \text{diag}(\mathbf{B}_{11R}, \mathbf{B}_{12R}, \dots, \mathbf{B}_{1kR}), \quad (2.38)$$

$$\begin{aligned}
\mathbf{B}_{11R} &= \mathbf{D}_{1R} + \sum_{j=2}^k p_j \mathbf{T}'_{11} \mathbf{W}_{j11}^{-1} \mathbf{T}_{11}; \\
\mathbf{B}_{1iR} &= \text{diag}(m_i + p_i - 1, m_i + p_i - 2, \dots, m_i), \quad i = 2, \dots, k; \\
\mathbf{D}_{1R} &= \text{diag}(m_1 + p_1 - 1, m_1 + p_1 - 2, \dots, m_i).
\end{aligned}$$

The MLE of $\boldsymbol{\Omega}$ given by (2.19) can be expressed as

$$\hat{\boldsymbol{\Omega}}_M = \mathbf{R}' \text{diag}(m_1 \mathbf{I}_{p_1}, m_2 \mathbf{I}_{p_2}, \dots, m_k \mathbf{I}_{p_k}) \mathbf{R},$$

and the unbiased estimate $\hat{\boldsymbol{\Omega}}_U$ of $\boldsymbol{\Omega}$ given by Proposition 2.4(b) can be expressed as

$\hat{\boldsymbol{\Omega}}_U = \mathbf{R}' \mathbf{U} \mathbf{R}$, where

$$\begin{aligned}
\mathbf{U} &= \text{diag}\left\{ (m_1 - p_1 - 1) \left(1 - \sum_{i=2}^k \frac{p_i}{m_i - p_1 - 1} \right) \mathbf{I}_{p_1}, (m_1 - p_1 - p_2 - 1) \mathbf{I}_{p_2}, \right. \\
&\quad \left. \dots, (m_k - p_1 - p_k - 1) \mathbf{I}_{p_k} \right\}.
\end{aligned}$$

Remark 2.4 *By Corollary 2.2, $\hat{\boldsymbol{\Omega}}_M$, $\hat{\boldsymbol{\Omega}}_U$, $\hat{\boldsymbol{\Omega}}_{1J}$ and $\hat{\boldsymbol{\Omega}}_{1R}$ are all inadmissible under the Stein loss L_1^* because they are all equivariant with respect to the group \mathcal{G} .*

Note that any estimate of $\boldsymbol{\Omega}$ having the form of $\mathbf{R}' \text{diag}(a_1, \dots, a_p) \mathbf{R}$ will be equivariant with respect to the group \mathcal{G} , where a_i is a constant, $i = 1, \dots, p$. Thus, by Corollary 2.2, any estimator having the form like $\mathbf{R}' \text{diag}(a_1, \dots, a_p) \mathbf{R}$ will be inadmissible under the Stein loss. However, it is unclear whether $\mathbf{R}' \text{diag}(a_1, \dots, a_p) \mathbf{R}$ is a Bayesian estimate, which also is different from the case with complete data in Sun and Sun (2005a).

2.6 Bayesian Estimators of Ω under the Entropy Loss

Theorem 2.3 *Suppose that $m_i + a_{ij} - 1 > 0$, $j = 1, \dots, p_i$, $i = 1, \dots, k$. Then under the entropy loss L_2^* , the Bayesian estimate of Ω with respect to the prior $p(\Psi)$ in (2.27) is given by*

$$\hat{\Omega}_2 = \mathbf{R}'\mathbf{B}_2\mathbf{R}, \quad (2.39)$$

where \mathbf{R} is given by (2.32), $\mathbf{B}_2 = \text{diag}(\mathbf{B}_{21}, \mathbf{B}_{22}, \dots, \mathbf{B}_{2k})$ with

$$b_{211} = \frac{1}{u_{11}}, \quad b_{21j} = \frac{1}{u_{1j}} \prod_{t=1}^{j-1} \frac{1}{1 + u_{1t}}, \quad j = 2, \dots, p_1 \quad (2.40)$$

and

$$\begin{aligned} b_{2i1} &= \frac{1}{\{1 + \text{tr}(\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1})\}u_{i1}}, \\ b_{2ij} &= \frac{1}{\{1 + \text{tr}(\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1})\}u_{ij}} \prod_{t=1}^{j-1} \frac{1}{1 + u_{it}}, \quad j = 2, \dots, p_i, \quad i = 2, \dots, k. \end{aligned} \quad (2.41)$$

Here $u_{ij} = (1 + 2b_{ij}t_{ij}^{-2})/(m_i + a_{ij} - 1)$ with t_{ij} being the j th diagonal element of \mathbf{T}_{ii} , $j = 1, \dots, p_i$, $i = 1, \dots, k$.

Proof. See Appendix 2E. □

Corollary 2.4 *Under the entropy loss L_2^* , the best \mathcal{G} -equivariant estimator of Ω is the same as the Bayesian estimator with respect to the left Haar invariant measure $\nu_{\mathcal{G}}^l(\Psi)$ and is given by*

$$\hat{\Omega}_{2B} = \mathbf{R}'\mathbf{B}_{2B}\mathbf{R}, \quad (2.42)$$

where \mathbf{R} is given by (2.32), $\mathbf{B}_{2B} = \text{diag}(\mathbf{B}_{21B}, \mathbf{B}_{22B}, \dots, \mathbf{B}_{2kB})$, and $\mathbf{B}_{2iB} = \text{diag}(d_{i1B}, \dots, d_{ip_iB})$ with

$$d_{ijB} = \begin{cases} \frac{(m_1 - j - 1)(m_1 - j)}{m_1 - 1}, & \text{if } i = 1, j = 1, \dots, p_1, \\ \frac{(m_i - p_1 - j - 1)(m_i - p_1 - j)}{(m_i - 1)\{1 + \text{tr}(\mathbf{B}_{21}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1}\mathbf{T}_{11})\}}, & \text{if } i = 2, \dots, k, j = 1, \dots, p_i. \end{cases} \quad (2.43)$$

Remark 2.5 Similar to Remark 2.3, the best \mathcal{G} -equivariant estimator $\hat{\mathbf{\Omega}}_{2B}$ is also minimax with respect to the entropy loss L_2^* .

Corollary 2.5 Consider the entropy loss L_2^* . (a) The Bayesian estimator $\hat{\mathbf{\Omega}}_{2J}$ of $\mathbf{\Omega}$ with respect to the Jeffreys prior $\pi_J(\mathbf{\Psi})$ is

$$\hat{\mathbf{\Omega}}_{2J} = \mathbf{R}'\mathbf{B}_{2J}\mathbf{R}, \quad (2.44)$$

where

$$\mathbf{B}_{2J} = \text{diag}(\mathbf{B}_{21J}, \mathbf{B}_{22J}, \dots, \mathbf{B}_{2kJ}), \quad (2.45)$$

$$\mathbf{B}_{21J} = (m_1 - p - 1)\mathbf{I}_{p_1}; \quad \mathbf{B}_{2iJ} = \frac{(m_1 - p - 1)(m_i - p_i - 1)}{\text{tr}(\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1}\mathbf{T}_{11}) + m_1 - p - 1}\mathbf{I}_{p_i}, \quad i = 2, \dots, k.$$

(b) The Bayesian estimator $\hat{\mathbf{\Omega}}_{2R}$ with respect to the reference prior $\pi_R(\mathbf{\Psi})$ is

$$\hat{\mathbf{\Omega}}_{2R} = \mathbf{R}'\mathbf{B}_{2R}\mathbf{R}, \quad (2.46)$$

where

$$\mathbf{B}_{2R} = \text{diag}(\mathbf{B}_{21R}, \mathbf{B}_{22R}, \dots, \mathbf{B}_{2kR}), \quad (2.47)$$

with

$$\begin{aligned} \mathbf{B}_{21R} &= \text{diag}\left(m_1 - 2, \frac{(m_1 - 2)^2}{m_1 - 1}, \dots, \frac{(m_1 - 2)^{p_1}}{(m_1 - 1)^{p_1 - 1}}\right); \\ \mathbf{B}_{2iR} &= \text{diag}\left(m_i - 2, \frac{(m_i - 2)^2}{m_i - 1}, \dots, \frac{(m_i - 2)^{p_i}}{(m_i - 1)^{p_i - 1}}\right) \\ &\quad / \{1 + \text{tr}(\mathbf{B}_{21R}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1}\mathbf{T}_{11})\}, \quad i = 2, \dots, k. \end{aligned}$$

Remark 2.6 *Similar to Remark 2.4, the MLE $\hat{\Omega}_M$ and the unbiased estimator $\hat{\Omega}_U$, $\hat{\Omega}_{2J}$ and $\hat{\Omega}_{2R}$ are still inadmissible under the entropy loss L_2^* .*

2.7 Bayesian Estimators of Ω under the Symmetric Loss

We need the following lemma, which is a direct result of Lemma 2.2 in Eaton and Olkin (1987).

Lemma 2.2 *Let $\mathcal{A} = \{\mathbf{A} \in \mathbf{R}^{p \times p} \mid \mathbf{A} \text{ is lower-triangular with positive diagonal elements}\}$. If Δ and Λ are both positive definite, then*

$$\min_{\mathbf{A} \in \mathcal{A}} \left\{ \text{tr}(\mathbf{A}\Delta\mathbf{A}') + \text{tr}((\mathbf{A}')^{-1}\Lambda\mathbf{A}^{-1}) \right\} = 2\text{tr}(\Lambda^{1/2}\Delta\Lambda^{1/2})^{1/2}$$

is achieved by taking \mathbf{A} as the inverse of Cholesky decomposition of $\Lambda^{-1/2}(\Lambda^{1/2}\Delta\Lambda^{1/2})^{1/2}\Lambda^{-1/2}$. Specifically, if Δ and Λ are both diagonal, then the minimum will be achieved at $\mathbf{A} = \Lambda^{1/4}\Delta^{-1/4}$.

Theorem 2.4 *Suppose that $m_i + a_{ij} - 1 > 0$, $j = 1, \dots, p_i$, $i = 1, \dots, k$. Then under the symmetric loss L_3^* , the Bayesian estimator of Ω with respect to the prior $p(\Psi)$ in (2.27) is given by*

$$\hat{\Omega}_3 = \mathbf{R}'\mathbf{B}_3\mathbf{R}, \tag{2.48}$$

where \mathbf{R} is given by (2.32), $\mathbf{B}_3 = \text{diag}(\mathbf{B}_{31}, \mathbf{B}_{32}, \dots, \mathbf{B}_{3k})$ with $\mathbf{B}_{31} = \mathbf{B}_{11}^{1/2}(\mathbf{B}_{11}^{1/2}\mathbf{B}_{21}^{-1}\mathbf{B}_{11}^{1/2})^{-1/2}\mathbf{B}_{11}^{1/2}$ and $\mathbf{B}_{3i} = \mathbf{B}_{1i}^{1/2}\mathbf{B}_{2i}^{1/2}$, $i = 2, \dots, k$, where \mathbf{B}_{1i} and \mathbf{B}_{2i} are given by Theorem 2.2 and Theorem 2.3, respectively, $i = 1, 2, \dots, k$.

Proof. See Appendix 2F. □

Corollary 2.6 *Under the symmetric loss L_3^* , the best \mathcal{G} -equivariant estimator of Ω is the same as the Bayesian estimator with respect to the left Haar invariant measure $\nu_{\mathcal{G}}^l(\Psi)$ and is given by*

$$\hat{\Omega}_{3B} = \mathbf{R}'\mathbf{B}_{3B}\mathbf{R}, \quad (2.49)$$

where $\mathbf{B}_{3B} = \text{diag}(\mathbf{B}_{31B}, \mathbf{B}_{32B}, \dots, \mathbf{B}_{3kB})$ with $\mathbf{B}_{31B} = \mathbf{B}_{11B}^{1/2}(\mathbf{B}_{11B}^{1/2}\mathbf{B}_{21B}^{-1}\mathbf{B}_{11B}^{1/2})^{-1/2}\mathbf{B}_{11B}^{1/2}$ and $\mathbf{B}_{3iB} = \mathbf{B}_{1iB}^{1/2}\mathbf{B}_{2iB}^{1/2}$, $i = 2, \dots, k$, where \mathbf{B}_{1iB} and \mathbf{B}_{2iB} are given by Corollary 2.2 and Corollary 2.4, respectively, $i = 1, 2, \dots, k$.

Remark 2.7 *Similar to Remarks 2.3 and 2.5, the best \mathcal{G} -equivariant estimator $\hat{\Omega}_2$ is also minimax with respect to the symmetric loss L_3^* .*

Corollary 2.7 *Consider the symmetric loss L_3^* . (a) The Bayesian estimator $\hat{\Omega}_{3J}$ of Ω with respect to the Jeffreys prior $\pi_J(\Psi)$ is*

$$\hat{\Omega}_{3J} = \mathbf{R}'\mathbf{B}_{3J}\mathbf{R}, \quad (2.50)$$

where

$$\mathbf{B}_{3J} = \mathbf{B}_{1J}^{1/2}(\mathbf{B}_{1J}^{1/2}\mathbf{B}_{2J}^{-1}\mathbf{B}_{1J}^{1/2})^{-1/2}\mathbf{B}_{1J}^{1/2}$$

with \mathbf{B}_{1J} and \mathbf{B}_{2J} being given by (2.36) and (2.45), respectively.

(b) The Bayesian estimate $\hat{\Omega}_{3R}$ with respect to the reference prior $\pi_R(\Psi)$ is

$$\hat{\Omega}_{3R} = \mathbf{R}'\mathbf{B}_{3R}\mathbf{R}, \quad (2.51)$$

where

$$\mathbf{B}_{3R} = \mathbf{B}_{1R}^{1/2}(\mathbf{B}_{1R}^{1/2}\mathbf{B}_{2R}^{-1}\mathbf{B}_{1R}^{1/2})^{-1/2}\mathbf{B}_{1R}^{1/2}$$

with \mathbf{B}_{1R} and \mathbf{B}_{2R} being given by (2.38) and (2.47), respectively.

Remark 2.8 *Similar to Remark 2.2 and 2.4, $\hat{\Omega}_M$, $\hat{\Omega}_U$, $\hat{\Omega}_{3J}$, and $\hat{\Omega}_{3R}$ are all inadmissible under the symmetric loss L_3^* .*

2.8 Estimating the Covariance Matrix

As immediate corollaries of our results on estimating the precision matrix, we now list the results for estimating covariance matrix under the star-shaped model with missing data.

Theorem 2.5 *Under the loss $L_i, i = 1, 2, 3$, the Bayesian estimator of Σ is given by*

$$\hat{\Sigma}_i = \mathbf{T}\mathbf{B}_i^{-1}\mathbf{T}', \quad (2.52)$$

where \mathbf{T} is given by (2.31) and $\mathbf{B}_i, i = 1, 2, 3$ is shown by Theorems 2.2, 2.3, and 2.4, respectively.

Corollary 2.8 *Under the loss $L_i, i = 1, 2, 3$, the \mathcal{G} -equivariant estimator of Σ , which also is the Bayesian estimator with respect to the left Haar invariant measure $\nu_{\mathcal{G}}^l(\Psi)$, is given by*

$$\hat{\Sigma}_{iB} = \mathbf{T}\mathbf{B}_{iB}^{-1}\mathbf{T}', \quad (2.53)$$

where \mathbf{T} is given by (2.31) and $\mathbf{B}_{iB}, i = 1, 2, 3$ is shown by Corollaries 2.2, 2.4, and 2.6, respectively.

Corollary 2.9 For $i = 1, 2, 3$, the Bayesian estimator of Σ with respect to the Jeffreys prior $\pi_J(\Psi)$ under the loss L_i is

$$\hat{\Sigma}_{iJ} = \mathbf{T}\mathbf{B}_{iJ}^{-1}\mathbf{T}',$$

and the Bayesian estimate of Σ with respect to the reference prior $\pi_R(\Psi)$ under the loss L_i is

$$\hat{\Sigma}_{iR} = \mathbf{T}\mathbf{B}_{iR}^{-1}\mathbf{T}',$$

where \mathbf{B}_{iJ} and \mathbf{B}_{iR} are shown by Corollaries 2.3, 2.5, and 2.7, respectively.

Remark 2.9 The MLE $\hat{\Sigma}_M$, the unbiased estimator $\hat{\Sigma}_U$, and the Bayesian estimates $\hat{\Sigma}_{iJ}$ and $\hat{\Sigma}_{iR}$ are all inadmissible under the loss L_i , $i = 1, 2, 3$.

Remark 2.10 The best \mathcal{G} -equivariant estimator $\hat{\Sigma}_i$ is minimax with respect to the loss L_i , $i = 1, 2, 3$.

2.9 Simulation Results

In this section, we will compare the risks of MLE $\hat{\Omega}_M$, the unbiased estimator $\hat{\Omega}_U$, the best equivariant estimator $\hat{\Omega}_{iB}$, the Bayesian estimate $\hat{\Omega}_{iJ}$, the Bayesian estimate $\hat{\Omega}_{iR}$ under each L_i^* , $i = 1, 2, 3$. Each risk will be denoted as $R_{iM}, R_{iU}, R_{iB}, R_{iJ}, R_{iR}$ under L_i^* , respectively.

Unlike the model with complete data studied by Sun and Sun (2005a), it is hard to derive a closed form expression for the risks of the above estimates under any L_i^* . So we compare their risks by simulation. Because all of these estimators are

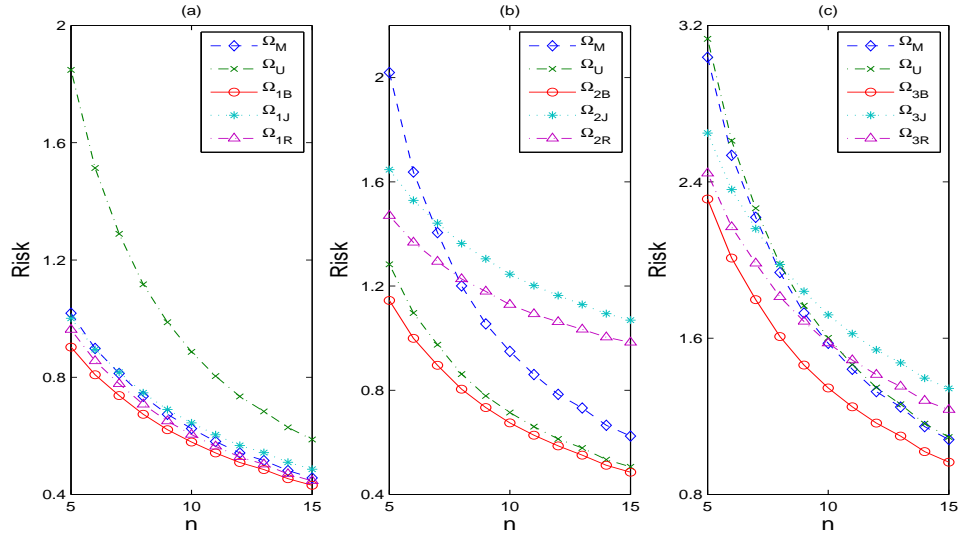


Figure 2.1: Risk Comparisons when $(p_1, p_2, p_3) = (2, 1, 1)$, $(n_1, n_2, n_3) = (3, 4, 5)$ and $5 \leq n \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

equivariant, we may take $\Sigma = \mathbf{I}_p$ in simulation. Some simulation results are plotted in Figures 2.1-2.9. From the simulation study, the improvement over the risks of $\hat{\Omega}_M$, $\hat{\Omega}_U$, $\hat{\Omega}_{iJ}$ and $\hat{\Omega}_{iR}$ by $\hat{\Omega}_{iB}$ under all three losses is quite substantial.

When $b_{ij} \neq 0$, then the corresponding Bayesian estimate $\hat{\Omega}_{Bi}$ of Ω with respect to the prior (2.27) under the loss L_i^* is not equivariant, and thus its risk depends on the covariance matrix Σ . We did a small study of risk comparison between the best equivariant estimate $\hat{\Omega}_{iB}$ and the Bayesian estimate $\hat{\Omega}_{Bi}$ with $a_{ij} = 2$ and $b_{ij} = 1$ when $(p_1, p_2, p_3) = (2, 2, 3)$ and $(n, n_1, n_2, n_3) = (10, 3, 5, 4)$. Figure 2.10 shows the results by taking $\Sigma = \text{diag}(\lambda, 1, 1, 1, 1, 1, 1)$ for the loss function L_i^* , $i = 1, 2, 3$. We can see that when λ is close to one, the improvement over the risks of the best equivariant estimate $\hat{\Omega}_{iB}$ by the Bayesian estimate $\hat{\Omega}_{Bi}$ under all three losses is quite substantial.

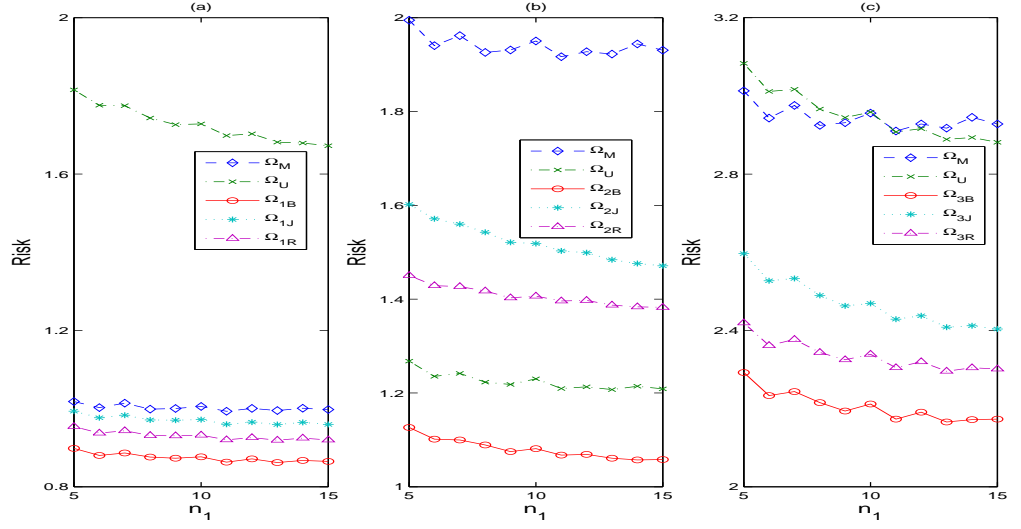


Figure 2.2: Risk Comparisons when $(p_1, p_2, p_3) = (2, 1, 1)$, $(n, n_2, n_3) = (5, 4, 5)$ and $5 \leq n_1 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

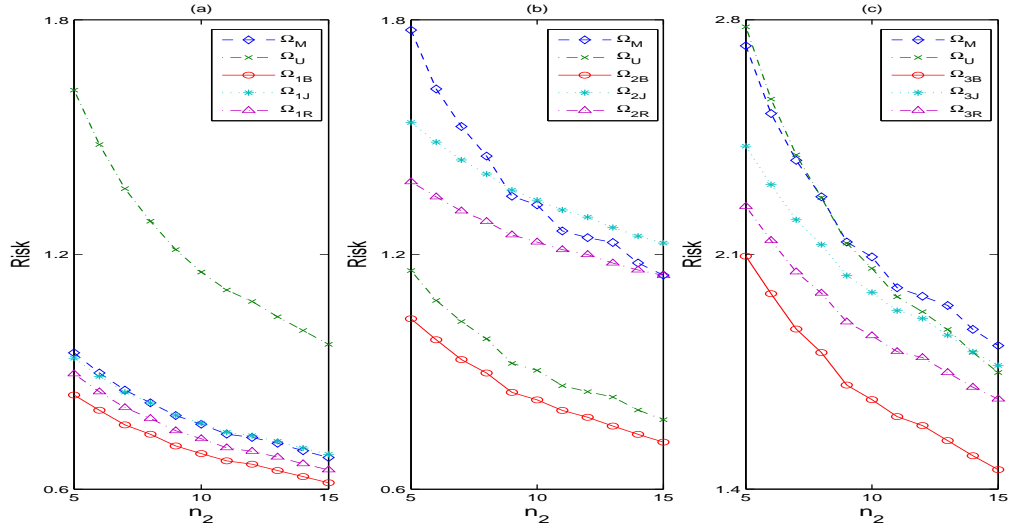


Figure 2.3: Risk Comparisons when $(p_1, p_2, p_3) = (2, 1, 1)$, $(n, n_1, n_3) = (5, 4, 5)$ and $5 \leq n_2 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

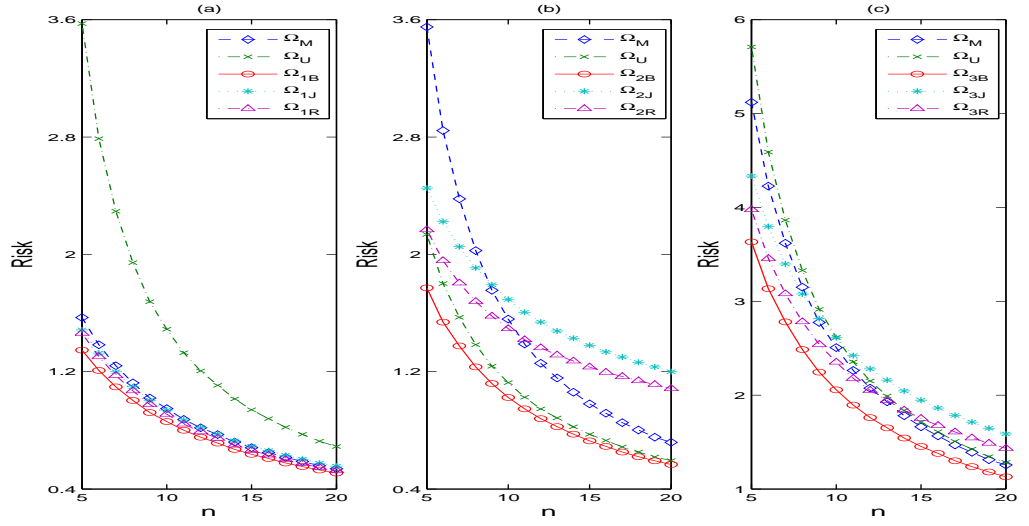


Figure 2.4: Risk Comparisons when $(p_1, p_2, p_3) = (2, 2, 1)$, $(n_1, n_2, n_3) = (3, 5, 4)$ and $5 \leq n \leq 20$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

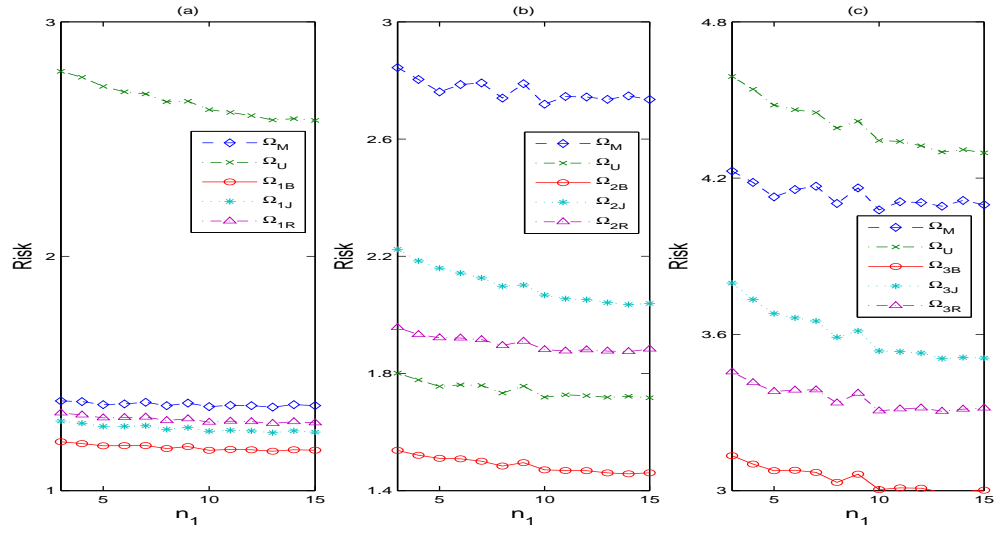


Figure 2.5: Risk Comparisons when $(p_1, p_2, p_3) = (2, 2, 1)$, $(n_1, n_2, n_3) = (6, 5, 4)$ and $3 \leq n_1 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

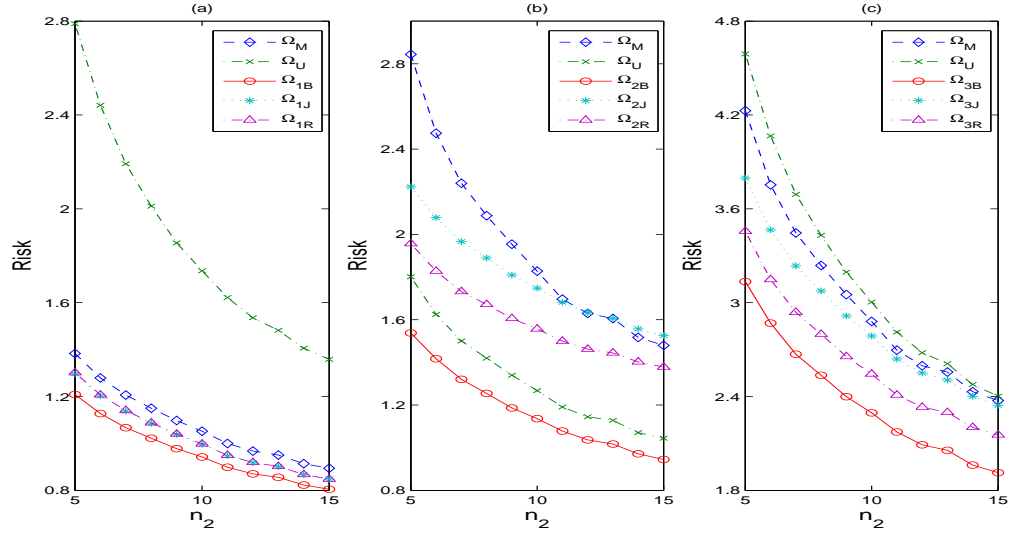


Figure 2.6: Risk Comparisons when $(p_1, p_2, p_3) = (2, 2, 1)$, $(n_1, n_2, n_3) = (6, 3, 4)$ and $5 \leq n_2 \leq 15$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

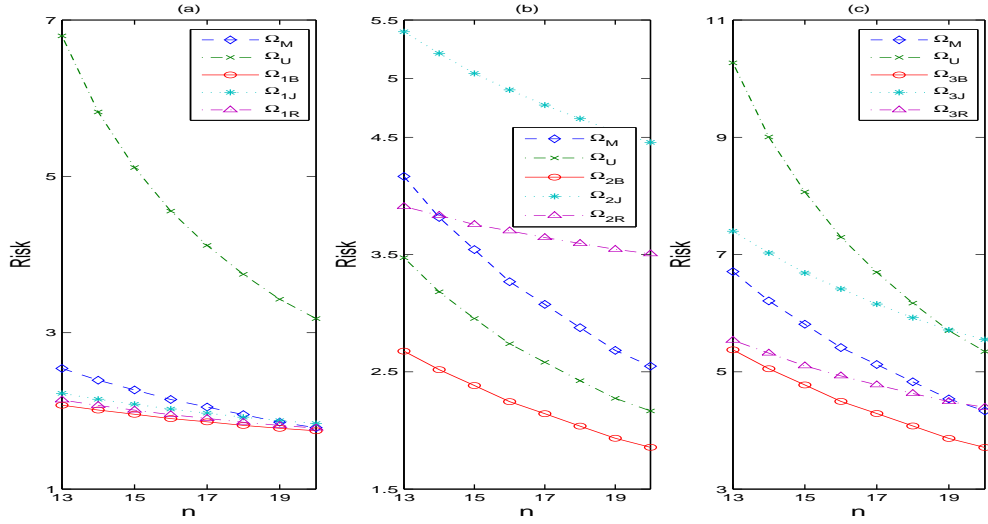


Figure 2.7: Risk Comparisons when $(p_1, p_2, p_3, p_4, p_5) = (2, 2, 3, 4, 1)$, $(n_1, n_2, n_3, n_4, n_5) = (3, 5, 6, 7, 4)$ and $13 \leq n \leq 20$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

