UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

OSAFU AUGUSTINE EGBON

Bayesian Spatial Process Models for Activation Patterns in Transcranial Magnetic Stimulation Mapping

Dissertation presented to the Department of Statistics - Des/UFSCar and to the Institute of Mathematical and Computer Sciences - ICMC-USP, as part of the requirements for obtaining the title of Doctor in Statistics - Interinstitutional Graduate Program in Statistics UFSCar-USP.

Supervisor: Prof. Francisco Louzada Neto

São Carlos
July 2023

UNIVERSIDADE FEDERAL DE SÃO CARLOS
CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA
PROGRAMA INTERINSTITUCIONAL DE PÓS-GRADUAÇÃO EM ESTATÍSTICA UFSCar-USP

OSAFU AUGUSTINE EGBON

Modelos de Processo Espacial Bayesiano para Padrões de Ativação em Mapeamento de Estimulação Magnética Transcraniana

Tese apresentada ao Departamento de Estatística – Des/UFSCar e ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Estatística - Programa Interinstitucional de Pós-Graduação em Estatística UFSCar-USP.

Orientador: Prof. Francisco Louzada Neto

São Carlos
Julho de 2023

UNIVERSIDADE FEDERAL DE SÃO CARLOS

Centro de Ciências Exatas e de Tecnologia
Programa Interinstitucional de Pós-Graduação em Estatística

Folha de Aprovação

Defesa de Tese de Doutorado do candidato Osafu Augustine Egbon, realizada em 07/07/2023.

Comissão Julgadora:

Prof. Dr. Francisco Louzada Neto (USP)

Prof. Dr. Diego Carvalho do Nascimento (UDA)

Prof. Dr. Paulo Henrique Ferreira da Silva (UFBA)

Prof. Dr. Christian Rudolf Heumann (LMU)

Prof. Dr. Ezra Gayawan (FUTA)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa Interinstitucional de Pós-Graduação em Estatística.

To
Celestina Egbon,
Mariella Bogoni, and
my family.

ACKNOWLEDGEMENTS

I am deeply grateful to God for His divine assistance, sound health, and provisions that have enabled me to complete my Ph.D. program successfully.

I express my sincere appreciation to my supervisor, Prof. Francisco Louzada Neto for his invaluable contribution, unwavering support, and guidance throughout my research journey. I also want to appreciate Dr. Diego Nascimento and Prof. Christian Heumann for their expertise which has been instrumental in shaping my research and helping me achieve my research goals.

I am grateful to Dr. Onno van der Groen and Prof. Dylan Edwards for data provision and to Dr. Oilson Gonzzato for his contribution at the early stage of the Shiny application development. I am also grateful to Assoc. Prof. Marcos Prates for his technical ideas that led to the improvement in the modeling and computational speed attained in this work, and to Dr. Ezra Gayawan for introducing me to novel modeling skills, which led to the improvement of this work.

I sincerely appreciate the Coordination for the Improvement of Higher Education Personnel – Brazil (CAPES) for the support of this research. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 and PDSE 88881.690056/2022 - 01.

I am indebted to Profs. Vera Tomazella, Carlos Diniz, Pedro Ramos, and the Professors of the PIPGES - UFSCar - USP program for teaching me statistical methods and computational techniques that have been invaluable in the development of my project.

I extend my heartfelt appreciation to my Dad, Mr. Patrick Egbon, and my Siblings, Mr. Efosa, Mrs. Osarenakhue, Mrs. Ayevbosa, Mrs. Osagiodagbon, and Mr. Oghosa, who provided me with all kinds of support, love, and prayers since the commencement of my Ph.D. program. Also, I would like to extend my appreciation to Mrs. Francismara Cruz and her parents for their love and prayers.

To my friends, Asrat Belachew, Alex Mota, Juan Bustamante, Alex Huayanay, Isaac Olmos, Danillo Assunção, Ritha Condori, Ximena Lozada, and colleagues of the PIPGES, and those who have shared a laugh, a meal, or a drink with me during this time.