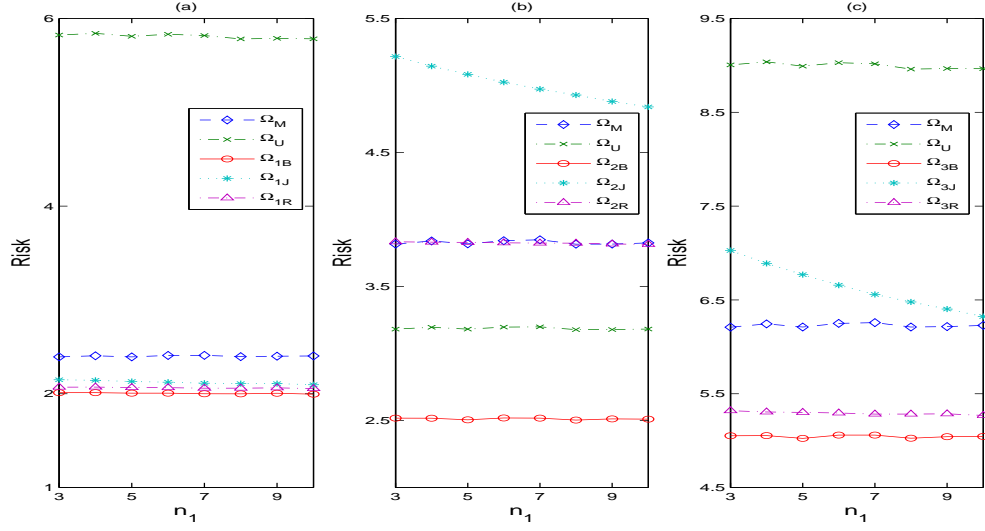


Figure 2.8: Risk Comparisons when $(p_1, p_2, p_3, p_4, p_5) = (2, 2, 3, 4, 1)$, $(n, n_2, n_3, n_4, n_5) = (14, 5, 6, 7, 4)$ and $3 \leq n_1 \leq 10$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

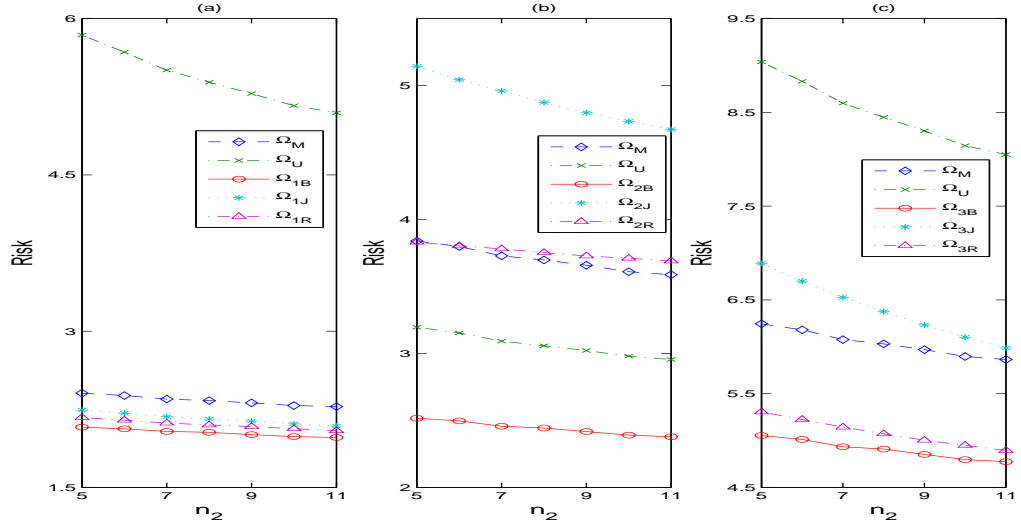


Figure 2.9: Risk Comparisons when $(p_1, p_2, p_3, p_4, p_5) = (2, 2, 3, 4, 1)$, $(n, n_1, n_3, n_4, n_5) = (14, 4, 6, 7, 4)$ and $5 \leq n_2 \leq 11$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

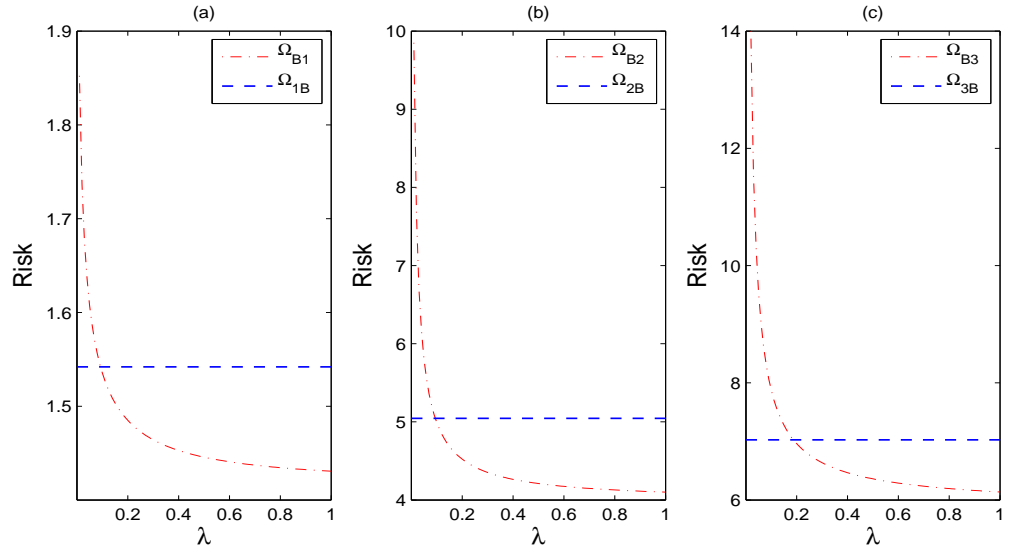


Figure 2.10: Risk Comparisons between Ω_{iB} and Ω_{Bi} when $(p_1, p_2, p_3) = (2, 2, 3)$, $(n, n_1, n_2, n_3) = (10, 3, 5, 4)$ and $\Sigma = \text{diag}(\lambda, 1, 1, 1, 1, 1, 1)$. (a) under L_1^* , (b) under L_2^* , and (c) under L_3^* .

2.10 Concluding Remarks

This chapter deals with the problem of estimating the covariance matrix and the precision matrix under the three common loss functions in the star-shaped model with missing data. Using a type of Cholesky decomposition of the precision matrix $\Omega = \Psi'\Psi$, we easily obtained the MLEs of the covariance matrix and the precision matrix. Also, by introducing a class of priors of Ψ , we get the closed forms of Bayesian estimators of Ω under the Stein loss, entropy loss and symmetric loss, respectively. This method is quite powerful in estimating the covariance matrix or the precision matrix.

Although our sample plan is restricted to taking observations from $\mathbf{X}, \mathbf{X}_1, (\mathbf{X}'_1, \mathbf{X}'_i)'$, $i = 2, \dots, k$, which is popular in economic studies, we can deal with other cases such

as taking observations from $\mathbf{X}, \mathbf{X}_1, (\mathbf{X}'_1, \mathbf{X}'_{i_1}, \mathbf{X}'_{i_2}, \dots, \mathbf{X}'_{i_j})'$, $2 \leq i_1 < \dots < i_j \leq k$ by applying the similar method. Of these cases, the monotone missing data pattern is not required when the covariance matrix has a special structure, which is different from the case of the covariance matrix with no restriction by Anderson (1957) and Liu (1999) and so on. In addition, for convenience, we assume that the sample sizes satisfying $n > p$, $n_1 > p_1$ and $n_i > p_1 + p_i$, $i = 2, \dots, k$ in this chapter. The essential conditions are $m_1 > p_1 + 1$ and $m_i > p_1 + p_i + 1$.

The investigation on a star-shaped model with missing data is, nevertheless, far from being complete, and there are many important and interesting questions to be further studied. An interesting but difficult problem is whether the best equivariant estimate $\hat{\boldsymbol{\Sigma}}_{iB}$ is admissible under the corresponding loss L_i .

Since the decomposition $\boldsymbol{\Omega} = \boldsymbol{\Psi}'\boldsymbol{\Psi}$ is quite powerful in estimating the covariance matrix and the precision matrix, we expect more results in this area by considering priors in terms of $\boldsymbol{\Psi}$ in the future.

Combining a star-shaped model with a Bayesian spatial model discussed in Chapters 3 and 4, one may consider to establish a new one, Bayesian multivariate spatial model with conditional independence structure, which may be applied to our current study of the MOFEP.

2.11 Appendix

Appendix 2A Proof of Proposition 2.3

(a) In fact, we will show $E(\hat{\Sigma}_{ii}^M) \neq \Sigma_{ii}$, $i = 2, \dots, k$. By Proposition 2.1,

$$\begin{aligned}\hat{\Sigma}_{ii}^M &= \frac{\mathbf{W}_{i22} - \mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}}{m_i} + \frac{\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}}{m_1} \\ &= \frac{\mathbf{W}_{i22}}{m_i} + \left(\frac{1}{m_1} - \frac{1}{m_i}\right)\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i} + \frac{\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}(\sum_{t=1}^k \mathbf{V}_{t11} - \mathbf{V}_{i11})\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}}{m_1}.\end{aligned}$$

Obviously, from (2.11), $E(\mathbf{W}_{i22}) = E(\mathbf{S}_{ii} + \mathbf{V}_{i22}) = m_i \Sigma_{ii}$, $i = 2, \dots, k$ and

$$E\{\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}(\sum_{t=1}^k \mathbf{V}_{t11} - \mathbf{V}_{i11})\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}\} = (m_1 - m_i)E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\Sigma_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}).$$

Because

$$\begin{pmatrix} \mathbf{S}_{11} & \mathbf{S}_{1i} \\ \mathbf{S}_{i1} & \mathbf{S}_{ii} \end{pmatrix} \sim W_{p_1+p_i} \left(n, \begin{pmatrix} \Sigma_{11} & \Sigma_{1i} \\ \Sigma_{i1} & \Sigma_{ii} \end{pmatrix} \right),$$

it follows

$$\mathbf{S}_{i1} \mid \mathbf{S}_{11} \sim N_{p_i, p_1}(\Sigma_{i1}\Sigma_{11}^{-1}\mathbf{S}_{11}, \Sigma_{ii-1} \otimes \mathbf{S}_{11}). \quad (2.54)$$

Similarly,

$$\mathbf{V}_{i12} \mid \mathbf{V}_{i11} \sim N_{p_i, p_1}(\Sigma_{i1}\Sigma_{11}^{-1}\mathbf{V}_{i11}, \Sigma_{ii-1} \otimes \mathbf{V}_{i11}), \quad (2.55)$$

and thus we have

$$(\mathbf{S}_{i1} + \mathbf{V}_{i12}) \mid (\mathbf{S}_{11}, \mathbf{V}_{i11}) \sim N_{p_i, p_1}(\Sigma_{i1}\Sigma_{11}^{-1}(\mathbf{S}_{11} + \mathbf{V}_{i11}), \Sigma_{ii-1} \otimes (\mathbf{S}_{11} + \mathbf{V}_{i11})). \quad (2.56)$$

So,

$$E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}) = E\{E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i} \mid (\mathbf{S}_{11}, \mathbf{V}_{i11}))\}$$

$$\begin{aligned}
&= E\{tr(\mathbf{W}_{i11}^{-1}\mathbf{W}_{i11})\boldsymbol{\Sigma}_{ii\cdot}\} + E(\boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{W}_{i11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i}) \\
&= tr(\mathbf{I}_{p_1})\boldsymbol{\Sigma}_{ii\cdot} + \boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}E(\mathbf{W}_{i11})\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i} \\
&= p_1\boldsymbol{\Sigma}_{ii\cdot} + m_i\boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i}
\end{aligned} \tag{2.57}$$

and

$$\begin{aligned}
&E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\boldsymbol{\Sigma}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}) = E\{E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\boldsymbol{\Sigma}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i} \mid (\mathbf{S}_{11}, \mathbf{V}_{i11}))\} \\
&= E\{tr(\mathbf{W}_{i11}^{-1}\boldsymbol{\Sigma}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i11})\boldsymbol{\Sigma}_{ii\cdot}\} + E(\boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}\mathbf{W}_{i11}\mathbf{W}_{i11}^{-1}\boldsymbol{\Sigma}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{i11}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i}) \\
&= tr\{E(\mathbf{S}_{11} + \mathbf{V}_{i11})^{-1}\boldsymbol{\Sigma}_{11}\}\boldsymbol{\Sigma}_{ii\cdot} + \boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i} \\
&= \frac{p_1}{m_i - p_1 - 1}\boldsymbol{\Sigma}_{ii\cdot} + \boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i}.
\end{aligned} \tag{2.58}$$

Hence,

$$\begin{aligned}
E(\hat{\boldsymbol{\Sigma}}_{ii}^M) &= \frac{E(\mathbf{W}_{i22})}{m_i} + \left(\frac{1}{m_1} - \frac{1}{m_i}\right)E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}) \\
&\quad + \frac{m_1 - m_i}{m_1}E(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}\boldsymbol{\Sigma}_{11}\mathbf{W}_{i11}^{-1}\mathbf{W}_{1i}) \\
&= \left[1 + p_1 \left\{\frac{1}{m_1} - \frac{1}{m_i} + \frac{m_1 - m_i}{m_1(m_i - p_1 - 1)}\right\}\right]\boldsymbol{\Sigma}_{ii} \\
&\quad + \left[1 - \frac{m_i}{m_1} - p_1 \left\{\frac{1}{m_1} - \frac{1}{m_i} + \frac{m_1 - m_i}{m_1(m_i - p_1 - 1)}\right\}\right]\boldsymbol{\Sigma}_{i1}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{1i},
\end{aligned} \tag{2.59}$$

which is not equal to $\boldsymbol{\Sigma}_{ii}$.

(b) Because

$$\begin{pmatrix} \mathbf{S}_{11} + \mathbf{V}_{i11} & \mathbf{S}_{1i} + \mathbf{V}_{i12} \\ \mathbf{S}_{i1} + \mathbf{V}_{i21} & \mathbf{S}_{ii} + \mathbf{V}_{i22} \end{pmatrix} \sim W_{p_1+p_i} \left(n + n_i, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{1i} \\ \boldsymbol{\Sigma}_{i1} & \boldsymbol{\Sigma}_{ii} \end{pmatrix} \right), \tag{2.60}$$

we have $\mathbf{W}_{ii\cdot} \sim W_{p_i}(m_i - p_1, \boldsymbol{\Sigma}_{ii\cdot})$, and thus

$$E(\hat{\boldsymbol{\Omega}}_{ii}^M) = \frac{m_i}{m_i - p_1 - p_i - 1}\boldsymbol{\Sigma}_{ii\cdot}^{-1} \neq \boldsymbol{\Omega}_{ii}, \quad i = 2, \dots, k,$$

which proves the second part. \square

Appendix 2B Proof of Proposition 2.4

(a) Obviously, $E(\hat{\Sigma}_{11}^M) = E(\mathbf{S}_{11} + \sum_{i=1}^k \mathbf{V}_{i11})/m_1 = \Sigma_{11}$. By (2.56),

$$E(\mathbf{S}_{i1} + \mathbf{V}_{i21} \mid \mathbf{S}_{11}, \mathbf{V}_{i11}) = \Sigma_{i1} \Sigma_{11}^{-1} (\mathbf{S}_{11} + \mathbf{V}_{i11})$$

and thus $E(\hat{\Sigma}_{i1}^M) = \Sigma_{i1}$ because of $E(\mathbf{S}_{11} + \mathbf{V}_{i11}) = m_i \Sigma_{11}$, $i = 2, \dots, k$. In addition,

for any $1 < i < j \leq k$,

$$\begin{aligned} \begin{pmatrix} \mathbf{S}_{i1} \\ \mathbf{S}_{j1} \end{pmatrix} \mid \mathbf{S}_{11} &\sim N_{p_i+p_j, p_1} \left(\begin{pmatrix} \Sigma_{i1} \\ \Sigma_{j1} \end{pmatrix} \Sigma_{11}^{-1} \mathbf{S}_{11}, \begin{pmatrix} \Sigma_{ii \cdot 1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{jj \cdot} \end{pmatrix} \otimes \mathbf{S}_{11} \right), \\ \mathbf{V}_{i21} \mid \mathbf{V}_{i11} &\sim N_{p_i, p_1}(\Sigma_{i1} \Sigma_{11}^{-1} \mathbf{V}_{i11}, \Sigma_{ii \cdot 1} \otimes \mathbf{V}_{i11}), \\ \mathbf{V}_{j21} \mid \mathbf{V}_{j11} &\sim N_{p_j, p_1}(\Sigma_{j1} \Sigma_{11}^{-1} \mathbf{V}_{j11}, \Sigma_{jj \cdot} \otimes \mathbf{V}_{j11}); \end{aligned} \quad (2.61)$$

then we get

$$\begin{aligned} &\begin{pmatrix} \mathbf{S}_{i1} + \mathbf{V}_{i21} \\ \mathbf{S}_{j1} + \mathbf{V}_{j21} \end{pmatrix} \mid (\mathbf{S}_{11}, \mathbf{V}_{i11}, \mathbf{V}_{j11}) \\ &\sim N_{p_i+p_j, p_1} \left(\begin{pmatrix} \Sigma_{i1} \Sigma_{11}^{-1} (\mathbf{s}_{11} + \mathbf{v}_{i11}) \\ \Sigma_{j1} \Sigma_{11}^{-1} (\mathbf{s}_{11} + \mathbf{v}_{j11}) \end{pmatrix}, \begin{pmatrix} \Sigma_{ii \cdot 1} \otimes (\mathbf{s}_{11} + \mathbf{v}_{i11}) & \mathbf{0} \\ \mathbf{0} & \Sigma_{jj \cdot} \otimes (\mathbf{s}_{11} + \mathbf{v}_{j11}) \end{pmatrix} \right). \end{aligned} \quad (2.62)$$

Also, by $E(\mathbf{W}_{11}) = m_1 \Sigma_{11}$, we can easily obtain that $E(\hat{\Sigma}_{ij}^M) = \Sigma_{i1} \Sigma_{11}^{-1} \Sigma_{1j} = \Sigma_{ij}$.

In addition, similar to (2.59), we can easily see that $E(\hat{\Sigma}_{ii}^U) = \Sigma_{ii}$, $i = 2, \dots, k$ and

the part (a) is proved.

(b) By (2.56),

$$\mathbf{W}_{i1} \mid \mathbf{W}_{i11} \sim N_{p_i, p_1}(\Sigma_{i1} \Sigma_{11} \mathbf{W}_{i11}, \Sigma_{ii \cdot 1} \otimes \mathbf{W}_{i11}), \quad (2.63)$$

and $(\mathbf{W}_{i1}, \mathbf{W}_{i11})$ is independent of $\mathbf{W}_{ii \cdot 1}$. Therefore,

$$\begin{aligned} E(\mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii \cdot 1}^{-1} \mid \mathbf{W}_{i11}) &= \mathbf{W}_{i11}^{-1} E(\mathbf{W}_{1i} \mid \mathbf{W}_{i11}) E(\mathbf{W}_{ii \cdot 1}^{-1}) \\ &= \frac{1}{m_i - p_1 - p_i - 1} \mathbf{W}_{i11}^{-1} \cdot \mathbf{W}_{i11} \Sigma_{11}^{-1} \Sigma_{1i} \cdot \Sigma_{ii \cdot 1}^{-1} \\ &= \frac{1}{m_i - p_1 - p_i - 1} \Sigma_{11}^{-1} \Sigma_{1i} \Sigma_{ii \cdot 1}^{-1} \end{aligned}$$

and

$$\begin{aligned}
& E(\mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii.1}^{-1} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} \mid \mathbf{W}_{i11}) \\
&= \frac{1}{m_i - p_1 - p_i - 1} \mathbf{W}_{i11}^{-1} E(\mathbf{W}_{1i} \boldsymbol{\Sigma}_{ii.1}^{-1} \mathbf{W}_{i1} \mid \mathbf{W}_{i11}) \mathbf{W}_{i11}^{-1} \\
&= \frac{p_i}{m_i - p_1 - p_i - 1} \mathbf{W}_{i11}^{-1} + \frac{1}{m_i - p_1 - p_i - 1} \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{1i} \boldsymbol{\Sigma}_{ii.1}^{-1} \boldsymbol{\Sigma}_{i1} \boldsymbol{\Sigma}_{11}^{-1}.
\end{aligned}$$

So we get

$$\begin{aligned}
E(\hat{\boldsymbol{\Omega}}_{11}^U) &= (m_1 - p_1 - 1) \left(1 - \sum_{i=2}^k \frac{p_i}{m_i - p_1 - 1} \right) E(\mathbf{W}_{11}^{-1}) \\
&\quad + \sum_{i=2}^k \frac{p_i}{m_i - p_1 - 1} \boldsymbol{\Sigma}_{11}^{-1} + \sum_{i=2}^k \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{1i} \boldsymbol{\Sigma}_{ii.1}^{-1} \boldsymbol{\Sigma}_{i1} \boldsymbol{\Sigma}_{11}^{-1} \\
&= \boldsymbol{\Sigma}_{11}^{-1} + \sum_{i=2}^k \boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{1i} \boldsymbol{\Sigma}_{ii.1}^{-1} \boldsymbol{\Sigma}_{i1} \boldsymbol{\Sigma}_{11}^{-1} = \boldsymbol{\Omega}_{11}; \\
E(\hat{\boldsymbol{\Omega}}_{1i}^U) &= -(m_i - p_1 - p_i - 1) E(\mathbf{W}_{i11}^{-1} \mathbf{W}_{1i} \mathbf{W}_{ii.1}^{-1}) = -\boldsymbol{\Sigma}_{11}^{-1} \boldsymbol{\Sigma}_{1i} \boldsymbol{\Sigma}_{ii.1}^{-1} = \boldsymbol{\Omega}_{1i}, \\
E(\hat{\boldsymbol{\Omega}}_{ii}^U) &= (m_i - p_1 - p_i - 1) E(\mathbf{W}_{ii.1}^{-1}) = \boldsymbol{\Sigma}_{ii.1}^{-1} = \boldsymbol{\Omega}_{ii}, \quad i = 2, \dots, k.
\end{aligned}$$

The proof is completed. □

Appendix 2C Proof of Proposition 2.5

Let $\boldsymbol{\theta} = (\psi_{111}, \psi_{121}, \psi_{122}, \dots, \psi_{1p_11}, \dots, \psi_{1p_1p_1}, \phi_{211}, \dots, \phi_{21p_1}, \psi_{211}, \dots, \phi_{k11}, \dots, \phi_{kp_kp_1}, \psi_{kp_k1}, \dots, \psi_{kp_kp_k})'$ and \mathbf{I}_i be the $i \times i$ identity matrix and \mathbf{e}_i be the $i \times 1$ vector with the i th element 1 and others 0. By the likelihood function $f(\mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k \mid \boldsymbol{\Psi})$ in the proof of Proposition 2.1, the Fisher information matrix of $\boldsymbol{\theta}$ is

$$\boldsymbol{\Lambda}(\boldsymbol{\theta}) = -E \left(\frac{\partial^2 \log f}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = \text{diag}(\boldsymbol{\Lambda}_{11}, \dots, \boldsymbol{\Lambda}_{1p_1}, \boldsymbol{\Lambda}_{21}, \dots, \boldsymbol{\Lambda}_{2p_2}, \dots, \boldsymbol{\Lambda}_{k1}, \dots, \boldsymbol{\Lambda}_{kp_k}), \quad (2.64)$$

where

$$\boldsymbol{\Lambda}_{1p_1} = m_1 \left\{ (\boldsymbol{\Psi}'_{11} \boldsymbol{\Psi}_{11})^{-1} + \frac{1}{\psi_{1p_1p_1}^2} \mathbf{e}_{p_1} \mathbf{e}'_{p_1} \right\},$$

$$\begin{aligned}
\mathbf{\Lambda}_{1j} &= m_1 \left\{ (\mathbf{I}_j \mathbf{0}) (\mathbf{\Psi}'_{11} \mathbf{\Psi}_{11})^{-1} (\mathbf{I}_j \mathbf{0})' + \frac{1}{\psi_{1jj}^2} \mathbf{e}_j \mathbf{e}_j' \right\}, \quad j = 1, 2, \dots, p_1 - 1, \\
\mathbf{\Lambda}_{ip_i} &= m_i \left\{ \left(\begin{pmatrix} \mathbf{\Psi}_{11} & \mathbf{0} \\ \mathbf{\Psi}_{i1} & \mathbf{\Psi}_{ii} \end{pmatrix}' \begin{pmatrix} \mathbf{\Psi}_{11} & \mathbf{0} \\ \mathbf{\Psi}_{i1} & \mathbf{\Psi}_{ii} \end{pmatrix} \right)^{-1} + \frac{m_i}{\psi_{ip_i p_i}^2} \mathbf{e}_{p_1+p_i} \mathbf{e}'_{p_1+p_i}, \quad i = 2, \dots, k, \\
\mathbf{\Lambda}_{ij} &= m_i (\mathbf{I}_{p_1+j} \mathbf{0}) \left\{ \left(\begin{pmatrix} \mathbf{\Psi}_{11} & \mathbf{0} \\ \mathbf{\Psi}_{i1} & \mathbf{\Psi}_{ii} \end{pmatrix}' \begin{pmatrix} \mathbf{\Psi}_{11} & \mathbf{0} \\ \mathbf{\Psi}_{i1} & \mathbf{\Psi}_{ii} \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{I}_{p_1+j} \\ \mathbf{0} \end{pmatrix} + \frac{m_i}{\psi_{ijj}^2} \mathbf{e}_{p_1+j} \mathbf{e}'_{p_1+j}, \right.
\end{aligned}$$

for $j = 1, 2, \dots, p_i - 1, i = 2, \dots, k$. Thus using the fact that $|\mathbf{B} + \mathbf{a}\mathbf{a}'| = |\mathbf{B}|(1 + \mathbf{a}'\mathbf{B}^{-1}\mathbf{a})$, where \mathbf{B} is invertible and \mathbf{a} is a vector, we can easily show that

$$|\mathbf{\Lambda}_{1j}| = 2m_1^j \prod_{t=1}^j \frac{1}{\psi_{1tt}^2}, \quad 1 \leq j \leq p_1, \quad (2.65)$$

$$|\mathbf{\Lambda}_{ij}| = 2m_i^{p_1+j} \prod_{t=1}^{p_1} \frac{1}{\psi_{1tt}^2} \cdot \prod_{s=1}^{p_i} \frac{1}{\psi_{iss}^2}, \quad 1 \leq j \leq p_i, 2 \leq i \leq k. \quad (2.66)$$

Hence the Jeffreys prior of $\mathbf{\Psi}$ (or $\boldsymbol{\theta}$) is $|\mathbf{\Lambda}(\boldsymbol{\theta})|^{\frac{1}{2}}$, which is proportional to that in (2.25). From (2.64)–(2.66), the reference prior of $\mathbf{\Psi}$ for the ordered group $\{\psi_{111}, (\psi_{121}, \psi_{122}), \dots, (\psi_{1p_11}, \dots, \psi_{1p_1p_1}), (\phi_{211}, \dots, \phi_{21p_1}, \psi_{211}), \dots, (\phi_{k11}, \dots, \phi_{kp_k p_1}, \psi_{kp_k 1}, \dots, \psi_{kp_k p_k})\}$ is easily obtained as (2.26), according to the algorithm in Berger and Bernardo (1992). \square

Appendix 2D Proof of Theorem 2.2

The Bayesian estimator of $\boldsymbol{\Omega}$ under the Stein loss L_1^* will be produced by minimizing the posterior risk

$$b_1(\hat{\boldsymbol{\Omega}}) = \int \left[\text{tr} \left\{ \hat{\boldsymbol{\Omega}}^{-1} (\mathbf{\Psi}' \mathbf{\Psi}) \right\} - \log |\hat{\boldsymbol{\Omega}}^{-1} (\mathbf{\Psi}' \mathbf{\Psi})| - p \right] f(\mathbf{\Psi} \mid \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) d\mathbf{\Psi},$$

where $f(\mathbf{\Psi} \mid \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k)$ is described in Theorem 2.1. Letting $\hat{\boldsymbol{\Omega}} = \hat{\mathbf{\Psi}}' \hat{\mathbf{\Psi}}$, where $\hat{\mathbf{\Psi}} \in \mathcal{G}$ and has the similar block partition as in (2.13). Thus the question becomes

to minimize

$$g_1(\hat{\Psi}) = \int tr\{(\Psi \hat{\Psi}^{-1})(\Psi \hat{\Psi}^{-1})'\} f(\Psi | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) d\Psi - \log |(\hat{\Psi}' \hat{\Psi})^{-1}|. \quad (2.67)$$

So we need to calculate the posterior expectation of $tr\{(\Psi \hat{\Psi}^{-1})(\Psi \hat{\Psi}^{-1})'\}$. Because

$$\Psi \hat{\Psi}^{-1} = \begin{pmatrix} \Psi_{11} \hat{\Psi}_{11}^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ (\Psi_{21} - \Psi_{22} \hat{\Psi}_{22}^{-1} \hat{\Psi}_{21}) \hat{\Psi}_{11}^{-1} & \Psi_{22} \hat{\Psi}_{22}^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ (\Psi_{31} - \Psi_{33} \hat{\Psi}_{33}^{-1} \hat{\Psi}_{31}) \hat{\Psi}_{11}^{-1} & \mathbf{0} & \hat{\Psi}_{33} \Psi_{33}^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ (\Psi_{k1} - \Psi_{kk} \hat{\Psi}_{kk}^{-1} \hat{\Psi}_{k1}) \hat{\Psi}_{11}^{-1} & \mathbf{0} & \mathbf{0} & \cdots & \hat{\Psi}_{kk} \Psi_{kk}^{-1} \end{pmatrix}, \quad (2.68)$$

we have

$$\begin{aligned} tr\{(\Psi \hat{\Psi}^{-1})(\Psi \hat{\Psi}^{-1})'\} &= \sum_{i=1}^k tr\{(\hat{\Psi}'_{ii})^{-1} \Psi'_{ii} \Psi_{ii} \hat{\Psi}_{ii}^{-1}\} \\ &+ \sum_{i=2}^k tr\{(\Psi_{i1} - \Psi_{ii} \hat{\Psi}_{ii}^{-1} \hat{\Psi}_{i1})(\hat{\Psi}'_{11} \hat{\Psi}_{11})^{-1} (\Psi_{i1} - \Psi_{ii} \hat{\Psi}_{ii}^{-1} \hat{\Psi}_{i1})'\}. \end{aligned} \quad (2.69)$$

By Theorem 2.1 (d) and Lemma 2.1 (a), it follows

$$E(\Psi'_{11} \Psi_{11} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) = (\mathbf{T}'_{11})^{-1} \mathbf{D} \mathbf{T}_{11}^{-1} \quad (2.70)$$

and

$$E(\Psi'_{ii} \Psi_{ii} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) = (\mathbf{T}'_{ii})^{-1} \mathbf{B}_{1i} \mathbf{T}_{ii}^{-1}, \quad i = 2, \dots, k. \quad (2.71)$$

Moreover, for any $2 \leq i \leq k$, by Theorem 2.1(c) and applying Theorem 2.3.5 in Gupta and Nagar (2000), we have

$$\begin{aligned} &E\{(\Psi_{i1} - \Psi_{ii} \hat{\Psi}_{ii}^{-1} \hat{\Psi}_{i1})(\hat{\Psi}'_{11} \hat{\Psi}_{11})^{-1} (\Psi_{i1} - \Psi_{ii} \hat{\Psi}_{ii}^{-1} \hat{\Psi}_{i1})' | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\} \\ &= tr\{(\hat{\Psi}'_{11} \hat{\Psi}_{11})^{-1} \mathbf{W}_{i11}^{-1}\} \mathbf{I}_{p_i} \end{aligned}$$

$$\begin{aligned}
& + E\{\Psi_{ii}(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})(\hat{\Psi}'_{11}\hat{\Psi}_{11})^{-1}(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})'\Psi'_{ii} \mid \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\} \\
& = tr\{(\hat{\Psi}'_{11})^{-1}\mathbf{W}_{i11}^{-1}\hat{\Psi}_{11}^{-1}\mathbf{I}_{p_i}\} \\
& + (\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})(\hat{\Psi}'_{11}\hat{\Psi}_{11})^{-1}(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})'(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}. \quad (2.72)
\end{aligned}$$

Combining (2.69)–(2.72), it yields

$$\begin{aligned}
& E\left[tr\{(\Psi\hat{\Psi}^{-1})(\Psi\hat{\Psi}^{-1})'\} \mid \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\right] \\
& = \sum_{i=1}^k tr\{(\hat{\Psi}'_{ii})^{-1}(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}\hat{\Psi}_{ii}^{-1}\} \\
& + \sum_{i=2}^k tr\left\{(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})(\hat{\Psi}'_{11}\hat{\Psi}_{11})^{-1}(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})'(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}\right\}. \quad (2.73)
\end{aligned}$$

(2.67) is equal to

$$\begin{aligned}
g_1(\hat{\Psi}) & = \sum_{i=1}^k tr\left\{(\hat{\Psi}'_{ii})^{-1}(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}\hat{\Psi}_{ii}^{-1}\right\} - \sum_{i=1}^k \log|(\hat{\Psi}'_{ii})^{-1}\hat{\Psi}_{ii}^{-1}| \\
& + \sum_{i=2}^k tr\left\{(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})(\hat{\Psi}'_{11}\hat{\Psi}_{11})^{-1}(\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1}\hat{\Psi}_{i1})'(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}\right\} \\
& \geq \sum_{i=1}^k tr\left\{(\hat{\Psi}'_{ii})^{-1}(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}\hat{\Psi}_{ii}^{-1}\right\} - \sum_{i=1}^k \log|(\hat{\Psi}'_{ii})^{-1}\hat{\Psi}_{ii}^{-1}|
\end{aligned}$$

and the equality holds if we take $\hat{\Psi}_{i1} = -\hat{\Psi}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}$, $i = 2, 3, \dots, k$. Consequently, $g_1(\hat{\Psi})$ attaches minimum at $\hat{\Psi}_{11} = \mathbf{G}_{11}\mathbf{T}_{11}^{-1}$, $\hat{\Psi}_{ii} = \mathbf{B}_{1i}^{1/2}\mathbf{T}_{ii}^{-1}$ and $\hat{\Psi}_{i1} = -\hat{\Psi}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1}$, $i = 2, 3, \dots, k$, where \mathbf{G}_{11} is the inverse of the Cholesky decomposition of \mathbf{B}_{11}^{-1} . This completes the proof. \square

Appendix 2E Proof of Theorem 2.3

Similarly to the proof of Theorem 2.2, the Bayesian estimator of $\mathbf{\Omega}$ under the entropy loss L_2^* will be produced by minimizing the posterior risk

$$b_2(\hat{\mathbf{\Omega}}) = \int [tr\{\hat{\mathbf{\Omega}}(\mathbf{\Psi}'\mathbf{\Psi})^{-1}\} - \log |\hat{\mathbf{\Omega}}(\mathbf{\Psi}'\mathbf{\Psi})^{-1}| - p] f(\mathbf{\Psi} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) d\mathbf{\Psi},$$

which is equivalent to minimize

$$g_2(\hat{\mathbf{\Psi}}) = \int tr\{(\hat{\mathbf{\Psi}}\mathbf{\Psi}^{-1})(\hat{\mathbf{\Psi}}\mathbf{\Psi}^{-1})'\} f(\mathbf{\Psi} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) d\mathbf{\Psi} - \log |\hat{\mathbf{\Psi}}'\hat{\mathbf{\Psi}}|.$$

Similar to (2.69), we have

$$\begin{aligned} tr\{(\hat{\mathbf{\Psi}}\mathbf{\Psi}^{-1})(\hat{\mathbf{\Psi}}\mathbf{\Psi}^{-1})'\} &= \sum_{i=1}^k tr\{\hat{\mathbf{\Psi}}_{ii}(\mathbf{\Psi}'_{ii}\mathbf{\Psi}_{ii})^{-1}\hat{\mathbf{\Psi}}'_{ii}\} \\ &+ \sum_{i=2}^k tr\{(\hat{\mathbf{\Psi}}_{i1} - \hat{\mathbf{\Psi}}_{ii}\mathbf{\Psi}_{ii}^{-1}\mathbf{\Psi}_{i1})(\mathbf{\Psi}'_{11}\mathbf{\Psi}_{11})^{-1}(\hat{\mathbf{\Psi}}_{i1} - \hat{\mathbf{\Psi}}_{ii}\mathbf{\Psi}_{ii}^{-1}\mathbf{\Psi}_{i1})'\}. \end{aligned} \quad (2.74)$$

From Theorem 2.1(d) and Lemma 2.1(b),

$$E\{(\mathbf{\Psi}'_{11}\mathbf{\Psi}_{11})^{-1} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\} = \mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11},$$

$$E\{(\mathbf{\Psi}'_{ii}\mathbf{\Psi}_{ii})^{-1} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\} = \mathbf{T}_{ii}\mathbf{B}_{2i}^{-1}\mathbf{T}'_{ii}/\{1 + tr(\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1})\}, i = 2, \dots, k$$

because of $m_i + a_{ij} - 1 > 0$, $j = 1, \dots, p_i$, $i = 1, \dots, k$. In addition, similar to (2.72),

for $i = 2, \dots, k$,

$$\begin{aligned} &E \left[tr\{(\hat{\mathbf{\Psi}}_{i1} - \hat{\mathbf{\Psi}}_{ii}\mathbf{\Psi}_{ii}^{-1}\mathbf{\Psi}_{i1})(\mathbf{\Psi}'_{11}\mathbf{\Psi}_{11})^{-1}(\hat{\mathbf{\Psi}}_{i1} - \hat{\mathbf{\Psi}}_{ii}\mathbf{\Psi}_{ii}^{-1}\mathbf{\Psi}_{i1})' | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k \right] \\ &= tr \left[E\{(\hat{\mathbf{\Psi}}_{i1} - \hat{\mathbf{\Psi}}_{ii}\mathbf{\Psi}_{ii}^{-1}\mathbf{\Psi}_{i1})'(\hat{\mathbf{\Psi}}_{i1} - \hat{\mathbf{\Psi}}_{ii}\mathbf{\Psi}_{ii}^{-1}\mathbf{\Psi}_{i1}) | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\} \right. \\ &\quad \left. E\{(\mathbf{\Psi}'_{11}\mathbf{\Psi}_{11})^{-1} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k\} \right] \\ &= tr \left\{ (\hat{\mathbf{\Psi}}_{i1} + \hat{\mathbf{\Psi}}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1})'(\hat{\mathbf{\Psi}}_{i1} + \hat{\mathbf{\Psi}}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1})\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11} \right\} \\ &\quad + tr(\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1})tr(\hat{\mathbf{\Psi}}_{ii}\mathbf{T}_{ii}\mathbf{B}_{2i}^{-1}\mathbf{T}'_{ii}\hat{\mathbf{\Psi}}'_{ii})/\{1 + tr(\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\mathbf{W}_{i11}^{-1})\}. \end{aligned}$$

So we have

$$\begin{aligned}
E[tr\{(\hat{\Psi}\Psi^{-1})(\hat{\Psi}\Psi^{-1})'\} | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k] &= \sum_{i=1}^k tr(\hat{\Psi}_{ii}\mathbf{T}_{ii}\mathbf{B}_{2i}^{-1}\mathbf{T}'_{ii}\hat{\Psi}'_{ii}) \\
&+ \sum_{i=2}^k tr\left\{(\hat{\Psi}_{i1} + \hat{\Psi}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1})'(\hat{\Psi}_{i1} + \hat{\Psi}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1})\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\right\} \quad (2.75)
\end{aligned}$$

and thus

$$\begin{aligned}
g_2(\hat{\Psi}) &= \sum_{i=1}^k \left\{ tr(\hat{\Psi}_{ii}\mathbf{T}_{ii}\mathbf{B}_{2i}^{-1}\mathbf{T}'_{ii}\hat{\Psi}'_{ii}) - \log |\hat{\Psi}'_{ii}\hat{\Psi}_{ii}| \right\} \\
&+ \sum_{i=2}^k tr\left\{(\hat{\Psi}_{i1} + \hat{\Psi}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1})'(\hat{\Psi}_{i1} + \hat{\Psi}_{ii}\mathbf{W}_{i1}\mathbf{W}_{i11}^{-1})\mathbf{T}_{11}\mathbf{B}_{21}^{-1}\mathbf{T}'_{11}\right\}.
\end{aligned}$$

Hence, we can readily see that $g_2(\hat{\Psi})$ is minimized at $\hat{\Psi}_{ii} = \mathbf{B}_{2i}^{1/2}\mathbf{T}_{ii}^{-1}$ for $i = 1, \dots, k$ and $\hat{\Psi}_{j1} = -\hat{\Psi}_{jj}\mathbf{W}_{j1}\mathbf{W}_{j11}^{-1}$ for $j = 2, \dots, k$. Thus the proof is completed. \square

Appendix 2F Proof of Theorem 2.4

The Bayesian estimator of $\mathbf{\Omega}$ under the symmetric loss L_3^* will be produced by minimizing the posterior risk

$$b_3(\hat{\mathbf{\Omega}}) = \int \left[tr\left\{\hat{\mathbf{\Omega}}(\Psi'\Psi)^{-1}\right\} + tr\left\{\hat{\mathbf{\Omega}}^{-1}(\Psi'\Psi)\right\} - 2p \right] f(\Psi | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) d\Psi,$$

which is equivalent to minimize

$$\begin{aligned}
g_3(\hat{\Psi}) &= \int \left[tr\left\{(\hat{\Psi}\Psi^{-1})(\hat{\Psi}\Psi^{-1})'\right\} + tr\left\{(\Psi\hat{\Psi}^{-1})(\Psi\hat{\Psi}^{-1})'\right\} \right] f(\Psi | \mathbf{S}, \mathbf{V}_1, \dots, \mathbf{V}_k) d\Psi \\
&- \log |\hat{\Psi}'\hat{\Psi}|.
\end{aligned}$$

Combining (2.75) with (2.73), it yields

$$g_3(\hat{\Psi}) = \sum_{i=1}^k \left[tr(\hat{\Psi}_{ii}\mathbf{T}_{ii}\mathbf{B}_{2i}^{-1}\mathbf{T}'_{ii}\hat{\Psi}'_{ii}) + tr\left\{(\hat{\Psi}'_{ii})^{-1}(\mathbf{T}'_{ii})^{-1}\mathbf{B}_{1i}\mathbf{T}_{ii}^{-1}\hat{\Psi}_{ii}^{-1}\right\} \right]$$

$$\begin{aligned}
& + \sum_{i=2}^k tr\{(\hat{\Psi}_{i1} + \hat{\Psi}_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1})'(\hat{\Psi}_{i1} + \hat{\Psi}_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}) \mathbf{T}_{11} \mathbf{B}_{21}^{-1} \mathbf{T}'_{11}\} \\
& + \sum_{i=2}^k tr\{(\mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1} \hat{\Psi}_{i1})(\hat{\Psi}'_{11} \hat{\Psi}_{11})^{-1}(\mathbf{W}_{i1} \mathbf{W}_{i11}^{-1} + \hat{\Psi}_{ii}^{-1} \hat{\Psi}_{i1})'(\mathbf{T}'_{ii})^{-1} \mathbf{B}_{1i} \mathbf{T}_{ii}^{-1}\} \\
& \geq \sum_{i=1}^k \left[tr(\hat{\Psi}_{ii} \mathbf{T}_{ii} \mathbf{B}_{2i}^{-1} \mathbf{T}'_{ii} \hat{\Psi}'_{ii}) + tr\{(\hat{\Psi}'_{ii})^{-1}(\mathbf{T}_{ii})^{-1} \mathbf{B}_{1i} \mathbf{T}_{ii}^{-1} \hat{\Psi}_{ii}^{-1}\} \right],
\end{aligned}$$

and the equality holds if we take $\hat{\Psi}_{i1} = -\hat{\Psi}_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}$, $i = 2, 3, \dots, k$. Thus, by Lemma 2.2, we can easily see that $g_3(\hat{\Psi})$ attaches minimum at $\hat{\Psi}_{11} = \mathbf{Q}_{31} \mathbf{T}_{11}^{-1}$ with \mathbf{Q}_{31} being the inverse of Cholesky decomposition of $\mathbf{B}_{11}^{-1/2}(\mathbf{B}_{11}^{1/2} \mathbf{B}_{21}^{-1} \mathbf{B}_{11}^{1/2})^{1/2} \mathbf{B}_{11}^{-1/2}$, $\hat{\Psi}_{ii} = \mathbf{B}_{3i}^{1/2} \mathbf{T}_{ii}^{-1}$ and $\hat{\Psi}_{i1} = -\hat{\Psi}_{ii} \mathbf{W}_{i1} \mathbf{W}_{i11}^{-1}$, $i = 2, 3, \dots, k$. Thus the proof is completed.

□

Chapter 3

An Efficient Simulation Algorithm for Bayesian Spatial Analysis

3.1 Introduction

Spatial data arises in many fields of application including ecological, environmental, and epidemiological settings. When formal inference is sought, point-referenced spatial data are often modelled using Gaussian random fields, specified by their mean function and covariance function. Inference based on maximum likelihood estimation has been implemented by Mardia and Marshall (1984). Customary likelihood asymptotic theories are not usually applicable (as discussed in Stein (1999)) and hence Bayesian modelling, which provides exact inference, becomes attractive. The main advantage of the Bayesian approach is that parameter uncertainty is fully accounted for when performing prediction and inference, even in small samples.

Simulation methods play an important role in full Bayesian analysis and MCMC algorithms are now widely used in all areas of statistics. MCMC methods are roughly as follows. Consider a desired probability distribution π . Assume for the moment that we can construct a Markov chain, such that its stationary probability distribution is π . We run a Markov chain forward till such time as its distribution is close to the stationary distribution, and then stop based on some criteria determined by diagnostics or other means. The best known MCMC methods are the Metropolis-Hastings algorithm and the Gibbs sampler. The Gibbs sampler is popularly applied in spatial area since the number of interest parameters is large. In addition, Metropolis-Hastings algorithm or slice sampler within the Gibbs sampler are often used because the full conditional distribution of the parameters θ controlling the correlation function is not of standard form, see Besag and Green (1993), Cowles (2003), Banerjee et al. (2004), Agarwal and Gelfand (2005) etc. For a review of MCMC methods such as Gibbs sampling, the Metropolis-Hastings algorithm, slice sampler etc, and their applications in spatial statistics, see Dellaportas and Roberts (2003).

However, an obvious problem with the MCMC methods is that it is not clear when one should stop running the chain and accept the final state as a realization of the probability distribution π . MCMC algorithms also usually suffer from slow convergence and highly serially correlated outputs, especially in the field of spatial statistics, (see Dellaportas and Roberts (2003)). Consequently, MCMC methods introduce bias into the sample. The simulation inference from correlated samples is generally less precise than from the same number of independent samples. In this chapter, we pro-

pose a new simulation algorithm based on the generalized Ratio-of-Uniforms method of Wakefield et al. (1991). Our new algorithm will create independent samples, rather than correlated samples.

This chapter is organized as follows. § 3.2 gives a brief review of Gaussian process models and a class of priors that we will use for Bayesian analysis. A new efficient algorithm for the posterior simulation is proposed based on the generalized Ratio-of-Uniforms method in § 3.3. As far as we know, this method has not been drawn much attention in the Bayesian spatial area, although it has been proposed by Wakefield et al. (1991) for about fifteen years. An illustrative simulated example will be provided in § 3.4. Some concluding remarks and discussions are given in § 3.5.

3.2 The Model

3.2.1 The likelihood

Let $\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{D}\}$, $\mathcal{D} \subseteq \mathbb{R}^2$, be the random field of interest. The data consist of n observations $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))'$ from a single realization of this random field, where $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ are known distinct sampling locations in \mathcal{D} . Based on \mathbf{Z} , the usual tasks of interest are estimation of the mean and covariance functions of the random field, and prediction of an unobserved random vector $\mathbf{Z}_0 = (Z(\mathbf{s}_{01}), Z(\mathbf{s}_{02}), \dots, Z(\mathbf{s}_{0k}))$, where $\mathbf{s}_{01}, \mathbf{s}_{02}, \dots, \mathbf{s}_{0k}$ are unsampled locations in the subregion \mathcal{D} .

We assume that $\mathbf{Z}(\cdot)$ is a Gaussian random field with $\mathbb{E}\{Z(\mathbf{s})\} = \sum_{j=0}^{p-1} \beta_j f_j(\mathbf{s}) =$

$\boldsymbol{\beta}'\mathbf{f}(\mathbf{s})$; $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})' \in \mathbb{R}^p$ are unknown regression parameters, $\mathbf{f}(\mathbf{s}) = (f_1(\mathbf{s}), \dots, f_p(\mathbf{s}))'$ are known location-dependent covariates, and $\text{cov}\{Z(\mathbf{s}), Z(\mathbf{u})\} = \sigma^2 K_{\boldsymbol{\theta}}(\|\mathbf{s} - \mathbf{u}\|)$ where $\|\cdot\|$ denotes Euclidean distance, $\sigma^2 = \text{var}\{Z(\mathbf{s})\}$, $K_{\boldsymbol{\theta}}(\|\mathbf{s} - \mathbf{u}\|) = \text{corr}\{Z(\mathbf{s}), Z(\mathbf{u})\}$, an isotropic correlation function, and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_c)' \in \Theta \subset \mathbb{R}^c$ are parameters controlling the range of correlation and the smoothness of the random field.

There are many correlation functions that are popularly used in spatial statistics, see Cressie (1993) or Banerjee et al. (2004) etc.

The likelihood of the model parameters $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$, based on the observed data $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))'$, is given by

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}; \mathbf{z}) = (2\pi\sigma^2)^{-n/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\right\} \quad (3.1)$$

where $\mathbf{X} = (x_{ij})$ is the known $n \times p$ matrix defined by $x_{ij} = f_j(\mathbf{s}_i)$, assumed to have full rank, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (K_{\boldsymbol{\theta}}(\|\mathbf{s}_i - \mathbf{s}_j\|))$, assumed to be positive definite for any $\boldsymbol{\theta} \in \Theta$.

3.2.2 The Priors

We consider a class of priors,

$$\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) \propto \frac{\pi(\boldsymbol{\theta})}{(\sigma^2)^a}, \quad a \in \mathbb{R}. \quad (3.2)$$

Slightly abusing terminology, we call $\pi(\boldsymbol{\theta})$ the marginal prior density of $\boldsymbol{\theta}$. This form of prior distribution, with $a = 1$, was proposed by Kitanidis (1986) and Handcock and Stein (1993). Berger et al. (2001) showed that the Jeffreys prior, the independent Jeffreys prior and the reference prior all are of this form.

Combining the likelihood (3.1) with the prior (3.2), we have the joint posterior distribution

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} \mid \mathbf{z}) \propto \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}; \mathbf{z}). \quad (3.3)$$

Obviously, it is very difficult to do analytical posterior analysis for (3.3). Simulation is therefore needed to estimate the parameters. The following integrated likelihood plays an important role in simulation study of the posterior distribution.

3.2.3 Integrated likelihood

As noted by Berger et al. (2001), it is possible to integrate the product of the likelihood and the prior over $(\boldsymbol{\beta}, \sigma^2)$ in a closed form for the prior distributions of the form (3.2). Indeed, a standard calculation yields

$$\int_{\mathbb{R}^p \times (0, \infty)} L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}; \mathbf{z}) \pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) d\boldsymbol{\beta} d\sigma^2 = L^I(\boldsymbol{\theta}; \mathbf{z}) \pi(\boldsymbol{\theta}), \quad (3.4)$$

where

$$L^I(\boldsymbol{\theta}; \mathbf{z}) \propto |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2} |\mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X}|^{-1/2} (S_{\boldsymbol{\theta}}^2)^{-((n-p)/2+a-1)} \quad (3.5)$$

and

$$S_{\boldsymbol{\theta}}^2 = (\mathbf{z} - \mathbf{X} \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X} \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}). \quad (3.6)$$

Note that $S_{\boldsymbol{\theta}}^2$ is the generalized residual sum of squares, and

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}} = (\mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} \mathbf{z} \quad (3.7)$$

is the generalized least squares estimator of $\boldsymbol{\beta}$, given $\boldsymbol{\theta}$. In fact, the right hand of (3.4) is proportional to the marginal posterior density $p(\boldsymbol{\theta} \mid \mathbf{z})$ of $\boldsymbol{\theta}$, which is crucial in

our new simulation scheme. This integrated likelihood is also of considerable interest to non-Bayesians, as discussed, for instance, in Berger et al. (1999).

Based on (3.4), the joint posterior distribution $p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} \mid \mathbf{z})$ is proper if and only if the marginal posterior density $p(\boldsymbol{\theta} \mid \mathbf{z})$ is proper, that is, $\int L^I(\boldsymbol{\theta}; \mathbf{z})\pi(\boldsymbol{\theta}) d\boldsymbol{\theta} < \infty$.

3.3 Posterior Simulation

3.3.1 Gibbs sampler

The Gibbs sampler is commonly used in spatial statistics. To implement the Gibbs sampler, we must assume that samples can be generated from each of the full conditional distributions. For our model, the full conditional distributions are:

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\theta}; \mathbf{z} \sim N_p(\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}), \quad (3.8)$$

$$\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{\theta}; \mathbf{z} \sim IG(n/2 + a - 1, (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})/2), \quad (3.9)$$

$$[\boldsymbol{\theta} \mid \boldsymbol{\beta}, \sigma^2; \mathbf{z}] \propto \exp\{-(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})/(2\sigma^2)\}\pi(\boldsymbol{\theta}). \quad (3.10)$$

where $IG(\alpha_1, \alpha_2)$ represents the inverse gamma distribution with shape parameter α_1 and scale parameter α_2 . For notational simplicity, we sometimes use the notation $[\boldsymbol{\alpha}_1 \mid \boldsymbol{\alpha}_2]$ to represent the conditional probability density of $\boldsymbol{\alpha}_1$ given $\boldsymbol{\alpha}_2$. Sampling from (3.8) and (3.9) is straightforward. However, sampling from (3.10) could be problematic because it is not a standard form.

Two methods may be considered to sample from (3.10). One is the Metropolis-Hastings algorithm. But there is no routine to choose an appropriate candidate

distribution for this algorithm. The other is the slice sampler, making use of suitable factorizations of the density to introduce auxiliary variables to simplify the sampling problem (Banerjee et al. (2004)). However, this method may not work well when the marginal prior $\pi(\theta)$ is not a standard distribution, for example, the reference prior obtained by Berger et al. (2001). Furthermore, both methods suffer from highly serially correlated draws and slow convergence in MCMC simulation. In §3.3.3, we propose a new simulation algorithm based on the generalized ratio-of-uniforms method.

3.3.2 The generalized Ratio-of-Uniforms method

The Ratio-of-Uniforms method introduced by Kinderman and Monahan (1977) is a flexible method that can be adjusted to a large variety of distributions. It has become a popular transformation method to generate nonuniform random variates, since it results in exact, efficient, fast, and easy-to-implement algorithms. The method is based on the following lemma.

Lemma 3.1 *Let h be a positive integrable function over \mathcal{X} , a subset of \mathbb{R} . Suppose that the variables (u, v) are uniformly distributed over*

$$C_h = \left\{ (u, v) : 0 < u \leq \left[h\left(\frac{v}{u}\right) \right]^{1/2} \right\}. \quad (3.11)$$

Then $x = v/u$ had density $h(x) / \int h(x)$.

This method is only suited for one-dimensional bell-shaped densities with tails that decrease at least as fast as x^{-2} . Wakefield et al. (1991) proposed a generalization of the Ratio-of-Uniforms method.

Lemma 3.2 *Let h be a positive integrable function over \mathcal{X} , a subset of \mathbb{R}^k . Suppose that the variables (u, v_1, \dots, v_k) are uniformly distributed over*

$$C_h(r) = \left\{ (u, v_1, \dots, v_k) : 0 < u \leq \left[h\left(\frac{v_1}{u^r}, \dots, \frac{v_k}{u^r}\right) \right]^{1/(rk+1)} \right\} \quad (3.12)$$

where $r \geq 0$ is a constant. Then $\mathbf{x} = (v_1/u^r, \dots, v_k/u^r)$ had density $h(\mathbf{x}) / \int h(\mathbf{x}) dx$.

For sampling random points uniformly distributed in $C_h(r)$, the Accept-Reject algorithm from a convenient k -dimensional enveloping rectangle $\mathcal{T}(r) = [0, a(r)] \times [b_1^-(r), b_1^+(r)] \times \dots \times [b_k^-(r), b_k^+(r)]$ is often used, where

$$\begin{aligned} a(r) &= \sup_{\mathbf{x}} \{h(\mathbf{x})\}^{1/(rk+1)}, \\ b_i^-(r) &= \inf_{\mathbf{x}} x_i \{h(\mathbf{x})\}^{r/(rk+1)}, \quad i = 1, \dots, k, \\ b_i^+(r) &= \sup_{\mathbf{x}} x_i \{h(\mathbf{x})\}^{r/(rk+1)}, \quad i = 1, \dots, k \end{aligned} \quad (3.13)$$

provided that all of them are finite.

When $k = 1$ and $h(x) = O(x^{-1-\varepsilon})$ as $x \rightarrow \infty$ for some $\varepsilon > 0$, it is easy to see that, by choosing r such that $r\varepsilon > 1$, $a(r)$, $b_1^-(r)$ and $b_1^+(r)$ defined by (3.13) are finite. Therefore, the above generalized ratio-of-uniforms method works for any bell-shaped density with tails that decrease at least as fast as $x^{-1-\varepsilon}$, $\varepsilon > 0$.

3.3.3 A new simulation algorithm

Our new simulation scheme is based on the following result.

Theorem 3.1 *The joint posterior distribution of $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$ has the following decomposition*

$$p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} \mid \mathbf{z}) = p(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{\theta}; \mathbf{z}) p(\sigma^2 \mid \boldsymbol{\theta}; \mathbf{z}) p(\boldsymbol{\theta} \mid \mathbf{z}), \quad (3.14)$$

where the conditional distribution of $\boldsymbol{\beta}$ given $\sigma^2, \boldsymbol{\theta}; \mathbf{z}$ is given by (3.8) and

$$\sigma^2 \mid \boldsymbol{\theta}; \mathbf{z} \sim IG((n-p)/2 + a, \mathbf{z}'\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}\mathbf{z}/2), \quad (3.15)$$

$$[\boldsymbol{\theta} \mid \mathbf{z}] \propto L^I(\boldsymbol{\theta}; \mathbf{z})\pi(\boldsymbol{\theta}), \quad (3.16)$$

with $L^I(\boldsymbol{\theta}; \mathbf{z})$ given by (3.5).

Proof. We just need to prove (3.15) because of (3.8) and (3.4). Since

$$\begin{aligned} & (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{z}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{z} + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}})'\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}) - \hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}'\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}, \end{aligned}$$

it follows

$$\begin{aligned} p(\sigma^2 \mid \boldsymbol{\theta}; \mathbf{z}) &= \int L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}; \mathbf{z})\pi(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}) d\boldsymbol{\beta} \\ &\propto (\sigma^2)^{-n/2-a} \int \exp\{-(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})/(2\sigma^2)\} d\boldsymbol{\beta} \\ &\propto (\sigma^2)^{-(n-p)/2-a} \exp[-\mathbf{z}'\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}\mathbf{z}/2]/(2\sigma^2) \end{aligned}$$

and hence (3.15). □

Let $h(\boldsymbol{\theta}) = L^I(\boldsymbol{\theta}; \mathbf{z})\pi(\boldsymbol{\theta})$. If there exists $r > 0$ such that all of $a(r), b_i^-(r), b_i^+(r), i = 1, \dots, k$ defined by (3.13) are finite, then we have the following algorithm for the posterior simulation.

Algorithm:

Step 1: Simulate

$$U \sim \text{Uniform } [0, a(r)],$$

$$V_i \sim \text{Uniform } [b_i^-(r), b_i^+(r)], \quad i = 1, \dots, k$$

and compute $\boldsymbol{\rho} = \mathbf{V}/U^r$ with $\mathbf{V} = (V_1, \dots, V_k)$;

Step 2: If $U \leq [L^I(\boldsymbol{\rho}; \mathbf{z})\pi(\boldsymbol{\rho})]^{1/(rk+1)}$, we accept $\boldsymbol{\rho}$ as a sample $\boldsymbol{\theta}$ from $L^I(\boldsymbol{\theta}; \mathbf{z})\pi(\boldsymbol{\theta})$;

otherwise go back to Step 1;

Step 3: For given $\boldsymbol{\theta}$, simulate σ^2 from

$$IG((n-p)/2 + a, \mathbf{z}'\{\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} - \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\}\mathbf{z}/2);$$

Step 4: For given $\boldsymbol{\theta}, \sigma^2$, simulate $\boldsymbol{\beta}$ from $N_p(\hat{\boldsymbol{\beta}}_{\boldsymbol{\theta}}, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\mathbf{X})^{-1})$;

Step 5: Go back to Step 1.

Note that the above algorithm will create *independent samples* from the joint posterior distribution of $\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}$, which is a key difference from other MCMC sampling methods such as Gibbs sampler. We found that the above algorithm works for a lot of choices of a and $\pi(\boldsymbol{\theta})$ in (3.2) in spatial data analysis. In the next section, we will provide a detailed example for applying the above algorithm.

Remark 3.1 For prediction of $Z(\mathbf{s}_0)$ at an unsampled site \mathbf{s}_0 , we just need to sample from $N(\mathbf{x}'_0\boldsymbol{\beta}^{(i)} + \gamma^{(i)'}\boldsymbol{\Sigma}_{\boldsymbol{\theta}^{(i)}}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}^{(i)}), \sigma^{2(i)}(1 - \gamma^{(i)'}\boldsymbol{\Sigma}_{\boldsymbol{\theta}^{(i)}}^{-1}\gamma^{(i)}))$ after we obtain the samples $(\boldsymbol{\beta}^{(i)}, \sigma^{2(i)}, \boldsymbol{\theta}^{(i)})$ from the joint posterior distribution $p(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta} \mid \mathbf{z})$, where $\gamma^{(i)} = (K_{\boldsymbol{\theta}^{(i)}}(\|\mathbf{s}_0 - \mathbf{s}_1\|), \dots, K_{\boldsymbol{\theta}^{(i)}}(\|\mathbf{s}_0 - \mathbf{s}_n\|))'$.