“Neurons that fire together wire together.”
(Donald Hebb)

ABSTRACT

EGBON, O. A. **Bayesian Spatial Process Models for Activation Patterns in Transcranial Magnetic Stimulation Mapping**. 2023. 160 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

In recent years, Spatial statistical models have been gaining rapid attention for solving problems in biological systems due to the improvement in spatial data collection. It has proven extremely important in unveiling spatial patterns and predicting biological processes. This project developed novel parametric and nonparametric Bayesian spatial statistical models to analyze data generated by the muscular responses elicited by Transcranial magnetic stimulation (TMS) pulses induced on the motor cortex of a patient. The goal is to unveil new insights into patients' response patterns important for achieving successful TMS therapy sessions. The first contribution of this project is a systematic review and meta-analysis of the existing Bayesian spatial models that could be considered for analyzing TMS datasets. The second contribution is the development of a user-friendly interface for performing Bayesian spatial modeling for analyzing TMS datasets based on state-of-the-art methods. The interface was documented in an R package, which is publicly available. The third contribution proposed novel spatial statistical models for integrating geostatistical datasets in the form of prior elicitation in a Bayesian analysis. The models were validated using simulation studies, and findings show that naively integrating geostatistical TMS datasets without ensuring the consistency of the data is detrimental to the desired inferences. The final contribution proposed a Bayesian nonparametric spatial model that leads to a non-stationary and non-Gaussian spatial process for the joint modeling of geostatistical TMS datasets. The method used a mixture of Dependent Dirichlet processes to share information across sub-spatial models. Two simulation studies were used to validate the model performance, and the result showed superior performance compared with independent and exchangeable models. The main finding of this work is that the primary motor cortex within the motor cortex region of the brain is responsible for the largest activation in the movement of the right first dorsal interosseous muscle. The finding also showed that the corticospinal excitability decreases with multiple TMS pulses on the motor cortex; however, it begins to regain its excitability strength after several stimulations. The findings from this project could guide TMS practitioners to improve patients' treatment experiences.

Keywords: Brain mapping, Dirichlet process, Gaussian process, Motor cortex, Prior elicitation.

RESUMO

EGBON, O. A. **Modelos de Processo Espacial Bayesiano para Padrões de Ativação em Mapeamento de Estimulação Magnética Transcraniana**. 2023. 160 p. Tese (Doutorado em Estatística – Programa Interinstitucional de Pós-Graduação em Estatística) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2023.

Nos últimos anos, os modelos estatísticos espaciais têm recebido rápida atenção para resolver problemas em sistemas biológicos devido ao aprimoramento na coleta de dados espaciais. Eles têm se mostrado extremamente importantes na revelação de padrões espaciais e na previsão de processos biológicos. Este projeto desenvolveu novos modelos estatísticos espaciais paramétricos e não paramétricos Bayesianos para analisar dados gerados pelas respostas musculares desencadeadas por pulsos de estimulação magnética transcraniana (TMS) induzidos no córtex motor de um paciente. O objetivo é descobrir novas perspectivas sobre os padrões de resposta dos pacientes, um fator importante para o sucesso das sessões de terapia com TMS. A primeira contribuição deste projeto é uma revisão sistemática e meta-análise dos modelos espaciais Bayesianos existentes que podem ser considerados para analisar conjuntos de dados de TMS. A segunda contribuição é o desenvolvimento de uma interface do usuário para realizar modelagem espacial Bayesianas para análise de conjuntos de dados de TMS com base em métodos de última geração. A interface foi documentada em um pacote R, que está disponível publicamente. A terceira contribuição propôs novos modelos estatísticos espaciais para integrar conjuntos de dados geoestatísticos na forma de elicitação de priori em uma análise Bayesiana. Os modelos foram validados usando estudos de simulação, e os resultados mostram que a integração ingênua de conjuntos de dados geoestatísticos de TMS sem garantir a consistência dos dados é prejudicial para as inferências desejadas. A contribuição final propôs um modelo espacial não paramétrico Bayesiano que leva a um processo espacial não estacionário e não gaussiano para a modelagem conjunta de conjuntos de dados geoestatísticos de TMS. O método utilizou uma mistura de processos de Dirichlet dependentes para compartilhar informações entre os submodelos espaciais. Dois estudos de simulação foram usados para validar o desempenho do modelo, e o resultado mostrou desempenho superior em comparação com modelos independentes e intercambiáveis. O principal resultado deste trabalho é que o córtex motor primário, dentro da região do córtex motor do cérebro, é responsável pela maior ativação no movimento do músculo interósseo dorsal do primeiro dedo direito. Os resultados também mostraram que a excitabilidade corticospinal diminui com múltiplos pulsos de TMS no córtex motor; no entanto, começa a recuperar sua força de excitabilidade após várias estimulações. Tais resultados podem orientar os profissionais de TMS a melhorar a experiência de tratamento dos pacientes.