3.4 An Illustrative Example

In this section, we analyze a simple spatial problem by applying our simulation algorithm described in § 3.3.

Suppose that we choose the correlation function to be exponential. Then

$$K_\theta(d) = \exp(-d/\theta), \quad \theta > 0. \quad (3.17)$$

Berger et al. (2001) obtained the reference prior $\pi^R(\boldsymbol{\beta}, \sigma^2, \theta)$, which is of the form (3.2) with

$$a = 1 \quad \text{and} \quad \pi^R(\theta) \propto \left\{ \text{tr}[\mathbf{W}_\theta^2] - \frac{1}{n-p} (\text{tr}[\mathbf{W}_\theta])^2 \right\}^{1/2}, \quad (3.18)$$

where

$$\mathbf{W}_\theta = \frac{\partial \boldsymbol{\Sigma}_\theta}{\partial \theta} \boldsymbol{\Sigma}_\theta^{-1} \mathbf{P}_\theta^\boldsymbol{\Sigma} \quad \text{and} \quad \mathbf{P}_\theta^\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}; \quad (3.19)$$

$(\partial/\partial\theta)\boldsymbol{\Sigma}_\theta$ denotes the matrix obtained by differentiating $\boldsymbol{\Sigma}_\theta$ element by element.

Berger et al. (2001) also proved that the posterior distribution $p(\boldsymbol{\beta}, \sigma^2, \theta \mid \mathbf{z})$ under the reference prior is proper.

Consider a 10×10 grid in $\mathcal{D} = [0, 1] \times [0, 1]$ with data points $\mathbf{z} = (z_{11}, z_{12}, \dots, z_{10,10})'$ and assume that $\mathbf{Z} \sim N_{100}(\boldsymbol{\mu}, \sigma^2 \boldsymbol{\Sigma}(\theta))$. The data \mathbf{Z} that fill the grid were simulated by taking $\boldsymbol{\mu} = \mathbb{E}[Z(\mathbf{s})] = 5.5 - 4.6x - 3.1y + 5.9x^2 - 6.5xy + 4.2y^2$, $\theta = 0.2$ and $\sigma^2 = 0.2$. So the real parameter $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_{p-1})' = (5.5, -4.6, -3.1, 5.9, -6.5, 4.2)'$. Our goal here is to estimate the parameters $(\boldsymbol{\beta}, \sigma^2, \theta)$ based on the simulated data \mathbf{z} .

To implement the first step of the algorithm, we need to study the analytical properties of the product of the integrated function $L^I(\theta; z)$ and the marginal prior density $\pi^R(\theta)$. Figure 3.1 and Figure 3.2 display the marginal prior density $\pi^R(\theta)$ and the marginal posterior density $\pi^R(\theta \mid \mathbf{z})$ up to a normalizing constant, respectively. By Lemma 1 of Berger et al. (2001), we have

$$\lim_{\theta \rightarrow 0^+} L^I(\theta; z) = c_0 > 0 \quad \text{and} \quad L^I(\theta; z) = O(1) \quad \text{as} \quad \theta \rightarrow \infty. \quad (3.20)$$

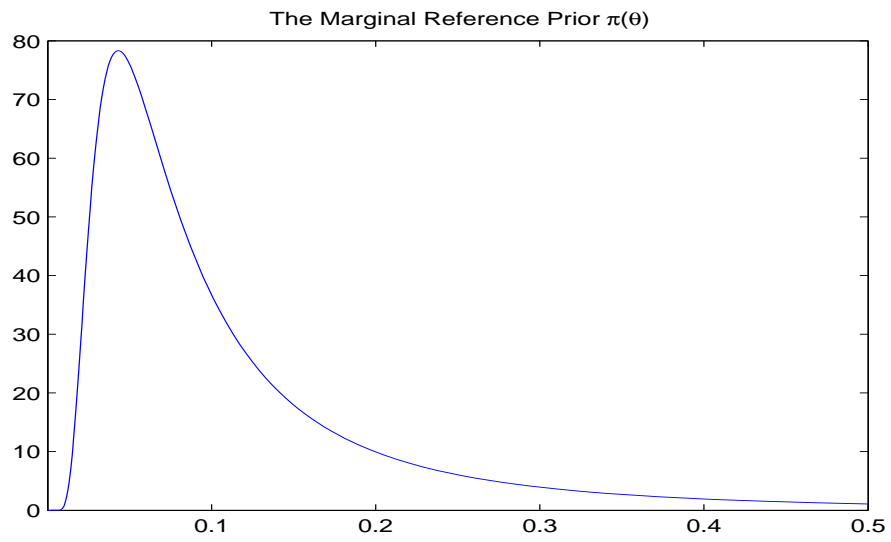


Figure 3.1: The Reference Prior $\pi^R(\theta)$.

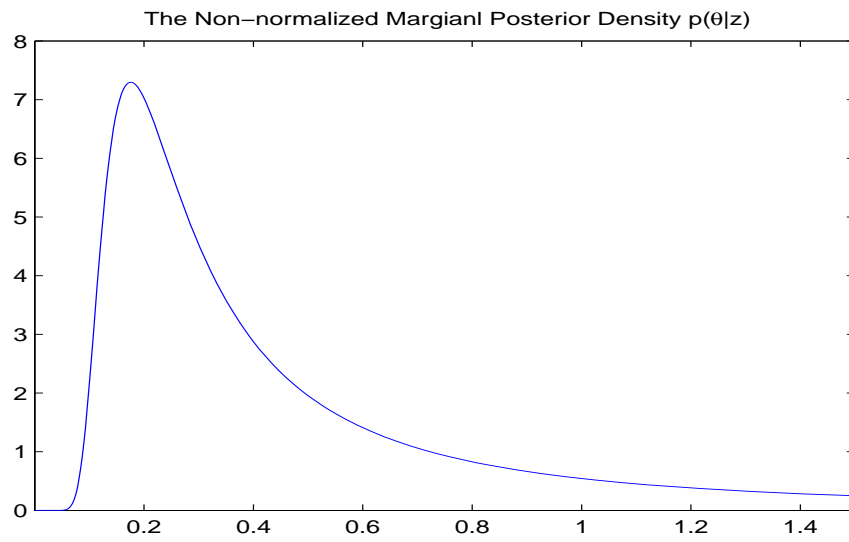


Figure 3.2: The Non-normalized Marginal Posterior Density of θ .

In addition, the proof of Corollary 1 of Berger et al. (2001) shows that

$$\pi^R(\theta) \leq \frac{C}{\theta^2} \quad (3.21)$$

if θ is large enough, where C is a constant. Therefore, by taking $r = 5$, we can easily show that $a(r), b_1^-(r), b_1^+(r)$ defined by (3.13) are finite. Obviously, $b_1^-(r) = 0$, and numerical calculation yields

$$a(r) = 1.3941, \quad b_1^+(r) = 1.0774. \quad (3.22)$$

The shaded area in Figure 3.3 shows the acceptance area when implementing the generalized Ratio-of-Uniforms method. Theoretically, the acceptance rate is the ratio of the shaded area to the area of the rectangle $[0, a(r)] \times [0, b^+(r)]$, which is about 0.38. In fact, we obtained a sample of size 10,000 with this algorithm through generating 26526 pairs of (u, v) , which means that the actual acceptance rate is 37.74 percent.

Remark 3.2 *If (u, v) are uniformly distributed over $C_h(\mu) = \{(u, v) : 0 < u \leq [h(\frac{v}{u} + \mu)]^{1/2}\}$, Then $x = v/u + \mu$ has density $h(x) / \int h(x) dx$, where μ is a constant. One may choose an appropriate μ to get higher acceptance rate for the above example.*

Figures 3.4 and 3.5 show histograms and density estimates for the scale parameter θ and the variance σ^2 , respectively.

Note that for any symmetric and unimodal distribution, its mean, median and mode will be equal if the mean exists. However, heavy-tailed distributions are popular in Bayesian spatial models. In this case, the posterior mean will sometimes be overly influenced by heavy tails of the posterior distribution. Banerjee et al. (2004)

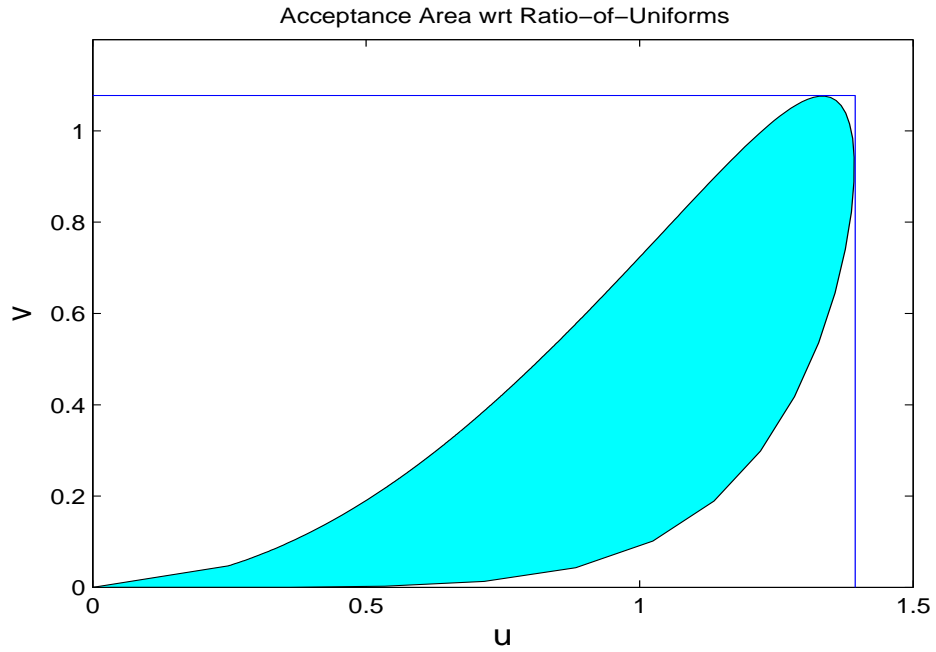


Figure 3.3: Acceptance Area based on the Generalized Ratio-of-Uniforms Method.

suggested that the posterior median would often be the best and safest point estimate for the corresponding parameter. We found that the posterior mode is the best for estimating the corresponding parameter in our setting. For estimating θ , the posterior mode, median and mean are 0.1980, 0.3663, 1.7217, respectively, while the real value of θ is 0.2. For estimating σ^2 , the posterior mode, median and mean are 0.1770, 0.3034, 1.4143, respectively, while the real value of σ^2 is 0.2.

Remark 3.3 *Loss functions are often used in estimation of parameters with both frequentist and Bayesian methods. People are reluctant to use loss functions, such as squared loss or entropy loss, for estimation of parameters in spatial statistics. The main reason is that it is very difficult to show the existence of high moments because the structure of a covariance matrix is complicated.*

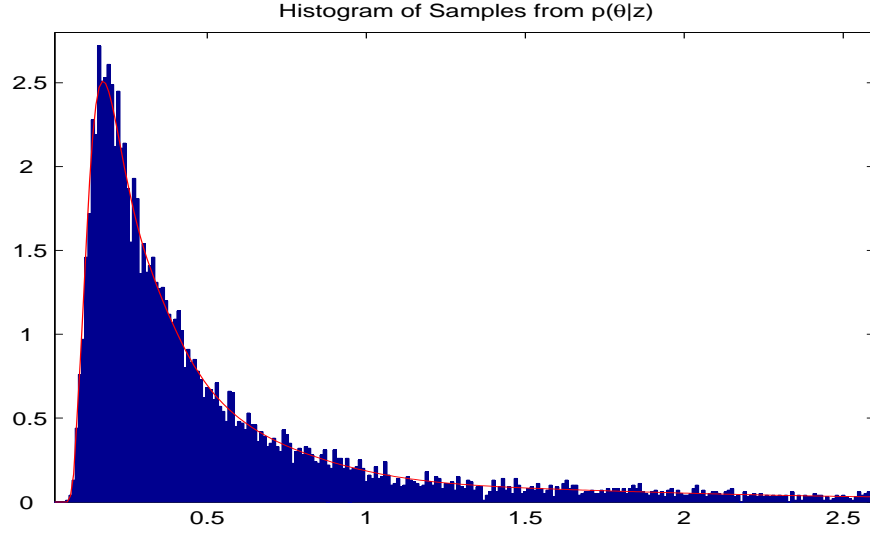


Figure 3.4: Histogram, Density Estimate for the Posterior of θ based on Simulation.

For estimating the regression coefficient $\boldsymbol{\beta}$, we found that the posterior distribution of β_i often widely spreads and thus it is difficult to estimate each of them accurately. For example, the posterior mean, median and mode of β_1 are -7.2179 , -7.2424 , -7.325 , respectively, while the real value of β_1 is -4.60 . The reason for this is that the structures of the design matrix \mathbf{X} and the correlation matrix $\boldsymbol{\Sigma}(\theta)$ strongly influences the shapes of the posterior distributions of the regression coefficients although there are 100 observations. In fact, it is well known that the generalized least square estimate of $\boldsymbol{\beta}$, which is the best linear unbiased estimate, follows the multivariate normal distribution with the mean $\boldsymbol{\beta}$ and the covariance matrix $\sigma^2[\mathbf{X}'\boldsymbol{\Sigma}^{-1}(\theta)\mathbf{X}]^{-1}$ when both σ^2 and θ are known. Thus for estimating β_1 , the standard deviation of the best linear unbiased estimate will be 1.4012 even though it is assumed that both σ^2 and θ are known.

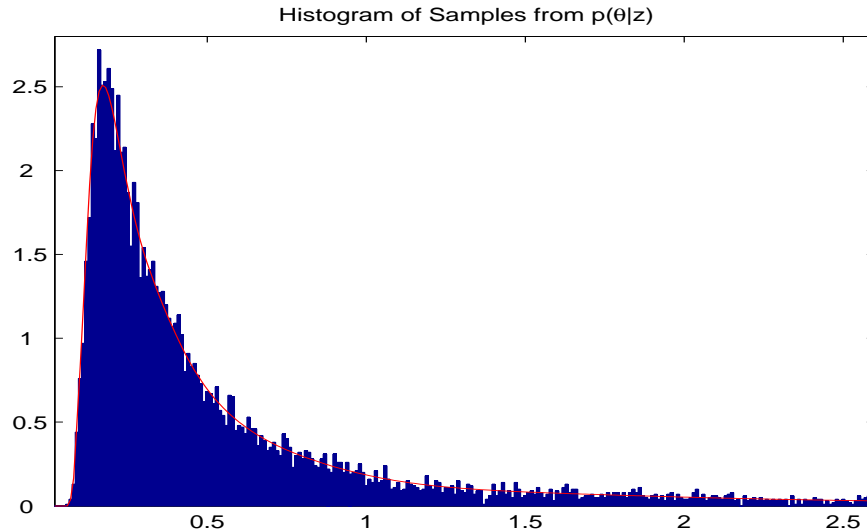


Figure 3.5: Histogram, Density Estimate for the Posterior of σ^2 based on Simulation.

Remark 3.4 *There are two different asymptotic frameworks in spatial models: increasing domain asymptotics, in which the minimum distance between sampling points is bounded away from zero and thus the spatial domain of observation is unbounded; and infill asymptotics, in which observations are taken ever more densely in a fixed and bounded domain. Asymptotic behaviors of spatial covariance parameters estimators can be quite different under the two frameworks. See Zhang (2004) and Zhang and Zimmerman (2005) etc.*

We also did a small comparison study for sampling from the posterior distribution $p(\theta \mid \mathbf{z})$ by using the Metropolis-Hastings algorithm and the Ratio-of-Uniforms algorithm. Note that the output sampling from the Metropolis-Hastings algorithm are correlated. But we can get approximately independent samples by thinning the output, that is, keeping every k th simulation draw and discarding the rest. For the above $p(\theta \mid \mathbf{z})$, we found that k should be at least 100 when $Gamma(1.4, 2)$

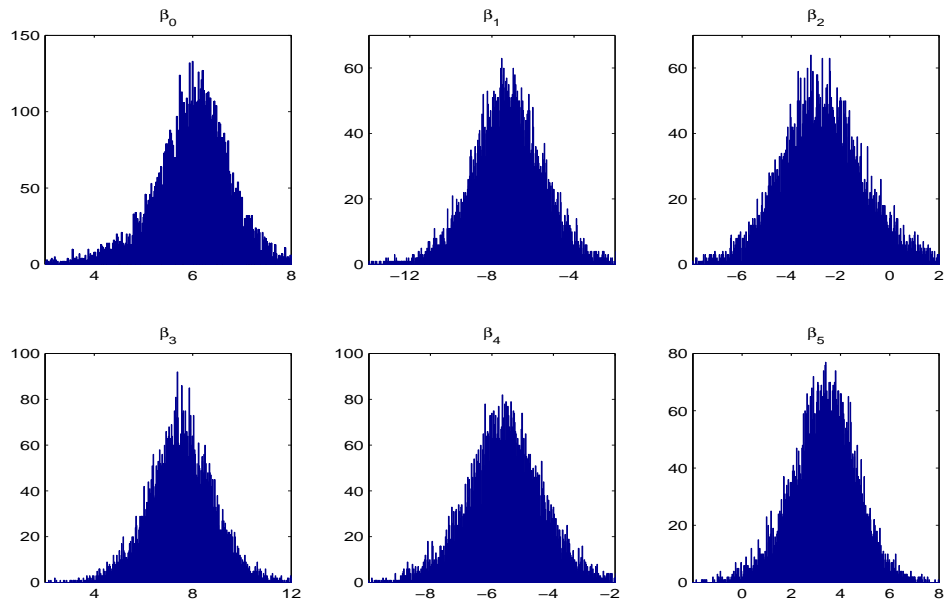


Figure 3.6: Posterior Densities of $\beta_i, i = 0, 1, \dots, 5$.

is chosen as a candidate distribution, which is close to $p(\theta | \mathbf{z})$. This means that to get 10,000 approximately independent samples, we need 1,000,000 samples from the Metropolis-Hastings algorithm, which takes about 30 hours on a 2.20GHz AMD Athlon XP 3200+ PC. However, it took just 46 minutes to get 10,000 independent samples by using our algorithm on the same computer.

3.5 Concluding Remarks and Discussions

Computational difficulties are often cited as a reason not to undertake Bayesian analysis. The simulation algorithm we proposed led to considerable simplification of the calculation necessary to undertake a full Bayesian spatial analysis. The proposed

simulation algorithm works well for a large class of priors including some popularly used priors in spatial area. In addition, the key difference between our algorithm and other MCMC algorithms is that our algorithm will create independent samples from the resulting posterior distribution. Hence we do not have drawbacks such as slow convergence or high autocorrelation that often appear in most MCMC algorithms.

The sufficient condition to implement our algorithm is that there exists a positive number r , such that all of $a(r), b_i^-(r), b_i^+(r), i = 1, \dots, k$ defined by (3.13) are finite, where $h(\boldsymbol{\theta})$ is the posterior density of $\boldsymbol{\theta}$ up to a normalized constant and $\boldsymbol{\theta}$ is the parameter controlling the correlation matrix. We are investigating whether this condition can be relaxed.

This chapter just dealt with the Gaussian random fields. The case of non-Gaussian data is often of interest. We will consider the non-Gaussian random fields in the future.

Chapter 4

Bayesian Spatial Reference

Analysis with Application to Site

Index Prediction

4.1 Introduction

The Missouri Ozark Forest Ecosystem Project (MOFEP) is an on-going, centuries-long experiment that is designed to monitor and assess the short and long-term effects of common management practices on Ozark ecosystems (see Brookshire and Shifley (1997), Shifley and Brookshire (2000) and Shifley and Kabrick (2002)). The MOFEP will provide a comprehensive evaluation of the impacts of operational forest management practices on a wide array of ecosystem attributes. The purpose of this chapter is to predict the site index at unmeasured locations.

Site index is a measure of forest productivity based upon the height at a specified base age of dominant or codominant trees. Height is used because it is correlated to site quality. In Missouri, a base age of 50 years is used. Most measured trees are not at the base age, so curves are used to convert heights of tree of any age to the base age height. A compendium of the published site index curves for eastern half of the United States was presented by Carmean et al. (1989). From this report, the equations by McQuilkin (1974) and McQuilkin (1978) were used for black oak, scarlet oak and white oak, and equations by Nash (1978) were used for shortleaf pine.

From 1993 and 1996, site index was determined on suitable trees at 648 permanent plots on the nine MOFEP sites. Trees considered suitable were canopy codominants having good form with no indication that they had been suppressed and showing the best growth potential. One to five candidate trees selected for site index determination were sampled outside of the 0.5-acre permanent vegetation plots but within 330 feet of vegetation plots. Candidate trees were also limited to four species – black oak, scarlet oak, white oak, and shortleaf pine because these four species are the most abundant commercial species in the region and reliable site index curves have been developed locally for them. The suitability of each tree was ranked by the field crew because good site index trees could not always be found near the vegetation plots.

For each site index tree, the distance and azimuth from the geo-referenced vegetation plot center were recorded and these can be later used to determine the location of each site index tree. Trees were assigned a ranking of their perceived quality for indicating site index. Tree heights were measured with clinometer to the nearest foot.

A single increment core was extracted at breast height and taken to the lab for age determination. Site index was determined using species, height, age at d.b.h., and published site index equations for species in the Missouri Ozarks (see, McQuilkin (1974), Nash (1978)).

Since the site index can only be available on some locations, the prediction of the site index at unmeasured locations is of ecological interest in practice. Recently, Bayesian approaches to analysis of spatial data have seen an upsurge in popularity, especially when the main goal is prediction, see Handcock and Stein (1993), De Oliveira et al. (1997), Ecker and Gelfand (1997) and many others. The main advantage of the Bayesian approach is that parameter uncertainty is fully accounted for when performing prediction and inference, even in small samples.

In this chapter, we will propose Bayesian spatial model to achieve this goal. We will just consider black oaks in sites one and two. However, our proposed method can easily be applied to other species and other sites in the MOFEP study. We will use the data collected in winter of 1995 by MOFEP technicians in the Missouri Department of Conservation (MDC) from each MOFEP vegetation plot (Shifley and Brookshire (2000)). § 4.2 will deal with how to set up an appropriate Bayesian spatial model and statistical inference of the parameters in the model will be given in § 4.3. Model validation will be briefly discussed in § 4.4. In § 4.5, we will discuss spatial prediction of the site index. Finally, some concluding remarks and discussions will be given in § 4.6.

4.2 The Gaussian Model

4.2.1 Data structure and the likelihood

We use the data collected by the Missouri Department Conservation in 1995. There are 173 sampled *black oaks* with good rank and high quality judged by technicians, irregularly located in sites one and two. See Figure 4.1. Site index of each tree was calculated based on the age and the height through the equation provided by McQuilkin (1974). The location of each site index tree was calculated based on the distance and azimuth from the geo-referenced vegetation plot center together with the coordinate of the geo-referenced vegetation plot center. In addition, some environmental characteristics, such as soil type, aspect, land type association, were measured at the location of each site index tree, which will be partly used as covariates in our model.

We choose the Gaussian process to model our spatial data. Let $\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{D}\}$, $\mathcal{D} \subseteq \mathbb{R}^2$, be the random field of interest. The data consist of n observations $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))'$ from a single realization of this random field, where $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ are known distinct sampling locations in \mathcal{D} . We assume that $\mathbf{Z}(\cdot)$ is a Gaussian random field with $\mathbb{E}\{Z(\mathbf{s})\} = \beta_0 + \beta_1 X_1(\mathbf{s}) + \dots + \beta_p X_p(\mathbf{s})$ and $\text{cov}\{Z(\mathbf{s}), Z(\mathbf{u})\} = \sigma^2 K_{\boldsymbol{\theta}}(\|\mathbf{s} - \mathbf{u}\|)$, where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)' \in \mathbb{R}^{p+1}$ are unknown regression parameters, $X_1(\mathbf{s}), \dots, X_p(\mathbf{s})$ are known location-dependent covariates, $\sigma^2 = \text{var}\{Z(\mathbf{s})\}$, and $K_{\boldsymbol{\theta}}(\|\mathbf{s} - \mathbf{u}\|) = \text{corr}\{Z(\mathbf{s}), Z(\mathbf{u})\}$ is an isotropic correlation function with $\|\cdot\|$ denoting Euclidean distance, and $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^c$ controlling the

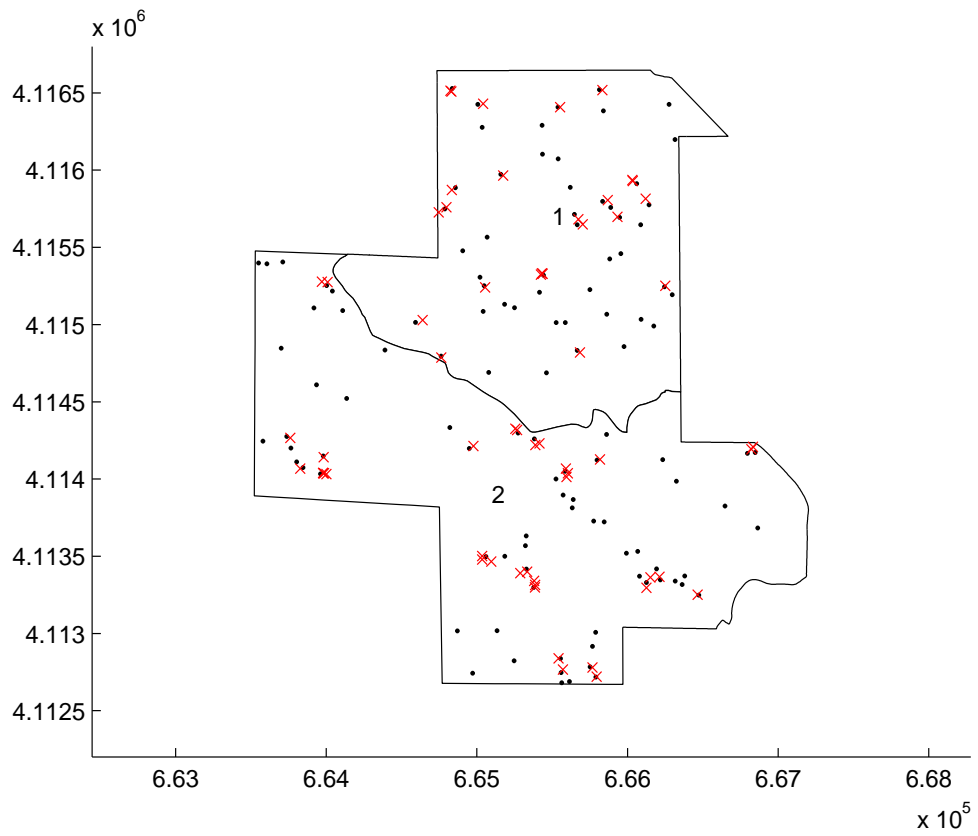


Figure 4.1: 173 Sampled Black Oaks in Sites One and Two of the MOFEP Study: Locations for Modelling Marked as “.” and Those for Model Validation Marked as “×”.

range of correlation and the smoothness of the random field. Thus the likelihood of the model parameters $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$, based on the observed data $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))'$, is given by

$$L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}; \mathbf{z}) = (2\pi\sigma^2)^{-n/2} |\boldsymbol{\Sigma}_{\boldsymbol{\theta}}|^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} (\mathbf{z} - \mathbf{X}\boldsymbol{\beta}) \right\}, \quad (4.1)$$

where $\mathbf{X} = (x_{ij})$ is the known $n \times (p+1)$ matrix with its first column as $\mathbf{1} = (1, \dots, 1)$, assumed to have full rank, and $\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = (K_{\boldsymbol{\theta}}(\|\mathbf{s}_i - \mathbf{s}_j\|))$, assumed to be positive definite for any $\boldsymbol{\theta} \in \Theta$.

For our data set, \mathcal{D} represents the area of sites one and two, and $Z(\mathbf{s})$ denotes the site index of black oak at location \mathbf{s} . Of 173 sampled black oaks, we choose $n = 113$ trees for modelling based on three considerations. First, the distances among 173 points located in sites one and two range from 3.32 meters to 3928.96 meters. In the spatial setting, if we add a location very close to an existing location, the data from the new location will not add much to the inference about the spatial model. Second, if the model contains two very close locations, then the associated correlation matrix $\boldsymbol{\Sigma}_{\boldsymbol{\theta}}$ will be nearly singular, which will result in numerical difficulties for parametric inference with either frequentist or Bayesian methods. Third, if the model has contained enough points, prediction at a new spatial location will improve very slowly with increasing sample size, as mentioned by Banerjee et al. (2004). For 113 locations chosen, the minimum distance among them is 50 meters. As a remedy, we will reserve some of the remaining 60 trees for empirical validation of the resulting predictions.

Covariates are chosen based on availability and ecological background. We pick up

three variables, aspect class, land type association (LTA) and soil depth, denoted as $X_1(\mathbf{s})$, $X_2(\mathbf{s})$ and $X_3(\mathbf{s})$ respectively. Aspect classes are described as ‘protected’ if the slope aspect is within 316 – 135 degrees, and ‘exposed’ if the slope aspect is within 136 – 315 degrees. There are two types of land type association, Current River Oak-Pine Woodland/Forest Hills (coded as OZ9b) and Current River Oak Forest Breaks (coded as OZ9e). See Shifley and Kabrick (2002) for detail. Soil depth was created by soil type and categorized into two classes, deep to very deep soil and shallow to moderate deep soil. See the detail relationship between soil types and soil characteristics in Kabrick et al. (2000). Table 4.1 summarizes the covariates used in the model.

Name	Symbol	Type	Categories	Description
aspect class	$X_1(\mathbf{s})$	categorical	2	1 – protected 0 – exposed
LTA	$X_2(\mathbf{s})$	categorical	2	1 – OZ9b 0 – OZ9e
soil depth	$X_3(\mathbf{s})$	categorical	2	1 – deep to very deep 0 – shallow to moderate deep

Table 4.1: Summary of Covariates

4.2.2 The correlation function

The previous subsection did not specify a correlation function for the spatial model. Actually, there are many correlation functions, such as spherical, power exponential, rational quadratic, that are popular in spatial statistics, see Cressie (1993), Chiles and Delfiner (1999), Stein (1999) and Banerjee et al. (2004), *etc.* Among the various families of correlation function which have been proposed, the Matérn family is particularly attractive. Its algebraic form is given by

$$K_{\theta,\nu}(d) = \frac{1}{2^{\nu-1}\Gamma(\nu)} \left(\frac{d}{\theta}\right)^{\nu} \mathcal{K}_{\nu}\left(\frac{d}{\theta}\right), \quad \theta > 0, \nu > 0, \quad (4.2)$$

where θ is the spatial range parameter, which measures how quickly the correlations of the random field decay with distance, ν is the smoothness parameter, which measures the differentiability of the random field, and $\mathcal{K}_{\nu}(\cdot)$ is the modified Bessel function of the second kind and of order ν ; see Abramowitz and Stegun (1965) for details on the behavior of this special function. What makes this family particularly attractive is that the corresponding process $Z(\cdot)$ is mean-square $[\nu] - 1$ times differentiable where $[\nu]$ represents the largest integer less than or equal to ν . So the Matérn family does allow for great flexibility in the smoothness of the random field while still keeping the number of parameters manageable. Furthermore, note that the exponential family is the special case with $\nu = 1$ and the Gaussian family is the case when $\nu \rightarrow \infty$.

The Matérn family was strongly recommended by Stein (1999) because of the parameter ν , which controls the smoothness of the random field. In what follows, we choose this family as the correlation function for our model.

4.2.3 The prior

The parameters in our model are the regression coefficient $\boldsymbol{\beta}$, the variance σ^2 , the range parameter θ and the smoothness parameter ν . The smoothness of a random field plays a critical role in spatial data analysis. However, there is often no basis for knowing a priori the degree of some physical process modelled as a random field. Thus, we will not assign a prior for the smoothness parameter ν and we assume for the moment that it is known. We will discuss how to estimate the smoothness parameter ν before we make Bayesian inference on the parameters $\boldsymbol{\beta}$, σ^2 and θ .

Selection of the prior is based upon previous knowledge of the model parameters. Often, there is little information available on the model parameters, which prompts the use of noninformative prior. In this chapter, we consider the reference prior for $(\boldsymbol{\beta}, \sigma^2, \theta)$ developed by Berger et al. (2001),

$$\pi(\boldsymbol{\beta}, \sigma^2, \theta) \propto \frac{\pi(\theta)}{\sigma^2}, \quad \boldsymbol{\beta} \in R^{p+1}, \sigma^2 > 0, \theta > 0, \quad (4.3)$$

where

$$\pi(\theta) \propto \left\{ \text{tr}[\mathbf{W}_\theta^2] - \frac{1}{n-p} (\text{tr}[\mathbf{W}_\theta])^2 \right\}^{1/2}, \quad (4.4)$$

with

$$\mathbf{W}_\theta = \frac{\partial \boldsymbol{\Sigma}_\theta}{\partial \theta} \boldsymbol{\Sigma}_\theta^{-1} \mathbf{P}_\theta^\boldsymbol{\Sigma} \quad \text{and} \quad \mathbf{P}_\theta^\boldsymbol{\Sigma} = \mathbf{I} - \mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}. \quad (4.5)$$

Here $(\partial/\partial\theta)\boldsymbol{\Sigma}_\theta$ denotes the matrix obtained by differentiating $\boldsymbol{\Sigma}_\theta$ with respect to θ element by element and \mathbf{I} is an identity matrix of order n .

The above reference prior is a non-informative and improper prior. The posterior

propriety under the reference prior (4.3) has fully been discussed by Berger et al. (2001). They prove that the posterior is always proper.

Remark 4.1 *The prior $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$ in (4.3) with $\pi(\theta) = 1$ was proposed by Kitanidis (1986) and Handcock and Stein (1993). However, as shown by Berger et al. (2001), it will result in an improper posterior.*

4.3 Statistical Inference for Parameters

4.3.1 Estimating the smoothness parameter

As we did not assign a prior for the smoothness parameter ν , we have to estimate this parameter before making Bayesian inference on $\boldsymbol{\beta}, \sigma^2$ and θ .

Because we are now considering the smoothness parameter ν , it is useful to explicitly recognize that the reference prior was defined with ν considered given, so we now write $\pi(\boldsymbol{\beta}, \sigma^2, \theta \mid \nu)$ instead of $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$, and $\pi(\theta \mid \nu)$ instead of $\pi(\theta)$ in this subsection. Although the reference prior (4.3) is improper, Berger et al. (2001) prove that the marginal prior $\pi(\theta \mid \nu)$ is proper. This makes it possible to apply the idea of Berger et al. (1998) to estimate the smoothness parameter ν . The procedure is as follows.

For each ν , the reference prior used is

$$\pi(\boldsymbol{\beta}, \sigma^2, \theta \mid \nu) = \frac{C(\nu)\pi(\theta \mid \nu)}{\sigma^2}, \quad (4.6)$$

where

$$C(\nu) = \frac{1}{\int_0^\infty \pi(\theta | \nu) d\theta}.$$

Note that computation of $C(\nu)$ must be done numerically. We compute $C(\nu)$ by the function *quad* with MATLAB software. Using the prior (4.6), we compute the marginal integrated likelihood for each ν as

$$\begin{aligned} m(\mathbf{z} | \nu) &= \int L(\boldsymbol{\beta}, \sigma^2, \theta, \nu; \mathbf{z}) \frac{C(\nu)\pi(\theta | \nu)}{\sigma^2} d\boldsymbol{\beta} d\sigma^2 d\theta \\ &= \int_0^\infty L^I(\theta, \nu; \mathbf{z}) C(\nu)\pi(\theta | \nu) d\theta \\ &= \int_0^\infty |\boldsymbol{\Sigma}_{\theta, \nu}|^{-1/2} |\mathbf{X}'\boldsymbol{\Sigma}_{\theta, \nu}^{-1}\mathbf{X}|^{-1/2} (S_{\theta, \nu}^2)^{-(n-p)/2} C(\nu)\pi(\theta | \nu) d\theta, \end{aligned} \quad (4.7)$$

where

$$\begin{aligned} L^I(\theta, \nu; \mathbf{z}) &= \int_{R^{p+1} \times (0, +\infty)} L(\boldsymbol{\beta}, \sigma^2, \theta, \nu; \mathbf{z}) \frac{1}{\sigma^2} d\boldsymbol{\beta} d\sigma^2 \\ &\propto |\boldsymbol{\Sigma}_{\theta, \nu}|^{-1/2} |\mathbf{X}'\boldsymbol{\Sigma}_{\theta, \nu}^{-1}\mathbf{X}|^{-1/2} (S_{\theta, \nu}^2)^{-(n-p)/2}, \end{aligned} \quad (4.8)$$

and

$$S_{\theta, \nu}^2 = (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\theta, \nu})' \boldsymbol{\Sigma}_{\theta, \nu}^{-1} (\mathbf{z} - \mathbf{X}\hat{\boldsymbol{\beta}}_{\theta, \nu}) \quad (4.9)$$

is the generalized residual sum of squares with

$$\hat{\boldsymbol{\beta}}_{\theta, \nu} = (\mathbf{X}'\boldsymbol{\Sigma}_{\theta, \nu}^{-1}\mathbf{X})^{-1} \mathbf{X}'\boldsymbol{\Sigma}_{\theta, \nu}^{-1}\mathbf{z} \quad (4.10)$$

being the generalized least squares estimator of $\boldsymbol{\beta}$ given θ and ν .

Based on the idea of Berger et al. (1998), ν can be estimated by maximizing the marginal density $m(\mathbf{z} | \nu)$, that is,

$$\hat{\nu} = \arg \max_{\nu} m(\mathbf{z} | \nu).$$

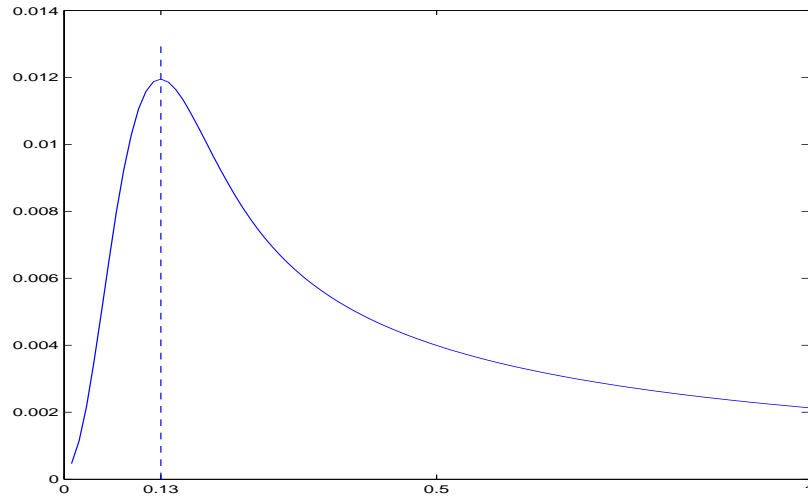


Figure 4.2: The marginal density $m(\mathbf{z} | \nu)$ in terms of ν

Figure 4.2 shows the the marginal integrated likelihood $m(\mathbf{z} | \nu)$ for our dataset. It follows that for our model, the most likely value of ν is 0.13. Thus we will assume $\nu = 0.13$ for the posterior simulation in the next subsection.

Remark 4.2 *Using various integrated likelihood methods for eliminating nuisance parameters has been studied by several authors, such as Liseo (1993), Berger et al. (1999), Liseo (2004) and the references therein.*

4.3.2 Posterior simulation

Standard Bayesian theory tells us that the posterior distribution is determined by,

$$p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z}) \propto L(\boldsymbol{\beta}, \sigma^2, \theta; \mathbf{z})\pi(\boldsymbol{\beta}, \sigma^2, \theta), \quad (4.11)$$

where $L(\boldsymbol{\beta}, \sigma^2, \theta; \mathbf{z})$ is given by (4.1) and $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$ by (4.3), assuming that the smoothness parameter ν is 0.13. The following theorem, which is a special case of Theorem 3.1, will play an important role in the simulation of the posterior distribution.

Theorem 4.1 *The joint posterior distribution of $\boldsymbol{\beta}, \sigma^2, \theta$ has the following decomposition*

$$p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z}) = p(\boldsymbol{\beta} | \sigma^2, \theta; \mathbf{z})p(\sigma^2 | \theta; \mathbf{z})p(\theta | \mathbf{z}), \quad (4.12)$$

where

$$(\boldsymbol{\beta} | \sigma^2, \theta; \mathbf{z}) \sim N_p(\hat{\boldsymbol{\beta}}_\theta, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}), \quad (4.13)$$

$$(\sigma^2 | \theta; \mathbf{z}) \sim IG((n-p)/2 + 1, \mathbf{z}'\{\boldsymbol{\Sigma}_\theta^{-1} - \boldsymbol{\Sigma}_\theta^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\}\mathbf{z}/2), \quad (4.14)$$

$$[\theta | \mathbf{z}] \propto L^I(\theta; \mathbf{z})\pi(\theta) \quad (4.15)$$

with $L^I(\theta; \mathbf{z})$ and $\hat{\boldsymbol{\beta}}_\theta$ being given by (4.8) and (4.10), respectively, except assuming the smoothness parameter ν known.

Sampling from (4.13) and (4.14) is straightforward. We will apply the generalized Ratio-of-Uniforms method proposed by Wakefield et al. (1991) to sample θ from (4.15) because it is not of standard form.

In order to apply the generalized Ratio-of-Uniforms method to sample θ from (4.15), it is important to study the analytical properties of the function $L^I(\theta; \mathbf{z})\pi(\theta)$. Figure 4.3 and Figure 4.4 show the plots of $\pi(\theta)$ and $L^I(\theta; \mathbf{z})\pi(\theta)$ with $\nu = 0.13$, respectively. The following theorem is essential to generate samples from $p(\theta | \mathbf{z})$ by the generalized Ratio-of-Uniforms method.

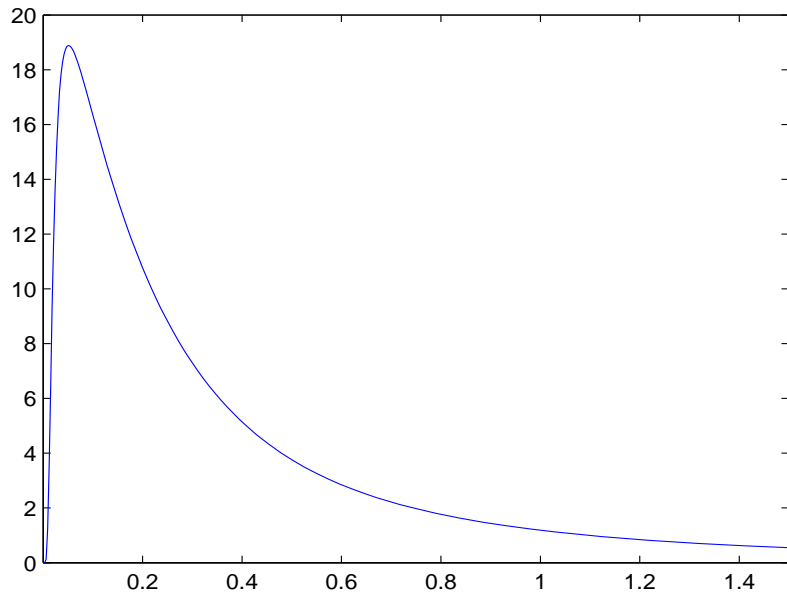


Figure 4.3: The Marginal Prior $\pi(\theta)$ with $\nu = 0.13$.

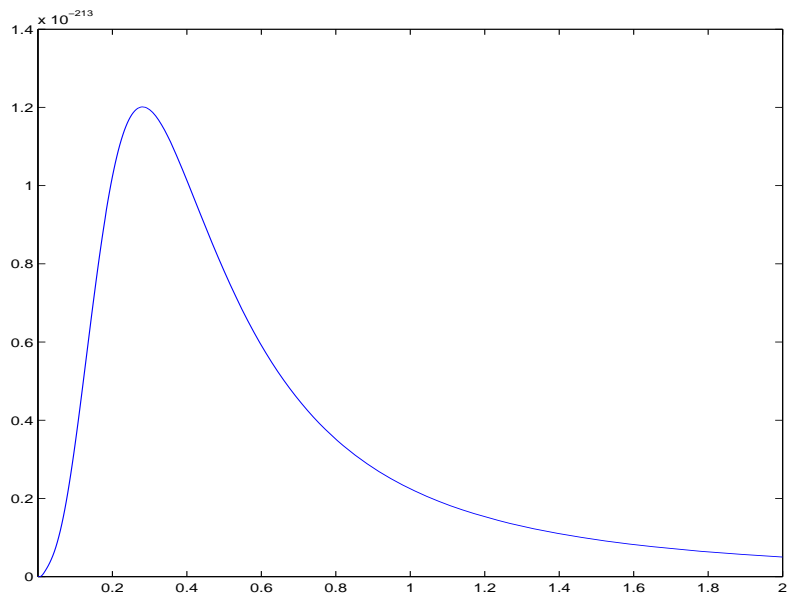


Figure 4.4: The Plot of $L^I(\theta; \mathbf{z})\pi(\theta)$ with $\nu = 0.13$.

Theorem 4.2 For $\nu \neq 1$, there exists a positive number r , such that both $[L^I(\theta; \mathbf{z})\pi(\theta)]^{1/(r+1)}$ and $\theta[L^I(\theta; \mathbf{z})\pi(\theta)]^{r/(r+1)}$ are bounded on $(0, +\infty)$.

Proof. For the Matérn family and the reference prior $\pi(\boldsymbol{\beta}, \sigma^2, \theta)$, Lemma 1 in Berger et al. (2001) showed that $L^I(\theta; \mathbf{z}) \rightarrow c_0$ as $\theta \rightarrow 0^+$ where c_0 is a positive number, $L^I(\theta; \mathbf{z}) = O(1)$ as $\theta \rightarrow +\infty$ and $\pi(\theta) \rightarrow 0^+$ as $\theta \rightarrow 0$. We consider three cases in the following.

(i) $0 < \nu < 1$. Corollary 1 in Berger et al. (2001) stated that

$$\pi(\theta) \propto \frac{1}{\theta^{1+2(1-\nu)}}$$

and thus we may take any $r > 1/[2(1 - \nu)]$.

(ii) ν is greater than 1 and is non-integer. Similarly, we have

$$\pi(\theta) \propto \frac{1}{\theta^{1+2(\nu-1)}}$$

and thus we may take any $r > 1/[2(\nu - 1)]$.

(iii) ν is greater than 1 but is integer. We have

$$\pi(\theta) \propto \frac{1 + 2(\nu - 1)|\log(\theta)|}{\theta^{1+2(\nu-1)}}$$

and thus we may still take $r > 1/[2(\nu - 1)]$. Thus the theorem follows. \square

By choosing an appropriate number r , the algorithm of the generalized Ratio-of-Uniforms method works as follows (notice that $\inf_{\theta} \theta[L^I(\theta; \mathbf{z})\pi(\theta)]^{r/(r+1)} = 0$):

Algorithm for simulation from $p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z})$:

Step 1: Compute

$$a(r) = \sup_{\theta > 0} [L^I(\theta; \mathbf{z})\pi(\theta)]^{1/(r+1)},$$

$$b(r) = \sup_{\theta > 0} \theta [L^I(\theta; \mathbf{z})\pi(\theta)]^{r/(r+1)};$$

Step 2: Simulate

$$U \sim \text{Uniform } [0, a(r)],$$

$$V \sim \text{Uniform } [0, b(r)],$$

and compute $\rho = V/U^r$;

Step 3: If $U \leq [L^I(\rho; \mathbf{z})\pi(\rho)]^{1/(r+1)}$, we accept ρ as a sample θ from $L^I(\theta; \mathbf{z})\pi(\theta)$;

otherwise go back to Step 2;

Step 4: For each θ , simulate

$$\sigma^2 \sim IG((n-p)/2 + 1, \mathbf{z}'\{\boldsymbol{\Sigma}_\theta^{-1} - \boldsymbol{\Sigma}_\theta^{-1}\mathbf{X}(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\}\mathbf{z}/2);$$

Step 5: For each (θ, σ^2) , simulate

$$\boldsymbol{\beta} \sim N_{p+1}(\hat{\boldsymbol{\beta}}_\theta, \sigma^2(\mathbf{X}'\boldsymbol{\Sigma}_\theta^{-1}\mathbf{X})^{-1});$$

Step 6: Go back to Step 2.

For our dataset, we got $\nu = 0.13$ and thus we may take any $r > 0.5747$ from Theorem 4.2. For simplicity, we choose $r = 1$, which means that the basic Ratio-of-Uniforms method proposed by Kinderman and Monahan (1977) can be applied. In this case, $a(1) = 0.1349$ and $b(1) = 0.0584$. The acceptance area with the Ratio-of-Uniforms method is shown in Figure 4.5 and the theoretical acceptance rate for the

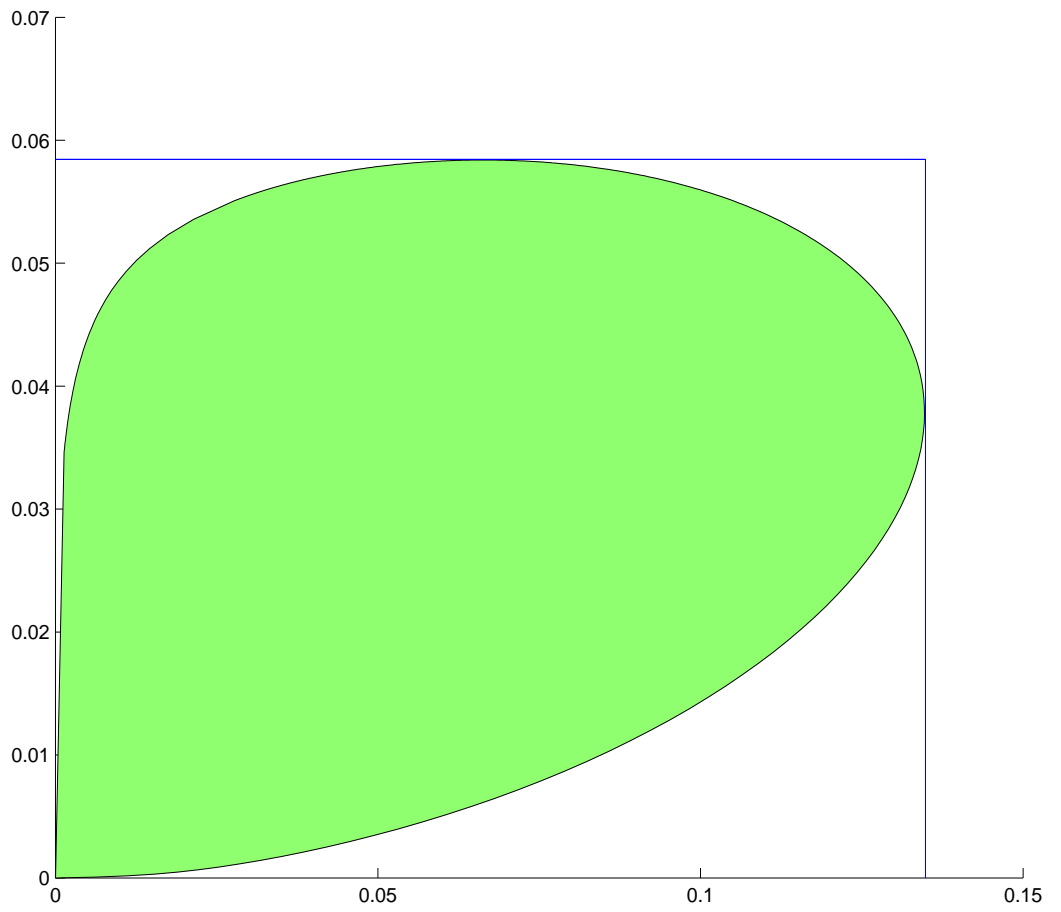


Figure 4.5: The Acceptance Area with the Ratio-of-Uniforms Method

simulation is around 76 percent. In fact, we obtained 10,000 samples by sampling 13101 pairs of (u, v) , which means the actual acceptance rate is 76.33 percent in our simulation. The simulation took about 52 minutes on a 2.20GHz AMD Athlon XP 3200+ PC. In addition, notice that our simulation produces independent samples for $(\boldsymbol{\beta}, \sigma^2, \theta)$, which is advantageous over other traditional MCMC methods when making inference of parameters or prediction.

Remark 4.3 *The acceptance rate of the generalized Ratio-of-Uniforms method depends on the value r chosen. It is unclear how to get the best choice of r . However, our experience shows that a good choice is the integer that is closest to $1/(2|\nu - 1|)$ and greater than or equal to 1.*

Remark 4.4 *If the smoothness parameter ν is exactly equal to one, it is still unclear whether the generalized Ratio-of-Uniforms method can be applied.*

Figure 4.6 gives the histogram of θ based on these 10,000 independent samples. We can see that the marginal posterior density of the range parameter θ is positively skewed and heavy-tailed, which commonly appears in spatial area. This property usually results in difficulties in spatial simulation because it is often difficult to get samples in the tail. See Banerjee et al. (2004) and Møller (2003) and the references therein. The histograms of σ^2 and $\beta_0, \beta_1, \beta_2, \beta_3$ are shown in Figures 4.7 and 4.8, respectively. Table 4.2 shows some posterior quantities from the simulation.

From Figure 4.8 or Table 4.2, we know that both the aspect class and the soil depth are very significant. The land type association is less significant in the model.

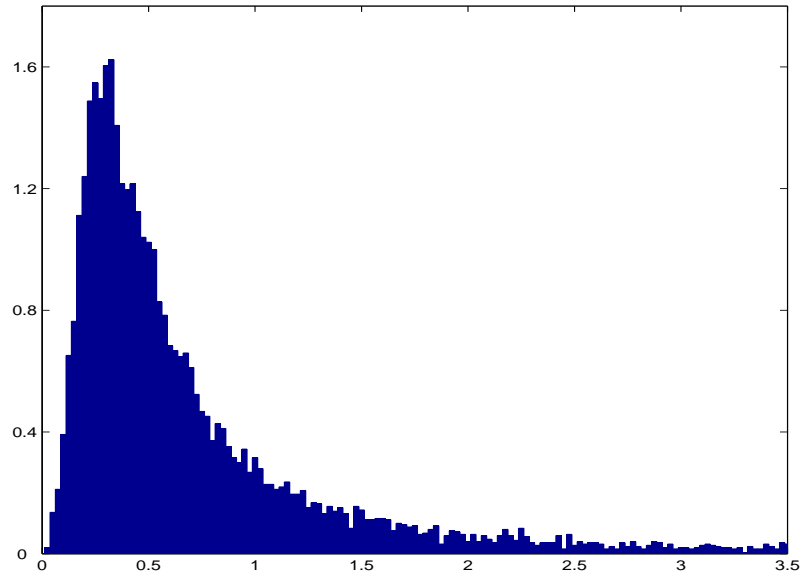


Figure 4.6: Histogram of θ for the Model in § 4.2.

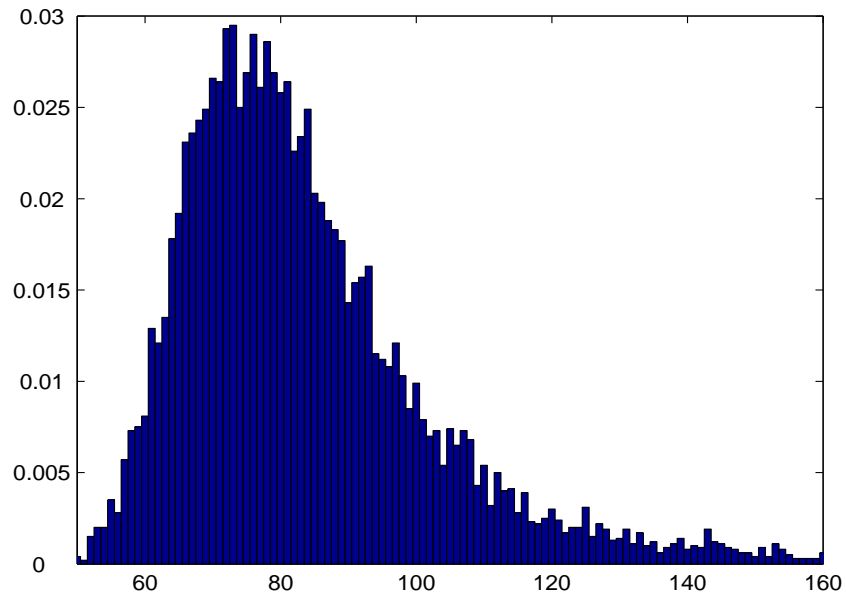


Figure 4.7: Histogram of σ^2 for the Model in § 4.2.

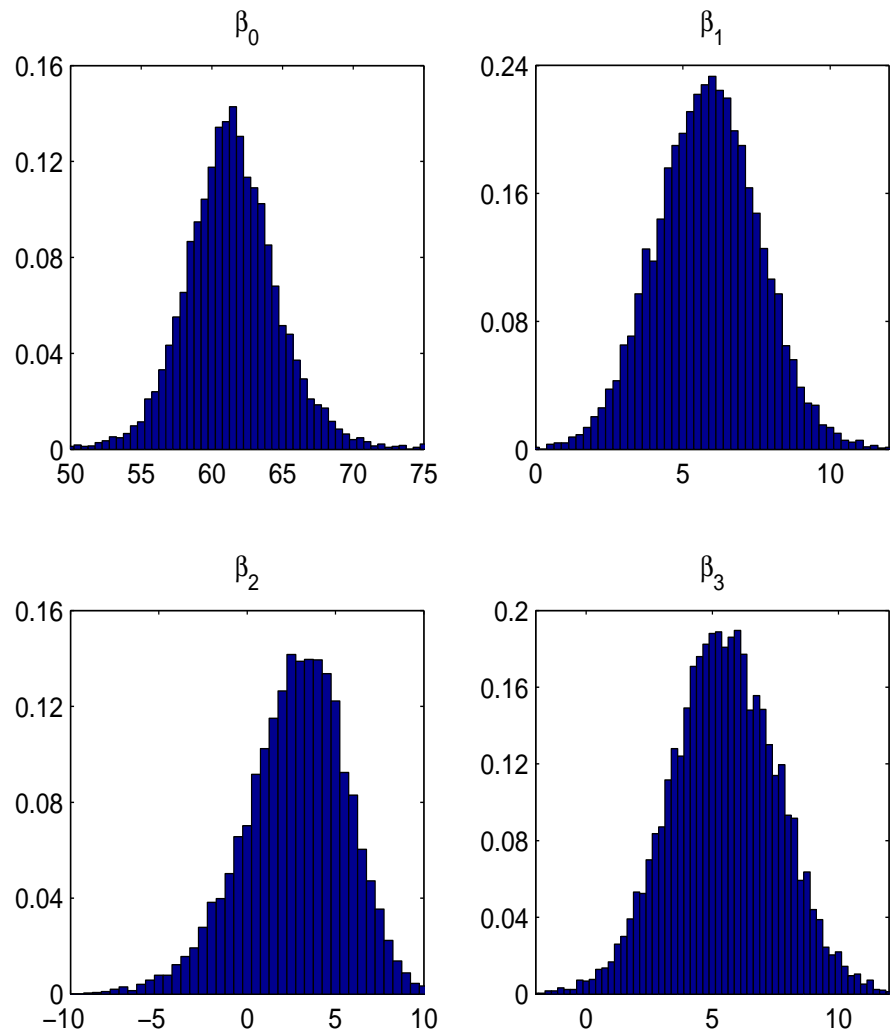


Figure 4.8: Histograms of $\beta_0, \beta_1, \beta_2, \beta_3$ for the Model in § 4.2.

parameter	median	mean	standard deviation	95% Bayesian Credible Interval
β_0	61.34	61.40	3.48	[54.91, 68.36]
β_1	5.86	5.84	1.74	[2.44, 9.27]
β_2	2.93	2.73	2.92	[-3.65, 7.93]
β_3	5.48	5.47	2.11	[1.40, 9.59]
σ^2	79.99	84.88	22.70	[58.34, 135.54]
θ	0.50	1.20	10.78	[0.12, 5.37]

Table 4.2: Posterior Quantities of $(\boldsymbol{\beta}, \sigma^2, \theta)$.

Figure 4.6 also tells us that the approximate mode of the marginal posterior density of the range parameter θ is around 0.28, which is smaller than the posterior median 0.50 . As mentioned in § 3.4, we would like to choose the posterior mode as an appropriate point estimate of the range parameter θ for our model. This implies that the *effective range*, which is the distance at which the correlation drops to only 0.05, is about 0.450 kilometers or 450 meters.

4.4 Model Validation

Assessing model adequacy is very important and fundamental in Bayesian data analysis, since the analysis can be misleading when the model is not adequate. The literature on Bayesian model adequacy is very extensive; for example, see Box (1980), Geisser (1993), Gelfand et al. (1992), Dey et al. (1997), and many others. When the

observations are independent, the cross-validation approach, in which the predictive distribution is used in various ways to assess model adequacy, are popularly used. The main idea of this cross-validation approach is to validate conditional predictive distribution arising from single observation deletion against observed responses. For spatial problems, the observations measured at different locations are correlated. If the sample size is small, this approach is still applicable, for example, see De Oliveira et al. (1997). However, this approach is not appropriate for a large spatial dataset because of very expensive computation. Now we will use some of 60 samples trees that are not contained in the model for assessing model adequacy.

From the 60 sampled trees that are outside the model, we pick 29 trees, such that the distance between each of the 29 trees and any of the 113 trees in the model is at least 20 meters. We then calculate the predictive distribution of the site index at each of 29 locations as follows:

$$p(z_0 | \mathbf{z}) = \int p(z_0 | \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z}) d\boldsymbol{\beta} d\sigma^2 d\theta, \quad (4.16)$$

where $p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z})$ is the posterior distribution described by (4.11) and $p(z_0 | \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta)$ is determined by

$$p(\mathbf{z}, z_0 | \boldsymbol{\beta}, \sigma^2, \theta) \sim N_{n+1} \left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{x}'(\mathbf{s}_0)\boldsymbol{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma}_\theta & \mathbf{k}_\theta \\ \mathbf{k}'_\theta & 1 \end{pmatrix} \right). \quad (4.17)$$

Here $\mathbf{x}(\mathbf{s}_0)$ is the covariate vector evaluated at \mathbf{s}_0 and $\sigma^2 \mathbf{k}_\theta$ is $n \times 1$ vector of covariances of z_0 with $(z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))$. In fact, we have

$$p(z_0 | \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) \sim N(\mu_0, \sigma_0^2) \quad (4.18)$$

where

$$\mu_0 = \mathbf{x}'(\mathbf{s}_0)\boldsymbol{\beta} + \mathbf{k}'_{\theta}\boldsymbol{\Sigma}_{\theta}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}), \quad \sigma_0^2 = \sigma^2(1 - \mathbf{k}'_{\theta}\boldsymbol{\Sigma}_{\theta}^{-1}\mathbf{k}_{\theta}). \quad (4.19)$$

Based on the samples generated from the posterior distribution $p(\boldsymbol{\beta}, \sigma^2, \theta \mid \mathbf{z})$ in the previous section, we can easily obtain the samples from the predictive distribution and then calculate the 95 percent Bayesian credible interval for each z_0 at location s_0 . The detailed algorithm will be described in the next section. Of 29 locations for the purpose of model validation, only 6 observed measurements are outside their corresponding 95 percent Bayesian credible intervals. In addition, no overprediction or underprediction tendency is noted. So the model proposed in § 4.2 seems to perform adequately. We will also justify our model by discussing the variability of the predictive distribution in the next section.

4.5 Spatial Prediction of the Site Index

4.5.1 Bayesian prediction at one location

Modelling point-referenced data is not only useful for identifying significant covariates but for producing smooth maps of the outcome by predicting it at unsampled locations. Spatial prediction is usually referred to as kriging and is popularly used in spatial problems. In the classical framework a lot of energy is devoted to the determination of the optimal estimates to plug into the predictive equation. However, as pointed out by Le and Zidek (1992), classical kriging methodology fails to incorporate parameter uncertainty when performing prediction and inference. This deficiency

leads to unwarranted confidence in the interpolated values and, essentially, to seemingly valid decisions or regulatory actions which are, in fact, unjustified. Bayesian approaches to spatial interpolation avoid this deficiency by considering uncertainty about the parameters in the model.

For each unmeasured site \mathbf{s}_0 , we obtain its covariate information from the GIS database provided by the Missouri Department of Conservation. Given samples from the posterior, simulation of realizations from (4.16) is straightforward. In fact, by (4.16) and (4.18), we just need to add

Step 5:* for given $(\boldsymbol{\beta}, \sigma^2, \theta)$, simulate z_0 from $N(\mu_0, \sigma_0^2)$ with (μ_0, σ_0^2) given by (4.19);

before *Step 6* of the algorithm in § 4.3.2 and thus we get independent samples of z_0 from the posterior predictive distribution $p(z_0 | \mathbf{z})$ in (4.16).

Figure 4.9 shows the histogram of the z_0 at (665614.875, 4115457.5), which is close to the center of site one. Figure 4.10 shows the histograms of the z_0 at other four locations, which are near the border of size one. Each histogram is based on a sample of size 5000. In fact, we found that the sample size 300 is good enough to get the approximate predictive density of z_0 . We can also see that each posterior predictive distribution is unimodal and approximately symmetric, which implies that its mean, median and mode are almost same. In addition, the variation of each predictive distribution is relatively small, comparing to the posterior distribution of the variance σ^2 . For example, the standard deviation of the predictive distribution at (665564.875, 4115337.5) is just about 2.80 for our 5000 samples. This also gives us

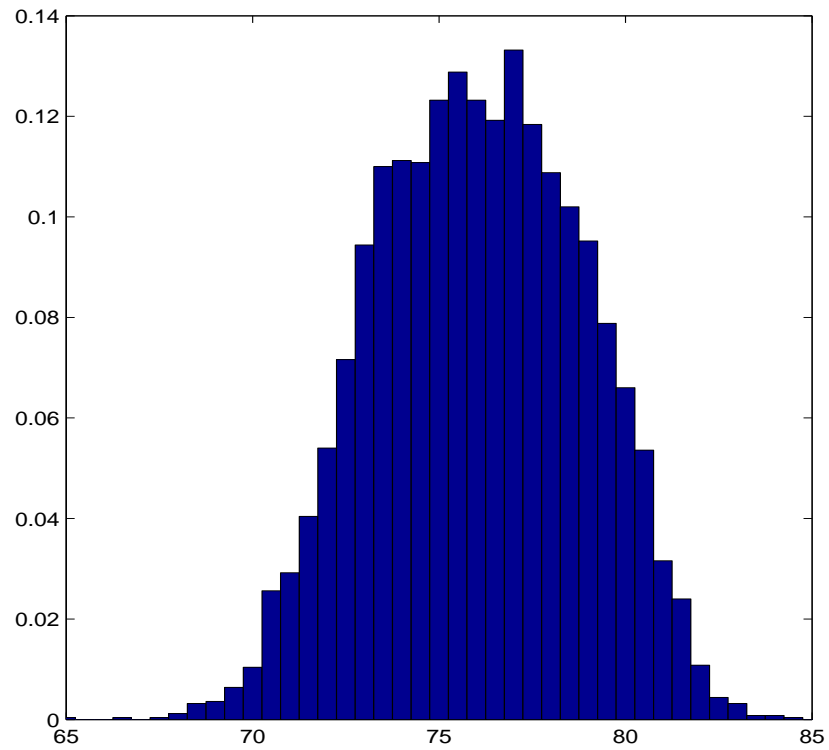


Figure 4.9: Histogram of z_0 Sampling from the Predictive Distribution at (665614.875, 4115457.5): Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth– Deep)

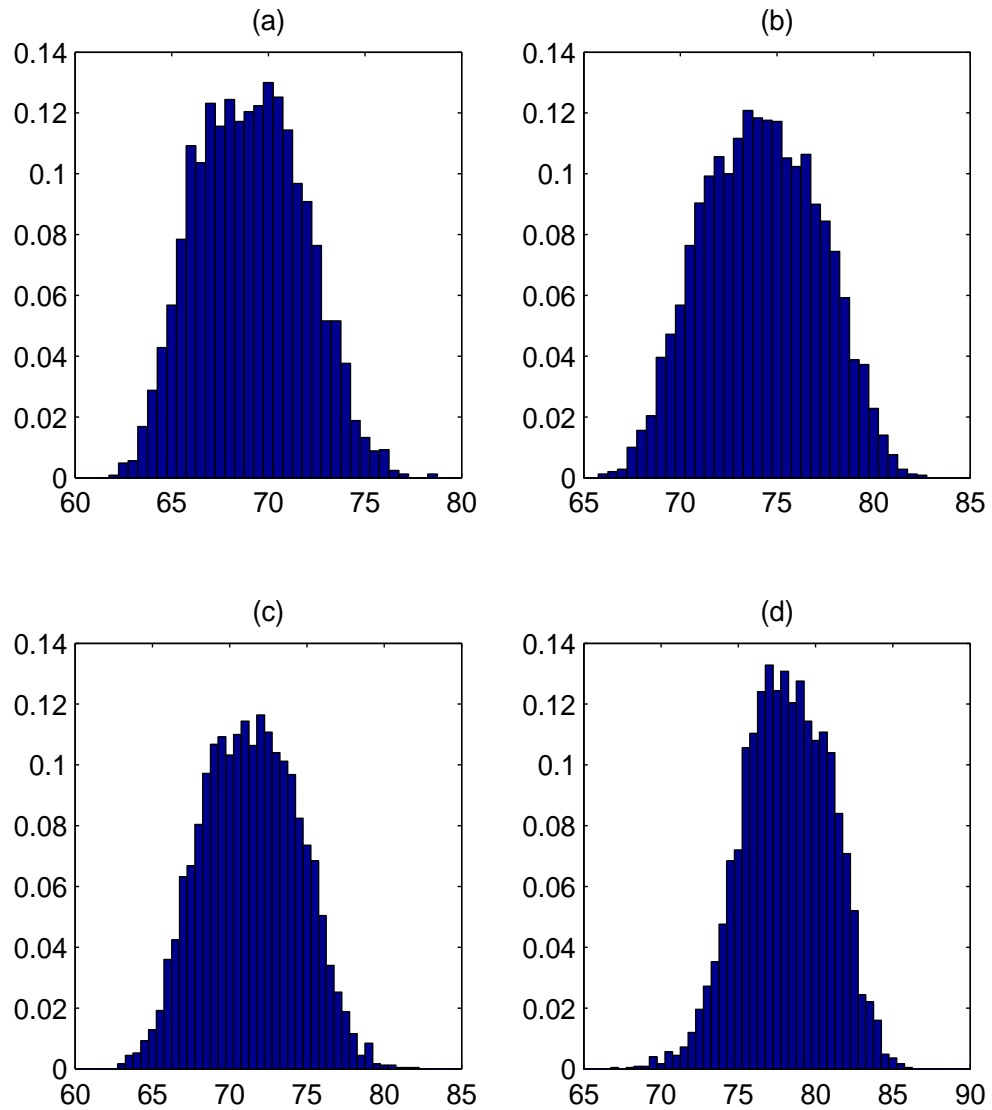


Figure 4.10: Histograms of the Predictive Distribution at Four Different Locations: (a) Location (666194.875, 4116397.5), Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep to Very Deep); (b) Location (666124.875, 4114517.5), Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep to Very Deep); (c) Location (664474.875, 4115407.5), Covariate Information (Aspect Class–Protected, Land Type Association–OZ9e and Soil Depth–Shadow to Deep); (d) Location (665564.875, 4115337.5), Covariate Information (Aspect Class–Exposed, Land Type Association–OZ9b and Soil Depth–Deep to Very Deep)

confidence that our proposed model is appropriate.

In order to get a prediction map of the site index, we make a grid of 10 meters by 10 meters on site one and thus there are 38,744 unmeasured locations. Figures 4.11, 4.12 and 4.13 present the covariate information for the aspect class, the land type association and the soil depth in site one, respectively. Based on the prediction procedure described in this section, Figure 4.14 shows a prediction map for site one. The map is made based on the median of 300 samples generated from the posterior predictive distribution at each grid point. The whole simulation using the above method took about 125 hours on a 2.20GHz AMD Athlon XP 3200+ PC. For convenience, we also provide a map of the standard deviation of the site index prediction in Figure 4.15.

4.5.2 Bayesian prediction of spatial block average

Sometimes it is of interest to predict the average site index $Z(\mathcal{B})$ over a subregion $\mathcal{B} \subset \mathcal{D}$, that is,

$$Z(\mathcal{B}) = |\mathcal{B}|^{-1} \int_{\mathcal{B}} Z(\mathbf{s}) d\mathbf{s}, \quad (4.20)$$

where $|\mathcal{B}|$ denotes the area of \mathcal{B} . The integration in (4.20) is the average of random variables, and hence a random or stochastic integral. To predict $Z(\mathcal{B})$ based on the dataset \mathbf{z} , we require the predictive distribution, which is now

$$p(z(\mathcal{B}) | \mathbf{z}) = \int p(z(\mathcal{B}) | \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z}) d\boldsymbol{\beta} d\sigma^2 d\theta \quad (4.21)$$

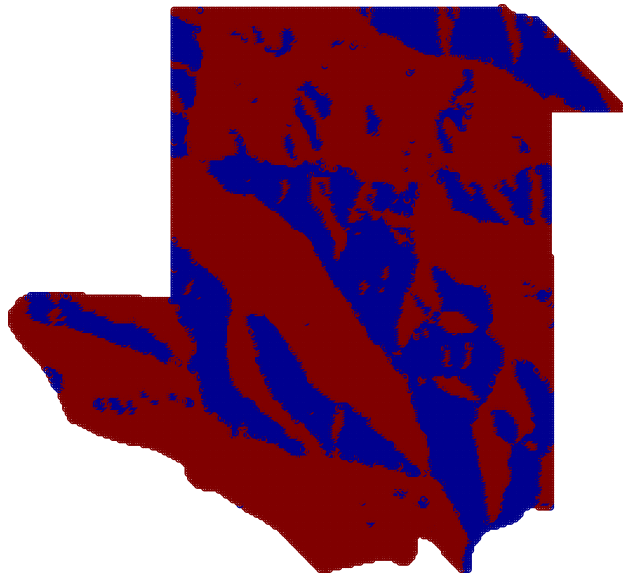


Figure 4.11: The Map of the Aspect Class at Site One: the Red Represents the Area with Protected and the Blue Represents the Area with Exposed



Figure 4.12: The Map of the Land Type Association at Site One: the Red Represents Current River Oak-Pine Woodland/Forest Hills (OZ9b) and the Blue Represents Current River Oak Forest Breaks (OZ9e).

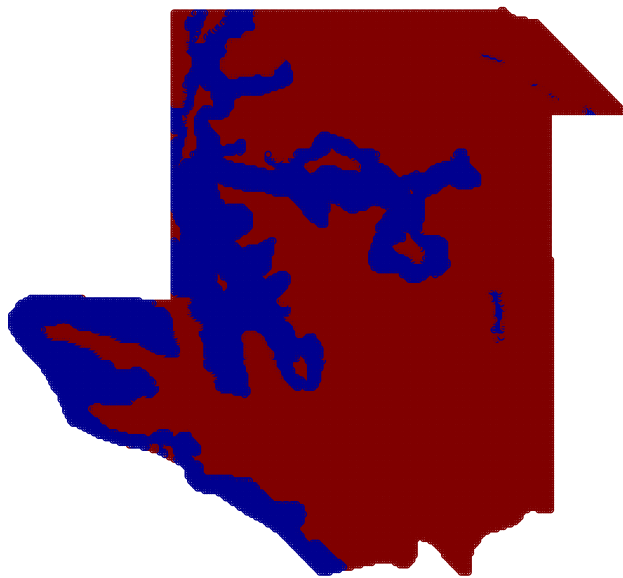


Figure 4.13: The Map of the Soil Depth at Site One: the Red Represents the Soil Depth Varying from Deep to Very Deep and the Blue Represents the Soil Depth Varying from Shallow to Deep.

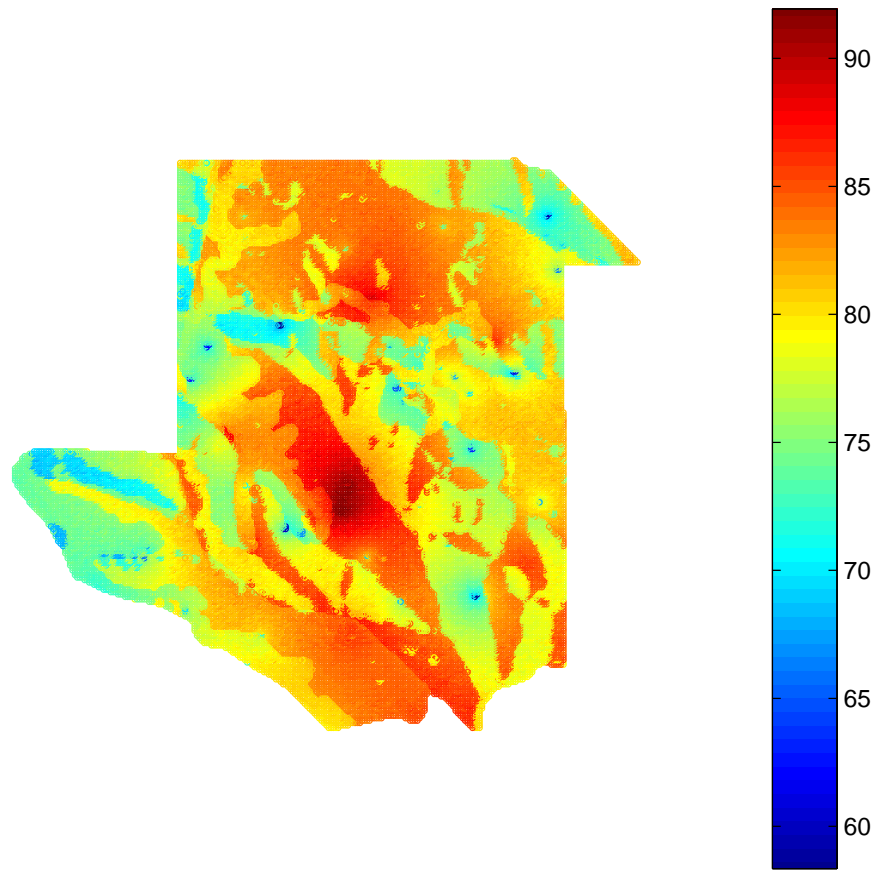


Figure 4.14: The Prediction Map of the Site Index at Site One

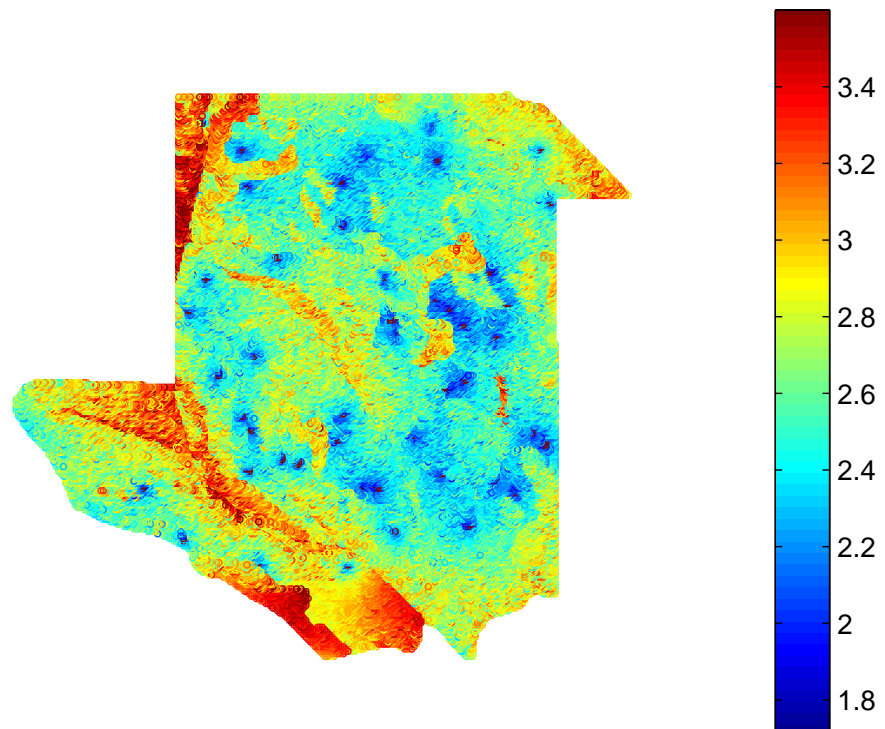


Figure 4.15: The Map of Standard Deviation of the Site Index Prediction at Site One

Under the Gaussian process in § 4.2, we have

$$p(\mathbf{z}, z(\mathcal{B}) \mid \boldsymbol{\beta}, \sigma^2, \theta) \sim N_{n+1} \left(\begin{pmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{x}'_{\mathcal{B}}\boldsymbol{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \boldsymbol{\Sigma}_{\theta} & \mathbf{k}_{\theta, \mathcal{B}} \\ \mathbf{k}'_{\theta, \mathcal{B}} & 1 \end{pmatrix} \right). \quad (4.22)$$

where

$$\mathbf{x}_{\mathcal{B}} = |\mathcal{B}|^{-1} \int_{\mathcal{B}} X(\mathbf{s}) d\mathbf{s} \quad (4.23)$$

and the i th element of $\mathbf{k}_{\theta, \mathcal{B}}$ is given by

$$(\mathbf{k}_{\theta, \mathcal{B}})_i = |\mathcal{B}|^{-1} \int_{\mathcal{B}} K_{\theta}(\mathbf{s}_i - \mathbf{s}) d\mathbf{s}. \quad (4.24)$$

Analogously to (4.18), we get

$$p(z(\mathcal{B}) \mid \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) \sim N(\mu_{\mathcal{B}}, \sigma_{\mathcal{B}}^2), \quad (4.25)$$

where

$$\mu_{\mathcal{B}} = \mathbf{x}'_{\mathcal{B}}\boldsymbol{\beta} + \mathbf{k}'_{\theta, \mathcal{B}}\boldsymbol{\Sigma}_{\theta}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta}), \quad \sigma_{\mathcal{B}}^2 = \sigma^2(1 - \mathbf{k}'_{\theta, \mathcal{B}}\boldsymbol{\Sigma}_{\theta}^{-1}\mathbf{k}_{\theta, \mathcal{B}}). \quad (4.26)$$

Riemannian approximation to integrate over \mathcal{B} for (4.23) and (4.24) may be difficult in practice, especially when \mathcal{B} is irregularly shaped. Instead, noting that each integration in (4.23) and (4.24) is an expectation with respect to a uniform distribution over the subregion \mathcal{B} , we may use Monte Carlo integration. In particular, we draw a set of locations $\mathbf{s}_{j, \mathcal{B}}$, $j = 1, \dots, m$, distributed independently and uniformly over \mathcal{B} and then replace $\mathbf{x}_{\mathcal{B}}$ and $(\mathbf{k}_{\theta, \mathcal{B}})_i$ by

$$\tilde{\mathbf{x}}_{\mathcal{B}} = \frac{1}{m} \sum_{j=1}^m \mathbf{x}(\mathbf{s}_{j, \mathcal{B}}), \quad (\tilde{\mathbf{k}}_{\theta, \mathcal{B}})_i = \frac{1}{m} \sum_{j=1}^m K_{\theta}(\mathbf{s}_i - \mathbf{s}_{j, \mathcal{B}}),$$

respectively. See Gotway and Young (2002) or Banerjee et al. (2004) and the references therein for details. The simulation of the predictive distribution of (4.21) is then similar to that in the previous subsection.

4.5.3 Bayesian prediction of spatial CDFs

For the spatial process $Z(\cdot)$, we might wish to find the proportion of area in a subregion $\mathcal{B} \subset \mathcal{D}$ that has site index above some level w . This proportion is the random variable

$$G_{\mathcal{B}}(w) = \Pr(\mathbf{s} \in \mathcal{B} : Z(\mathbf{s}) \geq w) = |\mathcal{B}|^{-1} \int_{\mathcal{B}} I(Z(\mathbf{s}) \geq w) d\mathbf{s}, \quad (4.27)$$

where $I(A)$ is an indicator function. Prediction of $G_{\mathcal{B}}(w)$ is equivalent to that of the spatial cumulative distribution function (CDF) $F_{\mathcal{B}}(w)$ defined by

$$F_{\mathcal{B}}(w) = 1 - G_{\mathcal{B}}(w) = \Pr(\mathbf{s} \in \mathcal{B} : Z(\mathbf{s}) < w) = |\mathcal{B}|^{-1} \int_{\mathcal{B}} I(Z(\mathbf{s}) < w) d\mathbf{s}, \quad (4.28)$$

which has been studied by many authors such as Lahiri (1999) Lahiri et al. (1999) and Zhu et al. (2002) etc.

Although that (4.27) may be studied analytically in some cases, it is difficult to work with in practice. We can still apply the idea of the Monte Carlo integration in § 4.5.2 to approximate $G_{\mathcal{B}}(w)$ first and then predict such an approximation through Bayesian analysis. See Banerjee et al. (2004) for details.

Similar discussion can be applied to the prediction of spatial extremes and their extent, see Cressie et al. (2004) and Craigmile et al. (2004) and the references therein.

4.6 Concluding Remarks and Discussions

In this chapter, we discuss how to model the site index dataset provided by the Missouri Department of Conservation and then predict the site index at unsampled

locations. As an example, we choose black oaks in sites one and two for analysis. Based on ecological background and availability, we choose three variables, aspect class, land type association and soil depth as covariates. To allow great flexibility of the smoothness of the random field, we choose the Matérn family as the correlation function. We also choose the reference prior as an appropriate prior because there is no previous knowledge of the parameters in the model and thus an appropriate Bayesian spatial model is established. An efficient algorithm based on the generalized Ratio-of-Uniforms method is applied for the posterior simulation. One advantage of the algorithm is that this simulation method can generate independent samples from the required posterior distribution, which is much more efficient for both statistical inference of the parameters and prediction of the site indexes at unsampled locations. Our results show that aspect class and soil depth are both significant while land type association is less significant.

It appears that the prediction map Figure 4.14 is not smooth enough, especially around the observed data locations. The main reason for this is that we did not consider the measurement error of the site index of a tree in our model and thus theoretically, the prediction errors at measured locations were zero(see Figure 4.15). However, the measurement errors always existed for the site indexes of trees (some may be high) because it is impossible to measure heights of trees accurately and in addition, the formula used for calculation from the tree height to the site index is approximate. One good solution is to propose a spatial model with the nugget effect (measurement error) as follows.

Suppose that we have observe n observations $\mathbf{Z} = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))'$ from a single realization of this random field $\{Z(\mathbf{s}), \mathbf{s} \in \mathcal{D}\}$, where $\mathcal{D} \subseteq \mathbb{R}^2$ is the random field of interest. Assume that $\mathbf{Z}(\cdot)$ is a Gaussian random field with $\mathbb{E}\{Z(\mathbf{s})\} = \beta_0 + \beta_1 X_1(\mathbf{s}) + \dots + \beta_p X_p(\mathbf{s})$. and

$$\text{cov}\{Z(\mathbf{s}), Z(\mathbf{u})\} = \begin{cases} \sigma^2 + \tau^2, & \text{if } \mathbf{s} = \mathbf{u}, \\ \sigma^2 K_{\theta, \nu}(\|\mathbf{s} - \mathbf{u}\|), & \text{if } \mathbf{s} \neq \mathbf{u}, \end{cases} \quad (4.29)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)' \in \mathbb{R}^{p+1}$ are unknown regression parameters, $X_1(\mathbf{s}), \dots, X_p(\mathbf{s})$ are known location-dependent covariates, $\sigma^2 + \tau^2 = \text{var}\{Z(\mathbf{s})\}$, and $K_{\theta, \nu}(\|\mathbf{s} - \mathbf{u}\|)$ is the Matérn correlation function given by (4.2). τ^2 is often called the nugget effect in spatial areas. Therefore, the likelihood of the model parameters $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu)$, based on the observed data $\mathbf{z} = (z(\mathbf{s}_1), \dots, z(\mathbf{s}_n))'$, is given by

$$L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu; \mathbf{z}) = (2\pi)^{-n/2} |\mathbf{G}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})' \mathbf{G}^{-1}(\mathbf{z} - \mathbf{X}\boldsymbol{\beta})\right\}, \quad (4.30)$$

where $\mathbf{X} = (x_{ij})$ is the known $n \times (p+1)$ matrix with its first column as $\mathbf{1} = (1, \dots, 1)$, assumed to have full rank, and

$$\mathbf{G} = \tau^2 \mathbf{I} + \sigma^2 \boldsymbol{\Sigma} \quad (4.31)$$

with $\boldsymbol{\Sigma} = (K_{\theta, \nu}(\|\mathbf{s}_i - \mathbf{s}_j\|))_{n \times n}$.

Because there is one more parameter, the nugget effect τ^2 in the likelihood function (4.30) than in (4.1), it appears impossible to develop the corresponding reference prior, even although the smoothness parameter ν is assumed to be known. As suggested by Agarwal and Gelfand (2005), we may consider the following prior

$$\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\theta)\pi(\nu), \quad (4.32)$$

where

$$\boldsymbol{\beta} \sim N_p(\mathbf{0}, d\mathbf{I}_p),$$

$$\sigma^2 \sim \text{IG}(a_1, b_1),$$

$$\tau^2 \sim \text{IG}(a_2, b_2),$$

$$\theta \sim \text{IG}(a_3, b_3),$$

$$\nu \sim \text{U}(0, 2),$$

where a_i, b_i and d are specified. $\text{IG}(a, b)$ refers to an inverse gamma distribution with shape parameter a and scale parameter b , and $\text{U}(a, b)$ is a uniform distribution on the interval (a, b) . For these specifications, the prior $\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu)$ given by (4.32) is proper and therefore the corresponding posterior distribution should be proper as well.

Note that for $\text{IG}(a, b)$, the variance is infinite when the shape parameter a is 2. The hyperparameters are chosen so that the prior is proper but vague enough so that inference is driven by the data. In practice, we may choose $d = 100$, $a_1 = a_2 = a_3 = 2$. b_1, b_2, b_3 may be chosen such that each of them is close to the MLE or REML of its corresponding parameter.

Since the algorithm in Chapter 3 is not applicable, a new simulation algorithm is needed. We may apply the Gibbs sampling for the posterior simulation and thus the full conditional distributions are needed. The full conditional distribution of $\boldsymbol{\beta}$ can be easily obtained as

$$(\boldsymbol{\beta} \mid \sigma^2, \tau^2, \theta, \rho; \mathbf{z}) \sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}),$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} + d^{-1}\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{G}^{-1}\mathbf{z}, \quad (4.33)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{G}^{-1}\mathbf{X} + d^{-1}\mathbf{I}_p)^{-1}. \quad (4.34)$$

with \mathbf{G} given by (4.31). However, the full conditional distributions of $\sigma^2, \tau^2, \theta, \nu$ are not standard forms. We will update each of them with the slice sampler, which can be found in Section 5.3. The proposed algorithm is as follows.

Algorithm:

Step 1: For given $(\sigma^2, \tau^2, \theta, \nu)$, update $\boldsymbol{\beta}$ by simulating

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}),$$

where $\boldsymbol{\mu}_{\boldsymbol{\beta}}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\beta}}$ are given by (4.33) and (4.34), respectively;

Step 2: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu)$, simulate $R \sim \exp(1)$ and let

$$V = R - \log L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu; \mathbf{z});$$

Step 3: For given $(\boldsymbol{\beta}, \tau^2, \theta, \nu)$, update σ^2 by simulating $\sigma^2 \sim IG(a_1, b_1)$ until

$$V > -\log L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \nu; \mathbf{z}); \quad (4.35)$$

Step 4: For given $(\boldsymbol{\beta}, \sigma^2, \theta, \nu, V)$, update τ^2 by simulating $\tau^2 \sim IG(a_2, b_2)$ until

the condition (4.35) is satisfied;

Step 5: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \nu, V)$, update θ by simulating $\theta \sim IG(a_3, b_3)$ until

the condition (4.35) is satisfied;

Step 6: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, V)$, simulate $\nu \sim U(0, 2)$ until the condition

(4.35) is satisfied;

Step 7: Go back to Step 1 until we get the appropriate number of posterior samples.

In order to speed the updating processes, we may apply the shrinkage slice sampler described in § 5.3.2 in Steps 4, 5 and 6.

In addition, we may consider a multivariate Bayesian spatial model by considering the four species together to enhance the prediction.

Chapter 5

Bayesian Spatial Models with Repeated Measurements — with Application to the Herbaceous Data Analysis

5.1 Introduction

Ground layer vegetation is an important component of any forest ecosystem and has been shown to be a useful indicator of attributes such as disturbance history, site productivity, and potential responses to management (Whittaker, 1967, Host and Pregitzer, 1991). The study of the Missouri Ozark Forest Ecosystem Project (MOFEP) is the first comprehensive effort to describe and evaluate the upland ground flora of

the southeastern Missouri Ozarks.

MOFEP ground flora was sampled in the same plots used for the woody vegetation. There are 70 to 76 permanent 0.2-ha (0.5-ac) circular plots located within each of nine study sites for a total of 648 MOFEP vegetation plots. Pre-treatment ground flora data were collected each summer from 1991 through 1995. Principal data collection was conducted between early June and late August each year. Sampling sometimes extended through September because of either limited field personnel or intentional resampling to address questions about seasonal effects.

There were four subplots in each plot. For each subplot, there were four 1-m^2 quadrats located 6.7 meters from the subplot center at $45^\circ, 135^\circ, 225^\circ, 315^\circ$. See Figure 5.1.

Within each 1-m^2 quadrat, all vascular species with live foliage less than 1 meter tall above the ground were identified and assigned an estimate of percentage coverage. The sample included plants not rooted in the quadrat frame but with leaves hanging over it. Individual species coverages were estimated as if no other species were in the quadrat (*i.e.*, overlap among leaves of different species was ignored). In addition, the total percentage coverage by live vegetation less than 1 meter in each quadrat, including vascular and non-vascular, was measured. See Grabner (2000).

The main aim of this chapter is to study the spatial effect of the total vegetation coverage in the MOFEP sites. We may choose the center of a quadrat as the location for the measurement of total vegetation coverage in this quadrat and apply the typical spatial tool described in the previous two chapter or Cressie (1993) for analysis.

However, the centers of four quadrats in the same subplot are very close to each other and this results in the singularity of the covariance matrix. In addition, we can not differentiate the covariate information in four quadrates of a subplot because the finest resolution for covariate in the MOFEP is a grid cell of 10 m by 10 m. Due to this special structure of the data, we propose a new spatial model by considering the measurements in four quadrats of a subplot as repeated measurements. See the related work by Hooten et al. (2003).

This chapter is organized as follows. Section 5.2 describes the new spatial model. The full conditional distributions and the simulation scheme will be discussed in Section 5.3. The prediction will be discussed in Section 5.4. We will apply our new model to the data of total vegetation coverage in site one in the MOFEP. Finally, some comments and possible generalization will be given.

5.2 Model Development

5.2.1 Hierarchical Model

Suppose that $Z_1(\mathbf{s}_1), \dots, Z_r(\mathbf{s}_1), \dots, Z_1(\mathbf{s}_n), \dots, Z_r(\mathbf{s}_n)$ are rn observations over the domain \mathcal{D} , where $Z_j(\mathbf{s}_i)$ is the j th observations at location s_i , $i = 1, \dots, n; j = 1, \dots, r$. Suppose that

$$Z_j(\mathbf{s}_i) = \mathbf{x}(\mathbf{s}_i)' \boldsymbol{\beta} + W(\mathbf{s}_i) + \varepsilon_j(\mathbf{s}_i), \quad (5.1)$$

where $\mathbf{x}(\mathbf{s}_i)$ is a $p \times 1$ vector, representing location-dependent covariates, and $(W(\mathbf{s}_1), \dots, W(\mathbf{s}_n))$ are realizations from a second-order stationary Gaussian process with

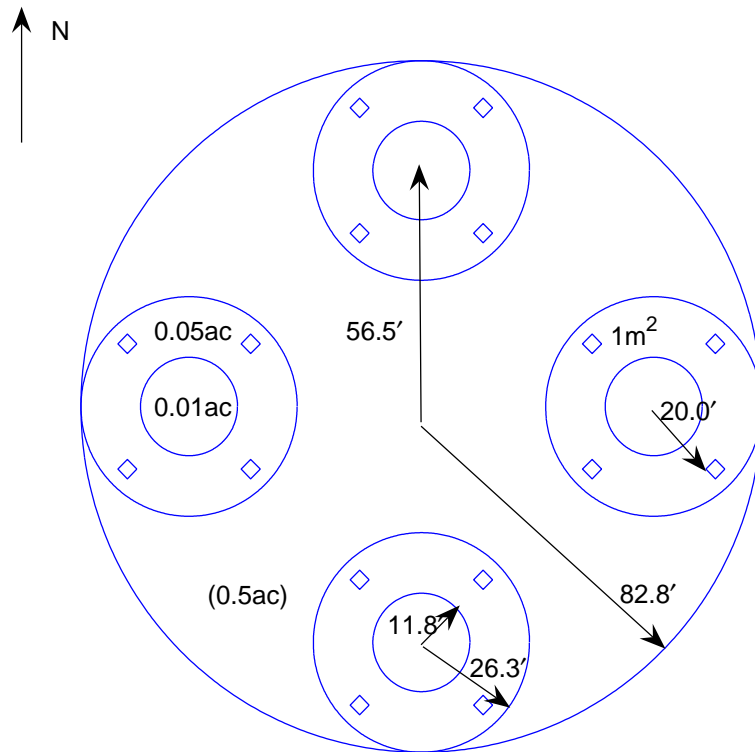


Figure 5.1: MOFEP Vegetation Plot Design.

covariance function $\sigma^2\mathbf{K}(\|\mathbf{s}_i - \mathbf{s}_j\|; \boldsymbol{\theta})$, that is,

$$(W(\mathbf{s}_1), \dots, W(\mathbf{s}_n))' \sim N_n(\mathbf{0}, \sigma^2\mathbf{K}(\boldsymbol{\theta})). \quad (5.2)$$

Here $\mathbf{K}(\boldsymbol{\theta}) = (K(\|\mathbf{s}_i - \mathbf{s}_j\|; \boldsymbol{\theta}))_{n \times n}$. For each location \mathbf{s}_i , we assume that

$$(\varepsilon_1(\mathbf{s}_i), \dots, \varepsilon_r(\mathbf{s}_i))' \sim N_r(\mathbf{0}, \sigma^2\tau^2\mathbf{H}(\rho)), \quad (5.3)$$

where $\mathbf{H}(\rho)$ is a specified matrix controlled by the parameter ρ . In addition, we assume the independence between $W(\mathbf{s}_i)$ and $\varepsilon_j(\mathbf{s}_i)$. In what follows, we work with the exponential family which is isotropic of the form $K(d; \theta) = \exp(-d/\theta)$ and we choose $H(\rho)$ as

$$\mathbf{H}(\rho) = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdots & \cdots & \cdots & \cdots \\ \rho & \rho & \cdots & 1 \end{pmatrix}_{r \times r}. \quad (5.4)$$

Note that θ is the spatial range parameter, which measures how quickly the correlations of the random field decay with distance, while ρ is the local correlation coefficient for repeated measurements. In addition, $\sigma^2\tau^2$ can be viewed as the variance of repeated measurement errors, which is often called the nugget effect when $r = 1$.

Remark 5.1 $H(\rho)$ in (5.4) is positive definite if and only if $\rho \in (-1/(r-1), 1)$.

The parameters of interest are $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ and we choose the functional form for the prior to be

$$\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho) = \pi(\boldsymbol{\beta})\pi(\sigma^2)\pi(\tau^2)\pi(\theta)\pi(\rho). \quad (5.5)$$

As for prior specifications, we assume

$$\boldsymbol{\beta} \sim N_p(\mathbf{0}, d\mathbf{I}_p), \quad (5.6)$$

$$\sigma^2 \sim \text{IG}(a_1, b_1), \quad (5.7)$$

$$\tau^2 \sim \text{IG}(a_2, b_2), \quad (5.8)$$

$$\theta \sim \text{IG}(a_3, b_3), \quad (5.9)$$

$$\rho \sim \text{U}(0, 1), \quad (5.10)$$

where a_i, b_i and d are specified. $\text{IG}(a, b)$ refers to an inverse gamma distribution with shape parameter a and scale parameter b , and $\text{U}(a, b)$ is a uniform distribution on the interval (a, b) . Based on my experience, it is reasonable to just assume ρ to be positive because it is the local correlation for repeated measurements that may be affected by local environmental conditions. However, it is often hard to justify this point from the observed data. For these specifications, the prior $\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ given by (5.5) is proper and therefore the corresponding posterior distribution should be proper as well.

Remark 5.2 *If $r = 1$, then the above model is exactly the same as the typical spatial model with nugget effect in Cressie (1993) or Banerjee et al. (2004). In this sense, the proposed model generalizes the traditional spatial model with nugget effect.*

5.2.2 The Likelihood

Note that though $(Z_1(\mathbf{s}_i), \dots, Z_r(\mathbf{s}_i))$ follows a multivariate normal distribution given the $W(\mathbf{s}_i)$, a Gibbs sampler which also updates the latent $W(\mathbf{s}_i)$ will be sampling

$n+p+r$ dimensional posterior density. However, it is possible to marginalize explicitly over the $W(\mathbf{s}_i)$ s and it is almost always preferable to implement iterative simulation with a lower dimensional distribution. In fact, such marginalization is routinely used in conjunction with normal data and normal random effects and, for example, is used by Chib and Carlin (1999) in the context of longitudinal data models. Let

$$\mathbf{Z} = (Z_1(\mathbf{s}_1), \dots, Z_r(\mathbf{s}_1), \dots, Z_1(\mathbf{s}_n), \dots, Z_r(\mathbf{s}_n))'. \quad (5.11)$$

Then we can prove that

$$(\mathbf{Z} \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho) \sim N_{rn}(\boldsymbol{\mu}, \sigma^2 \mathbf{G}), \quad (5.12)$$

where

$$\boldsymbol{\mu} = (\mathbf{X} \otimes \mathbf{1}_r) \boldsymbol{\beta}, \quad (5.13)$$

$$\mathbf{G} = \mathbf{K}(\theta) \otimes (\mathbf{1}_r \mathbf{1}_r') + \tau^2 \mathbf{I}_n \otimes \mathbf{H}(\rho). \quad (5.14)$$

Here “ $\mathbf{A} \otimes \mathbf{B}$ ” denotes the Kronecker product of \mathbf{A} and \mathbf{B} , and $\mathbf{1}_r$ is an $n \times 1$ vector with all elements one. The likelihood function is thus given by

$$L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{z}) = \frac{|\mathbf{G}|^{-1/2}}{(\sigma^2)^{nr/2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{z} - \boldsymbol{\mu})' \mathbf{G}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \right\}. \quad (5.15)$$

5.3 The Posterior Distribution and Sampling Schemes

5.3.1 The Full Conditional Distributions

The posterior density of $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ is

$$p(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho \mid \mathbf{z}) \propto L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{z}) \pi(\boldsymbol{\beta}) \pi(\sigma^2) \pi(\tau^2) \pi(\theta) \pi(\rho), \quad (5.16)$$

where $L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{z})$ is the likelihood function given by (5.15), and $\pi(\boldsymbol{\beta}), \pi(\sigma^2), \pi(\tau^2), \pi(\theta)$ and $\pi(\rho)$ are determined by (5.6)–(5.10), respectively. The above posterior distribution is proper because the prior is proper. However, it is very complicated. We will apply the Gibbs sampling for the posterior simulation and therefore the full conditional distributions are needed. The full conditional distribution of $\boldsymbol{\beta}$ can be readily obtained as follows:

$$(\boldsymbol{\beta} \mid \sigma^2, \tau^2, \theta, \rho; \mathbf{z}) \sim N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta), \quad (5.17)$$

where

$$\boldsymbol{\mu}_\beta = [(\mathbf{X} \otimes \mathbf{1}_r)' \mathbf{G}^{-1} (\mathbf{X} \otimes \mathbf{1}_r) + d^{-1} \mathbf{I}_p]^{-1} (\mathbf{X} \otimes \mathbf{1}_r)' \mathbf{G}^{-1} \mathbf{z}, \quad (5.18)$$

$$\boldsymbol{\Sigma}_\beta = \sigma^2 [(\mathbf{X} \otimes \mathbf{1}_r)' \mathbf{G}^{-1} (\mathbf{X} \otimes \mathbf{1}_r) + d^{-1} \mathbf{I}_p]^{-1}. \quad (5.19)$$

The full conditional distribution of σ^2 is still an inverse gamma distribution given by

$$(\sigma^2 \mid \boldsymbol{\beta}, \tau^2, \theta, \rho; \mathbf{z}) \sim \text{IG} \left(\frac{rn}{2} + a_1, \frac{1}{2} (\mathbf{z} - \boldsymbol{\mu})' \mathbf{G}^{-1} (\mathbf{z} - \boldsymbol{\mu}) + b_1 \right). \quad (5.20)$$

However, the full conditional distributions of τ^2 , θ or ρ are not standard forms and very complicated, which involves matrix determinant and matrix inverse functions. One possible choice is Metropolis-Hastings algorithm. But it is hard to find appropriate proposal densities. Slice sampling (Neal, 2003), by contrast, is advantageous in this case. We outline the shrinkage slice sampler in the following subsection.

5.3.2 Shrinkage Slice Sampler

Suppose we want to sample $x \sim f(x)$. The main idea of slice sampling is to introduce an auxiliary variable y such that the joint distribution of x and y is uniform

over the region $U = \{(x, y) : 0 < y < f(x)\}$. Then x and y are alternately sampled from uniform distributions in a Markov chain. Furthermore, suppose the target density can be written as $f(\theta) = f_1(\theta)f_2(\theta)$ where $f_2(\theta)$ is a standard density easy to sample and $f_1(\theta)$ is a complex function difficult to sample. For example, $f_2(\theta)$ is a proper prior $\pi(\theta)$ easy to sample while $f_1(\theta)$ is the likelihood function $L(\theta; \mathbf{y})$. We can introduce an auxiliary variable u such that the joint density of θ and u is $p(\theta, u) = I(u < f_1(\theta))f_2(\theta)$. So the marginal density of θ is $f(\theta)$. The following is the slice sampler. Given the current state $\theta^{(t)}$,

Step 1. Sample $u_t \sim U(0, f_1(\theta^{(t)}))$,

Step 2. Sample $\theta^{(t+1)} \sim f_2(\theta)$ until $u_t < f_1(\theta^{(t+1)})$.

As Neal (2003) stated, slicing only the likelihood $f_1(\theta) = L(\theta; \mathbf{y})$ and doing Gibbs updates using draws from the prior $f_2(\theta) = \pi(\theta)$ along with the rejection sampling is truly “off the shelf”, requiring no tuning at all. It is from this point that the slice sampling is proposed to be advantageous over the Metropolis-Hastings algorithm.

A naive rejection sampling scheme (repeatedly sample from $f_2(\theta)$ until we get a point in the slice, that is, $u < f_1(\theta)$) need not give good results. An algorithm which shrinks the support of $f_2(\theta)$ so that it gives better approximation to the slice whenever there is a rejection is more appropriate. The scheme is called “shrinkage slice sampler” by Neal (2003) and described as follows. If a point θ^* drawn from $f_2(\theta)$ is not in the slice and is larger (smaller) than the current value $\theta^{(t)}$ which is in the slice, the next draw is made from $f_2(\theta)$ truncated with the upper (lower) bound being θ^* . The truncated interval keeps shrinking with each rejection until a point in the

slice is drawn.

This is another issue worth mentioning about slice sampling in spatial statistics. With regard to computation, for large data sets often evaluation of the likelihood $L(\boldsymbol{\theta}; \mathbf{y})$ will produce an underflow, preventing sampling from the uniform distribution for u given $\boldsymbol{\theta}$ in Step 1. However, $\log L(\boldsymbol{\theta}; \mathbf{y})$ will typically not be a problem to compute. So it is often safer to compute $\log f_1(\boldsymbol{\theta})$ rather than $f_1(\boldsymbol{\theta})$ itself. One can then use the auxiliary variable $v = \log(u) = \log f_1(\boldsymbol{\theta}) - e_1$, where e_1 is exponentially distributed with mean one.

5.3.3 Simulation Algorithm

Combining Gibbs sampling with the shrinkage slice sampler, we propose the following simulation algorithm for the posterior (5.16).

Algorithm:

Step 1: For given $(\sigma^2, \tau^2, \theta, \rho)$, update $\boldsymbol{\beta}$ by simulating

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta),$$

where $\boldsymbol{\mu}_\beta$ and $\boldsymbol{\Sigma}_\beta$ are given by (5.18) and (5.19), respectively;

Step 2: For given $(\boldsymbol{\beta}, \tau^2, \theta, \rho)$, update σ^2 by simulating

$$\sigma^2 \sim IG\left(\frac{rn}{2} + a_1, \frac{1}{2}(\mathbf{z} - \boldsymbol{\mu})' \mathbf{G}^{-1}(\mathbf{z} - \boldsymbol{\mu}) + b_1\right),$$

where $\boldsymbol{\mu}$ and \mathbf{G} are given by (5.13) and (5.14), respectively;

Step 3: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$, simulate $R \sim \exp(1)$ and let

$$V = R - \log L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{z});$$

Step 4: For given $(\boldsymbol{\beta}, \sigma^2, \theta, \rho; V)$, update τ^2 by simulating $\tau^2 \sim IG(a_2, b_2)$ until

$$V > -\log L(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{z}); \quad (5.21)$$

Step 5: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \rho; V)$, update θ by simulating $\theta \sim IG(a_3, b_3)$ until

the condition (5.21) is satisfied;

Step 6: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta; V)$, simulate $\rho \sim U(0, 1)$ until the condition

(5.21) is satisfied;

Step 7: Go back to Step 1 until we get the appropriate number of MCMC samples.

Notice that in Steps 4, 5 and 6, to speed the updating processes, we always apply the technique of the shrinkage slice sampler in § 5.3.2.

5.4 Prediction

The Bayesian approach to prediction provides a general methodology for taking into account the uncertainty about parameters on subsequent predictions. For an unmeasured location \mathbf{s}_0 , $Z(\mathbf{s}_0) \hat{=} Z_0$ can be predicted by the following predictive distribution

$$p(z_0 | \mathbf{z}) = \int p(z_0 | \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z}) d\boldsymbol{\beta} d\sigma^2 d\theta, \quad (5.22)$$

where $p(\boldsymbol{\beta}, \sigma^2, \theta | \mathbf{z})$ is the posterior distribution described by (5.16) and $p(z_0 | \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta)$ is determined by

$$p(\mathbf{z}, z_0 | \boldsymbol{\beta}, \sigma^2, \theta) \sim N_{n+1} \left(\begin{pmatrix} \boldsymbol{\mu} \\ \mathbf{x}'(\mathbf{s}_0)\boldsymbol{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} \mathbf{G} & \mathbf{k}_\theta \\ \mathbf{k}'_\theta & 1 \end{pmatrix} \right). \quad (5.23)$$

Here $\mathbf{x}(\mathbf{s}_0)$ is the covariate vector evaluated at \mathbf{s}_0 and \mathbf{k}_θ is given by

$$\mathbf{k}_\theta = (K(\|\mathbf{s}_0 - \mathbf{s}_1\|; \theta), \dots, K(\|\mathbf{s}_0 - \mathbf{s}_n\|; \theta))' \otimes \mathbf{1}_r.$$

Thus, we have

$$p(z_0 \mid \mathbf{z}, \boldsymbol{\beta}, \sigma^2, \theta) \sim N(\mu_0, \sigma_0^2), \quad (5.24)$$

where

$$\mu_0 = \mathbf{x}'(\mathbf{s}_0)\boldsymbol{\beta} + \mathbf{k}_\theta' \mathbf{G}^{-1}(\mathbf{z} - \boldsymbol{\mu}), \quad \sigma_0^2 = \sigma^2(1 - \mathbf{k}_\theta' \mathbf{G}^{-1} \mathbf{k}_\theta). \quad (5.25)$$

As for the simulation, we just need to add

*Step 6**: for given $\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho$, simulate z_0 from $N(\mu_0, \sigma_0^2)$ with μ_0, σ_0^2 given by (5.25);

before *Step 7* of the algorithm in § 5.3.3 for each location \mathbf{s}_0 and thus we get the samples of z_0 from the posterior predictive distribution $p(z_0 \mid \mathbf{z})$ in (5.22). The inference of $z(\mathbf{s}_0)$ at unmeasured location \mathbf{s}_0 is based the samples obtained.

5.5 Application to the Herbaceous Data Analysis

In this section, we analyze the herbaceous data collected in 1994 in site one of the MOFEP. For each quadrat, vegetation was identified and quantified on a percent coverage basis (Grabner, 2000). Let \mathcal{D} represent the area of site one and $C_j(\mathbf{s}_i)$ be the observed percent coverage of total vegetation in the i th quadrat of a subplot at

center \mathbf{s}_i . Let

$$Z_j(\mathbf{s}_i) = \log \frac{C_j(\mathbf{s}_i) + .0005}{1 - (C_j(\mathbf{s}_i) + .0005)}. \quad (5.26)$$

The transformation (5.26) without adding .0005 is popularly used in logistic regression because it makes the transformation from $(0, 1)$ to $(-\infty, \infty)$. As Webster and McBratney (1987) and Cressie (2004) did, we add 0.05% to each observed datum $C_j(\mathbf{s}_i)$ to avoid taking logarithm of zero. We fit the model described in § 5.2 for $\{Z_j(\mathbf{s}_i)\}_s$.

Since the total vegetation data were measured in 73 plots of site one, so we have $n = 292$ and $r = 4$. See Figure 5.2 for locations of the subplots.

Covariates were chosen based on ecological background and availability. We chose two covariates: the aspect class and the soil depth because of their important influence on composition (Grabner, 2002, page 137). The full description for the aspect class and the soil depth can be found in § 4.2.1.

The hyperparameters are chosen so that the priors for parameters $\boldsymbol{\beta}$, σ^2 , τ^2 and θ are proper but vague enough and thus the inference is driven by the data. For $\boldsymbol{\beta}$, we choose $d = 100$, implying that each β_i has a large variance of its prior distribution. Centering the normal priors for the regression coefficients at zero is reasonable because all the coefficients are expected to have small magnitudes due to the small values of the response variable (minimum -7.02 to maximum 1.33). Note that for an inverse gamma distribution with shape a and scale b , it has a mean b and an infinite variance when the shape parameter a is 2. The means of the priors of σ^2 , τ^2 and θ are chosen to be close to their REML estimates while the prior variances are infinite. So we

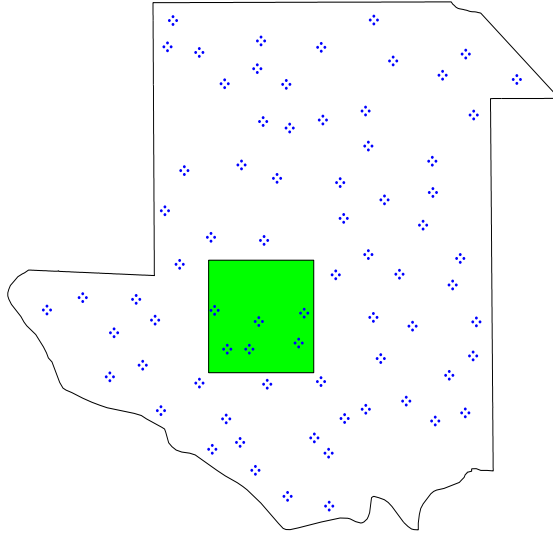


Figure 5.2: Locations of subplots in site one (the square area is for prediction)

choose $a_1 = a_2 = a_3 = 2$ and $b_1 = .48$, $b_2 = 2.50$, $b_3 = .25$.

The MCMC algorithm in § 5.3.3 was run for 10,000 iterations after a burn-in period of 1,000 iterations. It took about 230 hours on a 3.06GHz Intel Xeon server with MATLAB software. Resulting parameter distributions were summarized in the form of histograms. Figure 5.3 shows the marginal histograms for the parameters $\boldsymbol{\beta}$, in the model. The marginal posterior distributions for β_i are approximately symmetric. It can be inferred that soil depth is indeed an important factor influencing the total vegetation coverage because 95% Bayesian credible interval of β_2 is $(0.0526, 0.6951)$, which does not contain zero. The 95% Bayesian credible interval for β_1 is $(-0.1571, 0.4033)$.

This suggests that the aspect class is perhaps less important than the soil depth when modelling the total vegetation coverage. However, we will still make use of the aspect class for prediction later because $P(\beta_1 > 0 \mid \mathbf{z})$ is more than 0.85.

parameter	median	mean	standard deviation	95% Bayesian CI
β_0	-2.5066	-2.5065	0.2054	[-2.9109, -2.1095]
β_1	0.1267	0.1259	0.1438	[-0.1571, 0.4033]
β_2	0.3786	0.3789	0.1640	[0.0526, 0.6951]
σ^2	0.4464	0.4618	0.1213	[0.2556, 0.7198]
τ^2	2.9139	3.0465	0.9333	[1.6938, 5.4329]
θ	0.1398	0.1575	0.0776	[0.0673, 0.3498]
ρ	0.1508	0.1496	0.0497	[0.0450, 0.2435]

Table 5.1: Posterior Quantities of $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$.

Figure 5.4 also shows the marginal histograms for the parameters $\sigma^2, \tau^2, \theta^2, \rho$. Unlike β_i , the marginal posterior distributions for $\sigma^2, \tau^2, \theta^2, \rho$ are skewed. Table 5.1 shows the posterior quantities for parameters in the model.

Remember that the parameter θ gives the information of spatial effect while ρ provides the local correlation information for the measurements from four quadrats in a subplot. As suggested by Banerjee et al. (2004), we may choose posterior medians as their point estimates. So we have $\hat{\theta} = 0.1398$ and $\hat{\rho} = 0.1508$. Hence we get that the effective range, which is the distance at which the spatial correlation drops to

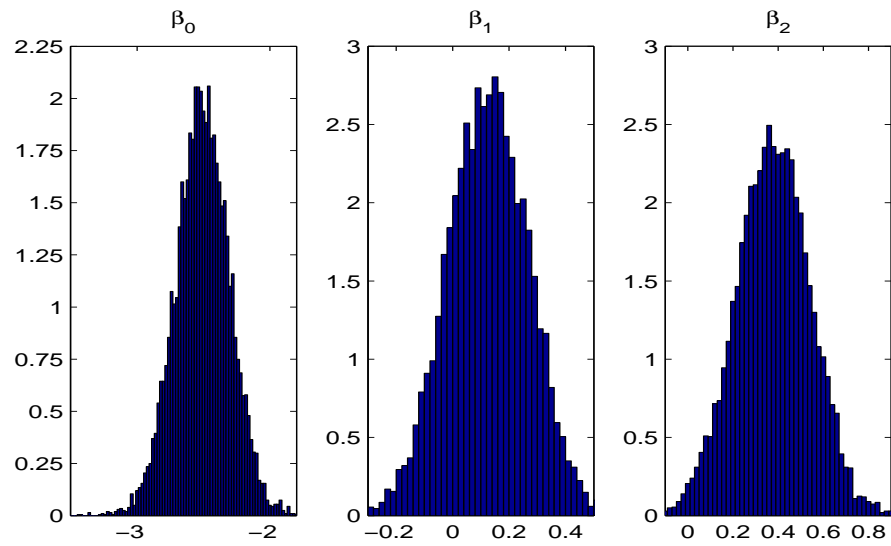


Figure 5.3: Histograms of β .

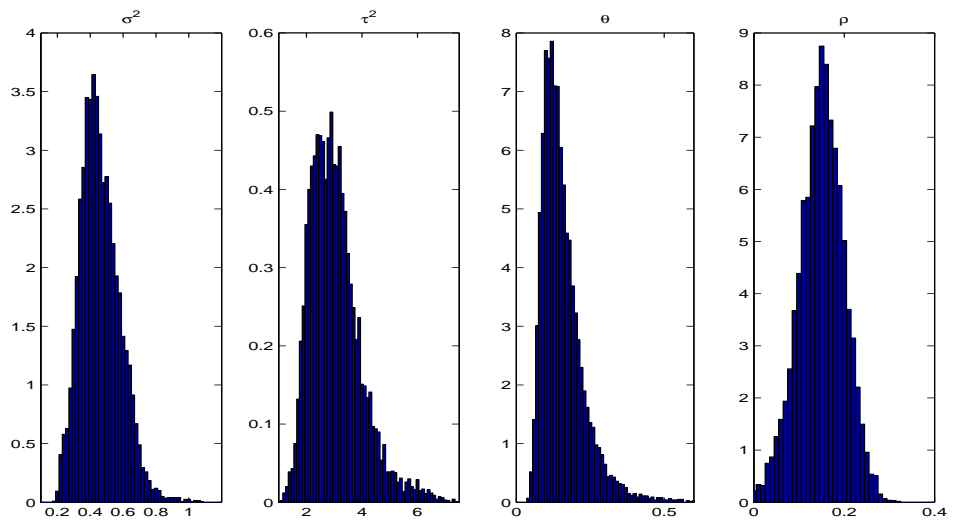


Figure 5.4: Histograms of $(\sigma^2, \tau^2, \theta, \rho)$.

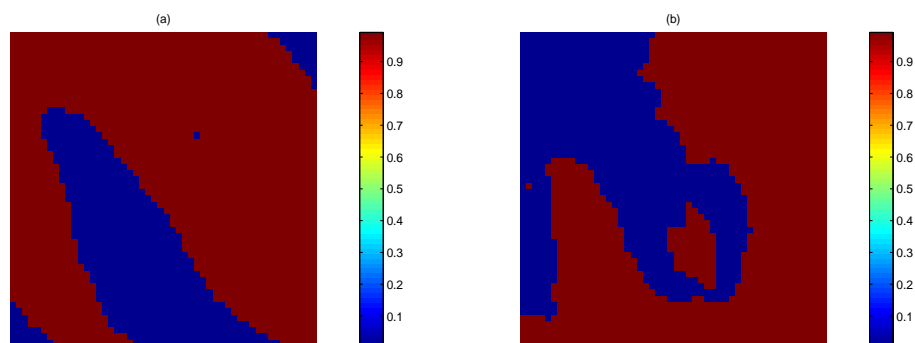


Figure 5.5: The Covariate Information in the Prediction Domain: (a) Aspect Class (1 = Protected, 0 = Exposed), (b) Soil Depth (1 = Deep to Very Deep, 0 = Shallow to Moderate Deep).

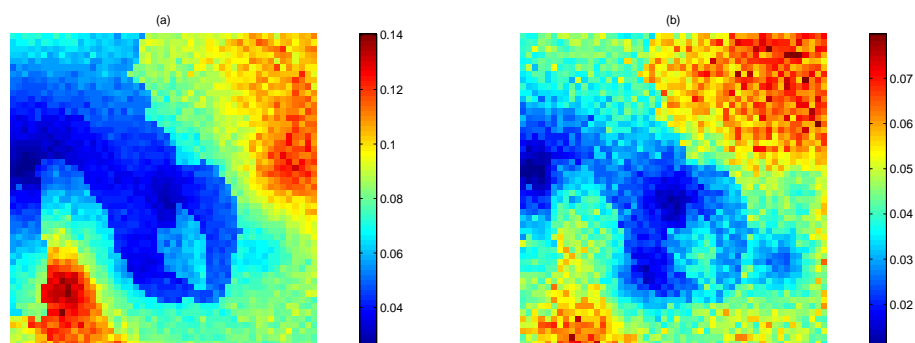


Figure 5.6: The Maps of Prediction and Prediction Error: (a) Prediction, (b) Standard Deviation of the Prediction.

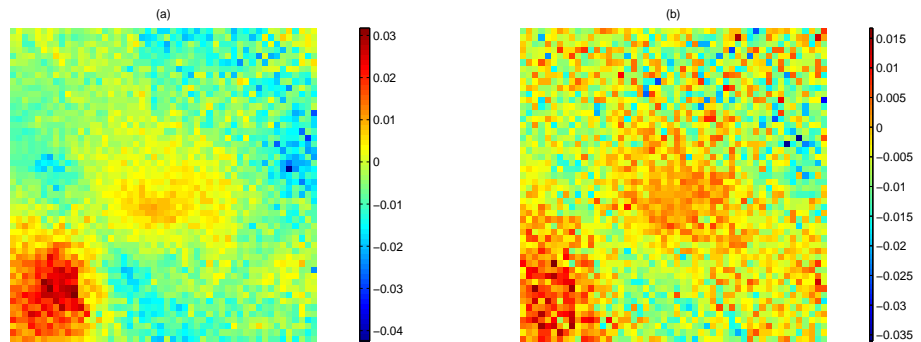


Figure 5.7: The Maps of Prediction Difference between Four Quadrats and Two Quadrats: (a) Prediction, (b) Standard Deviation of the Prediction

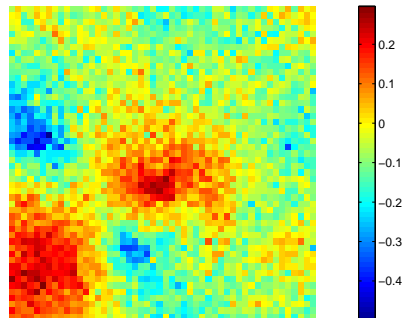


Figure 5.8: The Relative Difference of the Predictions between Four Quadrats and Two Quadrats.

only 0.05, is approximately 0.42 km. Notice that the maximum distance in site one is approximately 2.50 km and thus the above point estimate of the effective range implies that the strong spatial effect exists for the total vegetation coverage data in site one. In addition, the estimate of ρ shows that the measurements from four quadrats in a subplot are not strongly dependent but are not independent.

Remark 5.3 *As Dr. Dongchu Sun, one of my committee members, pointed out that for a prior of the parameter η , it may be vague enough for η itself while it is quite informative for $g(\eta)$, where $g(\eta)$ is a function of η . This is true. What we can do is to take care of parameters of interest first. Note that $\sigma^2\tau^2$ is one of interest parameters in the model because it can be viewed as the variance of repeated measurement errors. Although the prior of $\sigma^2\tau^2$ is not a standard form, the prior variance of $\sigma^2\tau^2$ is still infinite when σ^2 and τ^2 are assigned inverse Gamma distributions with shape parameters $a_i = 2, i = 1, 2$. In this case, hyperparameters b_1, b_2 would not affect the inference too much although we chose b_1, b_2 to be close to REMLs of σ^2 and τ^2 , respectively. The small study in the following supports this point. We generated 2000 samples from the posterior distribution with $d = 100, a_1 = a_2 = a_3 = 2$ and $b_1 = 0.30, b_2 = 1.00, b_3 = 0.2$. Figure 5.9 presents the comparison of the estimated posterior densities under two different cases. We found that the estimated posterior densities for each parameter are very close.*

Based on the theory in § 5.4, the total vegetation coverage at any location in site one may be predicted. We choose a square area of 500 meters by 500 meters in the southeastern part of site one for prediction. Like § 4.5, we made a grid cell size of

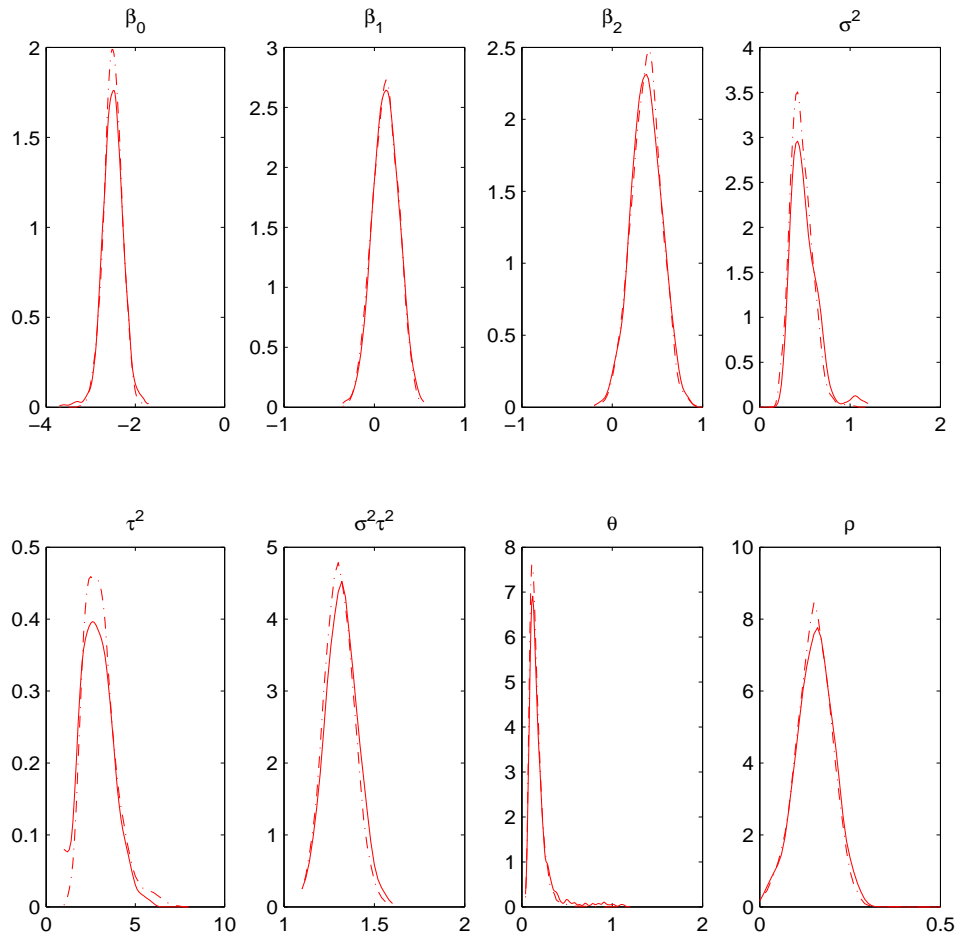


Figure 5.9: The Comparison of Posterior Densities: the Solid Line with the Case of $d = 100$, $a_1 = a_2 = a_3 = 2$ and $b_1 = 0.30, b_2 = 1.00, b_3 = 0.2$ and the Dashed Line with the Case of $d = 100$, $a_1 = a_2 = a_3 = 2$ and $b_1 = 0.48, b_2 = 2.50, b_3 = 0.25$.

10 meters by 10 meters, resulting in 2500 prediction locations in this domain. The covariate information in this prediction domain can be seen in Figure 5.5.

We choose the median of samples from the predictive distribution as a point estimate of $Z(\mathbf{s}_0)$ and thus get an estimate of the total vegetation coverage $C(\mathbf{s}_0)$ at location \mathbf{s}_0 from the transformation (5.26).

Since there are so many locations to be predicted, computation is very time-consuming due to inverse and determinant of large dimensional matrices if we make use of all 10,000 samples from the posterior distribution. We consider reducing the number of posterior samples. One way is to take just a few posterior samples, say, the first 500. However, as mentioned by Banerjee et al. (2004), the outputs created from the shrinkage slice sampler are usually highly autocorrelated, which may not give a good result for the predictive distribution when using a small size of posterior samples. The better way is to resort to *thinning*, which is simply retaining only every k th sampled value, where k is the approximate lag at which the autocorrelations in the chain become insignificant. For our 10000 posterior samples, we found that $k = 40$ is an appropriate number and thus 250 posterior sub-samples were obtained. A small study revealed that the difference of the sample medians from the predictive distribution incurred by using the 10000 total samples and using the 250 sub-samples is ignorable. For example, at location $\mathbf{s}_0 = (665000, 4115000)$, the sample median of the predictive distribution is about -2.8164 with the 10000 total posterior samples while it is about -2.8013 with just 250 “thinning” posterior sub-samples.

Figure 5.6(a) displays the prediction map of the total vegetation coverage in the

250,000m² square domain based on each median of 250 “thinning” posterior subsamples. The corresponding map of standard deviation can be seen in Figure 5.6(b). It took about 136 hours on a 2.80GHz Intel Pentium IV PC for the prediction simulation.

It is interesting to see whether we can reduce the number of quadrats in each subplot when the prediction of the total vegetation coverage is the primary purpose. So among four quadrats in each subplot, two of them, which are at 45° and 225°, are chosen and then we remodel the new data set again ($n = 292, r = 2$). The corresponding maps of the total vegetation coverage in the 250,000m² square domain can be seen in Figure 5.7. It looks like there is no meaningful difference between Figure 5.6(a) and Figure 5.7(a). We can also see the map of the relative difference based on the results of four quadrats and two quadrats in each subplot in Figure 5.8. If biologists agree that the difference is acceptable, we may just measure the total vegetation coverage in two quadrats of each subplot.

5.6 Discussion and Conclusion

We have developed a new spatial model taking into account several close measurements as repeated measurements in one location and applied to the analysis of the total vegetation coverage data in site one of the MOFEP. Our results show that the soil depth covariate is an important factor while the aspect class is less important when modelling the total vegetation coverage. We also show that the strong spatial effect does exist in the data discussed and the measurements in four quadrats of a

subplot are not strongly correlated but are not independent. In addition, prediction of the total vegetation coverage at unmeasured locations is established.

Model validation is very important when setting up a new model. This is worthy of further study in the future.

Possible extensions could be made as follows.

In the MOFEP, it is also important to model the vegetation coverage by an individual species. For modelling the coverage by an individual species, the model in § 5.2.1 is not applicable directly because plants occur infrequently and have low coverage. Consequently, the data contain a lot of zeros and 0.001s. For example, *Desmodium nudiflorum* is the most popular species in nine MOFEP sites. Despite being the most abundant species, there are about 41.6 percent quadrats where *Desmodium nudiflorum* is absent and about 8.7 percent quadrates where the coverage of *Desmodium nudiflorum* is less one percent (counted as 0.1 percent). However, we may make use of our model by introducing latent variables.

Let $Y_j(\mathbf{s}_i)$ be the observed percent coverage of an individual species and $Y_j^*(\mathbf{s}_i)$ be the unobserved true percent coverage in the i th quadrat of a subplot at center \mathbf{s}_i . It is reasonable to assume that

$$Y_j(\mathbf{s}_i) = \begin{cases} 0, & \text{if } 0 \leq Y_j^*(\mathbf{s}_i) < 0.001, \\ 0.001, & \text{if } 0.001 \leq Y_j^*(\mathbf{s}_i) < 0.01, \\ Y_j^*(\mathbf{s}_i), & \text{if } 0.01 \leq Y_j^*(\mathbf{s}_i) \leq 1. \end{cases} \quad (5.27)$$

Let $U_j(\mathbf{s}_i) = \Phi^{-1}(Y_j^*(\mathbf{s}_i))$ or $Y_j^*(\mathbf{s}_i) = \Phi(U_j(\mathbf{s}_i))$, where $\Phi(\cdot)$ is the cumulative probability function of the standard normal distribution. It is straightforward to see

that

$$Y_j(\mathbf{s}_i) = \begin{cases} 0, & \text{if } U_j(\mathbf{s}_i) < c_1, \\ 0.001, & \text{if } c_1 \leq U_j(\mathbf{s}_i) < c_2, \\ \Phi^{-1}(U_j(\mathbf{s}_i)), & \text{if } U_j(\mathbf{s}_i) \geq c_2. \end{cases} \quad (5.28)$$

where

$$c_1 = \Phi^{-1}(0.001) = -3.0902, \quad c_2 = \Phi^{-1}(0.01) = -2.3263. \quad (5.29)$$

Like $Z_j(\mathbf{s}_i)$, it is assumed that $\{U_j(\mathbf{s}_i)\}$ follow the hierarchical model in § 5.2.1. In what follows, we assume $r = 4$. Let

$$\mathbf{U} = (U_1(\mathbf{s}_1), \dots, U_r(\mathbf{s}_1), \dots, U_1(\mathbf{s}_n), \dots, U_r(\mathbf{s}_n))'. \quad (5.30)$$

Therefore,

$$(\mathbf{U} \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho) \sim N_{rn}(\boldsymbol{\mu}, \sigma^2 \mathbf{G}), \quad (5.31)$$

which is the same as the conditional distribution of \mathbf{Z} given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ in (5.12) of subsection 5.2.2, where $\boldsymbol{\mu}$ and \mathbf{G} are given by (5.13) and (5.14), respectively. So, for given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$, the conditional density of U is

$$p(\mathbf{u} \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho) = \frac{|\mathbf{G}|^{-1/2}}{(\sigma^2)^{nr/2}} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{u} - \boldsymbol{\mu})' \mathbf{G}^{-1} (\mathbf{u} - \boldsymbol{\mu}) \right\}. \quad (5.32)$$

Let

$$\mathbf{Y} = (Y_1(\mathbf{s}_1), \dots, Y_r(\mathbf{s}_1), \dots, Y_1(\mathbf{s}_n), \dots, Y_r(\mathbf{s}_n))'. \quad (5.33)$$

From (5.28), we have that for given U , the conditional density of Y is

$$p(\mathbf{y} \mid \mathbf{u}) = \prod_{i=1}^n \prod_{j=1}^r [I(y_j(\mathbf{s}_i) = 0, u_j(\mathbf{s}_i) < c_1)$$

$$\begin{aligned}
& + I(y_j(\mathbf{s}_i) = 0.001, c_1 \leq u_j(\mathbf{s}_i) < c_2) \\
& + I(y_j(\mathbf{s}_i) = \Phi^{-1}(u_j(\mathbf{s}_i)), u_j(\mathbf{s}_i) \geq c_2)], \tag{5.34}
\end{aligned}$$

where $I(A)$ is the indicator function of A . As for the prior of $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$, we may still choose the prior $\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ defined by (5.5)-(5.10). Combining (5.34) and (5.32) with the prior $\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$, we have the joint distribution of $(\mathbf{y}, \mathbf{u}, \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$

$$p(\mathbf{y}, \mathbf{u}, \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho) = p(\mathbf{y} | \mathbf{u})p(\mathbf{u} | \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)\pi(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho). \tag{5.35}$$

For the posterior simulation, we still prefer to the Gibbs sampler. (5.35) tells us that for given \mathbf{U} , updating $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ requires no information from \mathbf{y} and thus the algorithm in subsection 5.3.3 can be used.

However, for each MCMC cycle, we need update $\{U_j(\mathbf{s}_i)\}$ because they are latent variables. The full conditional distribution of \mathbf{U} is given by

$$p(\mathbf{u} | \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{y}) \propto p(\mathbf{y} | \mathbf{u})p(\mathbf{u} | \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho). \tag{5.36}$$

From (5.31)(5.34) and (5.36), we can see that the full conditional distribution of \mathbf{U} follows a truncated multivariate normal distribution and thus the idea of the simulation algorithm for in Griffiths (2002) may be borrowed. We propose the following algorithm for the posterior simulation.

Algorithm:

Step 1: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho; \mathbf{y})$, update \mathbf{U} by using the simulation

algorithm in Griffiths (2002) for a truncated multivariate normal distribution;

Step 2: For given $(\sigma^2, \tau^2, \theta, \rho, \mathbf{u})$, update $\boldsymbol{\beta}$ by simulating

$$\boldsymbol{\beta} \sim N_p(\boldsymbol{\mu}_{\boldsymbol{\beta}}^*, \boldsymbol{\Sigma}_{\boldsymbol{\beta}}^*),$$

where

$$\boldsymbol{\mu}_{\boldsymbol{\beta}}^* = [(\mathbf{X} \otimes \mathbf{1}_4)' \mathbf{G}^{-1} (\mathbf{X} \otimes \mathbf{1}_4) + d^{-1} \mathbf{I}_p]^{-1} (\mathbf{X} \otimes \mathbf{1}_4)' \mathbf{G}^{-1} \mathbf{u}, \quad (5.37)$$

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}}^* = \sigma^2 [(\mathbf{X} \otimes \mathbf{1}_4)' \mathbf{G}^{-1} (\mathbf{X} \otimes \mathbf{1}_4) + d^{-1} \mathbf{I}_p]^{-1}; \quad (5.38)$$

Step 3: For given $(\boldsymbol{\beta}, \tau^2, \theta, \rho, \mathbf{u})$, update σ^2 by simulating

$$\sigma^2 \sim IG\left(\frac{rn}{2} + a_1, \frac{1}{2}(\mathbf{u} - \boldsymbol{\mu})' \mathbf{G}^{-1} (\mathbf{u} - \boldsymbol{\mu}) + b_1\right),$$

where $\boldsymbol{\mu}$ and \mathbf{G} are given by (5.13) and (5.14), respectively;

Step 4: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho, \mathbf{u})$, simulate $R \sim \exp(1)$ and let

$$V = R - \log p(\mathbf{u} \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho), \quad (5.39)$$

where $p(\mathbf{u} \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho)$ is given by (5.32);

Step 5: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho, \mathbf{u}, V)$, update τ^2 by simulating $\tau^2 \sim IG(a_2, b_2)$

until

$$V > -\log p(\mathbf{u} \mid \boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \rho); \quad (5.40)$$

Step 6: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \rho, \mathbf{u}, V)$, update θ by simulating $\theta \sim IG(a_3, b_3)$ until

the condition (5.40) is satisfied;

Step 7: For given $(\boldsymbol{\beta}, \sigma^2, \tau^2, \theta, \mathbf{u}, V)$, simulate $\rho \sim U(0, 1)$ until the condition

(5.40) is satisfied;

Step 8: Go back to Step 1 until we get the appropriate number of MCMC samples.

Details including the prediction of the percent coverage of an individual species at unmeasured locations will be investigated in the future.

Bibliography

- ABRAMOWITZ, M. and STEGUN, I. (1965). *Handbook of Mathematical Functions*. New York.
- AGARWAL, D. K. and GELFAND, A. E. (2005). Slice sampling for simulation based fitting of spatial data models. *Statistics and Computing* **15**, 61–69.
- ANDERSON, T. W. (1957). Maximum likelihood estimates for a multivariate normal distribution when some observations are missing. *Journal of the American Statistical Association* **52**, 200–203.
- ANDERSSON, S. A. and PERLMAN, M. D. (1993). Lattice models for conditional independence in a multivariate normal distribution. *The Annals of Statistics* **21**, 1318–1358.
- BANERJEE, S., CARLIN, B. P. and GELFAND, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Chapman & Hall/CRC.
- BERGER, J. O. and BERNARDO, J. M. (1992). On the development of reference priors. In *Proceedings of the Fourth Valencia International Meeting, Bayesian Statistics*, vol. 4.

- BERGER, J. O., DE OLIVEIRA, V. and SANS, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association* **96**, 1361–1374.
- BERGER, J. O., LISEO, B. and WOLPERT, R. L. (1999). Integrated likelihood methods for eliminating nuisance parameters (Pkg: p1-28). *Statistical Science* **14**, 1–22.
- BERGER, J. O., PERICCHI, L. R. and VARSHAVSKY, J. A. (1998). Bayes factors and marginal distributions in invariant situations. *Sankhya, Series A, Indian Journal of Statistics* **60**, 307–321.
- BESAG, J. and GREEN, P. J. (1993). Spatial statistics and Bayesian computation (Disc: p53-102). *Journal of the Royal Statistical Society, Series B, Methodological* **55**, 25–37.
- BONDAR, J. V. and MILNES, P. (1981). Amenability: A survey for statistical applications of Hunt-Stein and related conditions on groups. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete [Became: @J(ProbTher)]* **57**, 103–128.
- BOX, G. E. P. (1980). Sampling and Bayes' inference in scientific modelling and robustness (C/R: p404-430). *Journal of the Royal Statistical Society, Series A, General* **143**, 383–430.
- BROOKSHIRE, B. and SHIFLEY, S. (1997). *Proceedings of the Missouri Ozark Forest Ecosystem Project Symposium: An Experimental Approach to Landscape Research*

- (eds.). General Technical Report NC-193, St. Paul, MN: The United States Department of Agriculture, Forest Service, North Central Experiment Station.
- CARMEAN, W. H., HAHN, J. T. and JACOBS, R. D. (1989). *Site index curves for forest tree species in the eastern United States*. USDA Forest Service, North Central Forest Experiment Station, General Technical Report NC-128.
- CHIB, S. and CARLIN, B. P. (1999). On MCMC sampling in hierarchical longitudinal models. *Statistics and Computing* **9**, 17–26.
- CHILES, J.-P. and DELFINER, P. (1999). *Geostatistics: modeling spatial uncertainty*. John Wiley & Sons, New York.
- COWLES, M. K. (2003). Efficient model-fitting and model-comparison for high-dimensional Bayesian geostatistical models. *Journal of Statistical Planning and Inference* **112**, 221–239.
- CRAIGMILE, P. F., CRESSIE, N., SANTNER, T. J. and RAO, Y. (2004). Bayesian inferences on environmental exceedances and their spatial locations. Tech. Rep. 729, Department of Statistics, The Ohio State University.
- CRESSIE, N. (2004). Block kriging for lognormal spatial processes. Tech. Rep. 739, Department of Statistics, The Ohio State University.
- CRESSIE, N., ZHANG, J. and CRAIGMILE, P. F. (2004). Geostatistical prediction of spatial extremes and their extent. Tech. Rep. 738, Department of Statistics, The Ohio State University.

- CRESSIE, N. A. C. (1993). *Statistics for spatial data (Revised edition)*. Wiley-Interscience.
- DE OLIVEIRA, V., KEDEM, B. and SHORT, D. A. (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association* **92**, 1422–1433.
- DELLAPORTAS, P. and ROBERTS, G. O. (2003). An introduction to MCMC. In *Spatial statistics and computational methods (Aalborg, 2001)*, vol. 173 of *Lecture Notes in Statist.* New York: Springer, pp. 1–41.
- DEMPSTER, A. M. (1972). Covariance selection. *Biometrics* **28**, 157–175.
- DEY, D. K., CHEN, M.-H. and CHANG, H. (1997). Bayesian approach for nonlinear random effects models. *Biometrics* **53**, 1239–1252.
- DOMONICI, F., PARMIGIANI, G. and CLYDE, M. (2000). Conjugate analysis of multivariate normal data incomplete observations. *The Canadian Journal of Statistics* **28**, 533–550.
- EATON, M. L. and OLKIN, I. (1987). Best equivariant estimators of a Cholesky decomposition. *The Annals of Statistics* **15**, 1639–1650.
- ECKER, M. D. and GELFAND, A. E. (1997). Bayesian variogram modeling for an isotropic spatial process. *Journal of Agricultural, Biological, and Environmental Statistics* **2**, 347–369.
- GEISSER, S. (1993). *Predictive inference. An introduction*. Chapman & Hall Ltd.

- GELFAND, A. E., DEY, D. K. and CHANG, H. (1992). Model determination using predictive distributions, with implementation via sampling-based methods (Disc: p160-167). In *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*.
- GOTWAY, C. A. and YOUNG, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association* **97**, 632–648.
- GRABNER, J. (2000). Ground layer vegetation in Missouri Ozark Forest Ecosystem Project: pre-treatment species composition, richness, and diversity. In *Missouri Ozark Forest Ecosystem Project: Site history, Soils, Landform, Woody and Herbaceous Vegetation, Down Wood, and Inventory Methods for Landscape Experiment*, S. Shifley and B. Brookshire, eds., vol. NC-208.
- GRABNER, J. (2002). *Patterns in upland forest vegetation in relation to geology, topography, and soils: an approach to ecological land classification in the Southeast Missouri Ozarks*. Master's thesis, University of Missouri-Columbia.
- GRIFFITHS, W. (2002). A Gibbs' sampler for the parameters of a truncated multivariate normal distribution. Working Paper 856, Department of Economics, University of Melbourne.
- GUPTA, A. K. and NAGAR, D. K. (2000). *Matrix Variate Distributions*. Chapman & Hall Ltd., New York.
- GUPTA, A. K. and OFORI-NYARKO, S. (1995). Improved minimax estimators of

- normal covariance and precision matrices. *Statistics [Formerly: @J(MaOpfStS)]* **26**, 19–25.
- HAFF, L. R. (1980). Empirical Bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics* **8**, 586–597.
- HANDCOCK, M. S. and STEIN, M. L. (1993). A Bayesian analysis of kriging. *Technometrics* **35**, 403–410.
- HOOTEN, M. B., LARSEN, D. R. and WIKLE, C. K. (2003). Predicting the spatial distribution of ground flora on large domains using a hierarchical Bayesian model. *Landscape Ecology* **18**, 487–502.
- HOST, G. E. and PREGITZER, K. S. (1991). Ecological species groups for upland forest ecosystems of northwestern lower Michigan. *Forest Ecology and Management* **43**, 87–102.
- KABRICK, J., MEINERT, D., NIGH, T. and GORLINSKY, B. J. (2000). Physical environment of the Missouri Ozark Forest Ecosystem Project Sites. In *Missouri Ozark Forest Ecosystem Project Site History, Soils, Landforms, Woodys and Herbaceous Vegetation, Down Wood, and Inventory Methods for Landscape Experiemnt*, vol. NC-208.
- KIEFER, J. (1957). Invariance, minimax sequential estimation, and continuous time process. *The Annals of Mathematical Statistics* **28**, 573–601.
- KINDERMAN, A. and MONAHAN, J. F. (1977). Computer generation of random

- variables using the ratio of uniforms deviates. *ACM Transactions on Mathematical Software* **3**, 257–260.
- KITANIDIS, P. K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research* **22**, 499–507.
- KONNO, Y. (1995). Estimation of a normal covariance matrix with incomplete data under Stein’s loss. *Journal of Multivariate Analysis* **52**, 308–324.
- KONNO, Y. (2001). Inadmissibility of the maximum likelihood estimator of normal covariance matrices with the lattice conditional independence. *Journal of Multivariate Analysis* **79**, 33–51.
- KRISHNAMOORTHY, K. and GUPTA, A. K. (1989). Improved minimax estimation of a normal precision matrix. *The Canadian Journal of Statistics* **17**, 91–102.
- KUBOKAWA, T. and KONNO, Y. (1990). Estimating the covariance matrix and the generalized variance under a symmetric loss. *Annals of the Institute of Statistical Mathematics* **42**, 331–343.
- LAHIRI, S. N. (1999). Asymptotic distribution of the empirical spatial cumulative distribution function predictor and prediction bands based on a subsampling method. *Probab. Theory and Related Fields* **114**, 55–84.
- LAHIRI, S. N., KAISER, M. S., CRESSIE, N. and HSU, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling (C/R: p97-110). *Journal of the American Statistical Association* **94**, 86–97.

- LAURITZEN, S. L. (1996). *Graphical models*. Oxford University Press.
- LE, N. D. and ZIDEK, J. V. (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging. *Journal of Multivariate Analysis* **43**, 351–374.
- LISEO, B. (1993). Elimination of nuisance parameters with reference priors. *Biometrika* **80**, 295–304.
- LISEO, B. (2004). The elimination of nuisance parameters. In *Dey D. and Rao, C. R. eds.*, vol. 25 of *Handbook of Statistics*. New York: Elsevier Sciences, pp. 1–34.
- LITTLE, R. J. A. and RUBIN, D. B. (1987). *Statistical analysis with missing data*. John Wiley & Sons, New York.
- LIU, C. (1999). Efficient ML estimation of the multivariate normal distribution from incomplete data. *Journal of Multivariate Analysis* **69**, 206–217.
- MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–146.
- MCQUILKIN, R. A. (1974). *Site index prediction tables for black, scarlet and white oaks in southeastern Missouri*. Research Paper NC-108. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station.
- MCQUILKIN, R. A. (1978). *How to estimate site index for oaks in the Missouri Ozarks*. St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Forest Experiment Station.

- MØLLER, J., ed. (2003). *Spatial statistics and computational methods*, vol. 173 of *Lecture Notes in Statistics*. New York: Springer-Verlag. Papers from the TMR and MaPhySto Summer School held at Aalborg University, Aalborg, August 19–22, 2001.
- NASH, A. J. (1978). *A method for classifying shortleaf pine sites in Missouri*. Research Bulletin 824. Columbia, MO: Missouri Agricultural Experiment Station.
- NEAL, R. M. (2003). Slice sampling. *Ann. Statist.* **31**, 705–767. With discussions and a rejoinder by the author.
- ROBERT, C. P. (1994). *The Bayesian choice: a decision-theoretic motivation*. Springer-Verlag Inc.
- SHIFLEY, S. and BROOKSHIRE, B. (2000). *Missouri Ozark Forest Ecosystem Project: site history, soils, landforms, woody and herbaceous vegetation, down wood, and inventory methods for the landscape experiment (eds.)*. General Technical Report NC-208, St. Paul, MN: The United States Department of Agriculture, Forest Service, North Central Research Station.
- SHIFLEY, S. and KABRICK, J. (2002). *Proceedings of the Missouri Ozark Forest Ecosystem Project Symposium: post-treatment results of the landscape experiment (eds.)*. General Technical Report NC-227, St. Paul, MN: U.S. Dept. of Agriculture, Forest Service, North Central Research Station.
- SINHA, B. K. and GHOSH, M. (1987). Inadmissibility of the best equivariant esti-

- mators of the variance-covariance matrix, the precision matrix, and the generalized variance under entropy loss. *Statistics & Decisions* **5**, 201–227.
- STEIN, M. L. (1999). *Interpolation of spatial data: some theory for kriging*. Springer-Verlag Inc, New York.
- SUN, D. and SUN, X. (2005a). Estimation of the multivariate normal precision and covariance matrices in a star-shaped model. *Annals of the Institute of Statistical Mathematics* **57**, 455–484.
- SUN, D. and SUN, X. (2006). Estimation of multivariate normal precision and covariance matrices in a star-shaped model with missing data. *Journal of Multivariate Analysis* **97**, 698–719.
- SUN, X. (1998). Improved estimation of the generalized precision under the squared loss. *Communications in Statistics, Part A – Theory and Methods* **27**, 2725–2742.
- SUN, X. (1999a). Improved estimation of the generalized variance under the entropy loss. *J. East China Norm. Univ. Natur. Sci. Ed.* , 23–32.
- SUN, X. (1999b). Improvement on the best affine equivariant estimation of the covariance matrix under the entropy loss. *Chinese J. Appl. Probab. Statist.* **15**, 168–175.
- SUN, X. and HE, Z. (2005). An efficient algorithm for bayesian spatial analysis via the generalized Ratio-of-Uniforms method. Preprint.

- SUN, X., HE, Z. and KABRICK, J. (2006). Bayesian spatial prediction of the site index in the study of the Missouri Ozark Forest Ecosystem Project. Preprint.
- SUN, X. and PANG, W. (2000). Improved estimation of the generalized precision under the entropy loss. *Acta Math. Appl. Sinica (English Ser.)* **16**, 162–170.
- SUN, X. and SUN, D. (2005b). Estimation of the Cholesky decomposition of the covariance matrix for a conditional independent normal model. *Statist. Probab. Lett.* **73**, 1–12.
- WAKEFIELD, J. C., GELFAND, A. E. and SMITH, A. F. M. (1991). Efficient generation of random variates via the ratio-of-uniforms method. *Statistics and Computing* **1**, 129–133.
- WEBSTER, R. and MCBRATNEY, A. B. (1987). Mapping soil fertility at broom's barn by simple kriging. *Journal of the Science of Food and Agriculture* **38**, 97–115.
- WHITTAKER, J. (1990). *Graphical models in applied multivariate statistics*. John Wiley & Sons, New York.
- WHITTAKER, R. (1967). Gradient analysis of vegetation. *Biological Review* **49**, 207–264.
- WONG, F., CARTER, C. and KOHN, R. (2002). Efficient estimation of covariance selection models. In *Bayesian Statistics 7. Proceedings of the Seventh Valencia International Meeting*.

- YANG, R. and BERGER, J. O. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics* **22**, 1195–1211.
- ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *J. Amer. Statist. Assoc.* **99**.
- ZHANG, H. and ZIMMERMAN, D. L. (2005). Toward reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92**, 921–936.
- ZHOU, X., SUN, X. and WANG, J. (2001). Estimation of the multivariate normal precision matrix under the entropy loss. *Annals of the Institute of Statistical Mathematics* **53**, 760–768.
- ZHU, J., LAHIRI, S. N. and CRESSIE, N. (2002). Asymptotic inference for spatial CDFS over time. *Statistica Sinica* **12**, 843–861.

VITA

Xiaoqian Sun was born on February 12, 1967 in Jiangsu Province, China. He graduated with B.A. in mathematics from Nanjing Normal University in 1987. He also received master's degree and doctoral degree in mathematical statistics in 1993 and 1999 respectively, both from East China Normal University, Shanghai. Before he came in the United States in 2002, he worked as an assistant/associate professor in Department of Mathematics, Huaiyin Teachers' College, Jiangsu for several years. He has accepted a position beginning August 2006 as an assistant professor in Department of Mathematical Sciences at Clemson University, South Carolina.