Palavras-chave: Córtex motor, Elicitação a priori, Mapeamento Cerebral, Processo Gaussiano,

Processo de Dirichlet.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

CHAPTER

1

INTRODUCTION

1.1 Background

The improvement in data collection methods has increased the availability of spatial data, and researchers are increasingly faced with the task to analyze them, which consequently makes spatial statistical models becoming more popular. A spatial statistical model is a large class of models widely applied to study spatial patterns of spatially correlated data and predictions on unobserved locations. It has been considered for solving problems in epidemiology (AUCHINCLOSS *et al.*, 2012), agriculture (SERNEELS; LAMBIN, 2001), medicine (XIE *et al.*, 2023; SCHLUTH *et al.*, 2023), and engineering (GARCIA *et al.*, 2016). Spatial statistical models are high-dimensional models, as a result, they are often considered within a Bayesian framework. The main benefit of the Bayesian technique over the frequentist counterpart is the possibility of estimating the complete distribution of the parameters that govern the model (BERGER; GARCÍA, 2004).

A Bayesian spatial statistical model is a form of the generalized linear mixed model that includes a spatial latent variable to account for spatial variation in spatially indexed data. The latent variable is modeled through a spatial prior, usually, multivariates, which takes account of spatial dependences across a domain under investigation. A common family of spatial prior distributions is the Gaussian Markov Random Field (CRESSIE, 1993; RUE; HELD, 2005), due to its friendly theoretical properties. However, the Student-t spatial process (ORDOÑEZ *et al.*, 2023), Poisson-Gaussian Mixture Process (GONÇALVES; AGUILAR; PRATES, 2022), and Dirichlet processes (GELFAND; KOTTAS; MACEACHERN, 2005) are other possible choices.

Prior elicitation is a key element in Bayesian inference. It provides a framework to combine beliefs in the form of prior probability distributions with experimental data. Prior knowledge is elicited from experts, as in the work of Hoegh *et al.* (2017) or elicited from historical data as in Feißt, Krisam and Kieser (2020). Due to the high dimensionality of spatial

models, prior elicitation from experts becomes difficult, thus elicitation from data becomes appealing. Prior distribution derived from historical data can provide increased precision of parameter estimates. Power and commensurate prior distributions (HOBBS *et al.*, 2011) are existing methods for eliciting knowledge from historical non-geostatistical spatial data and have been adopted in several fields. However, there is still a lack of existing methods for eliciting knowledge from geostatistical data. Due to the possibility that historical geostatistical data may have a different spatial window and exhibit some level of misalignment, it is not trivial to extend these classical methods to geostatistical data. Thus, efforts are required for developing prior elicitation frameworks for geostatistical data for solving practical problems.

In biological systems, there are often multiple observed spatial datasets available for a phenomenon under investigation. This can happen for a couple of reasons. Firstly, there may be a need to reobserve data after a significant event that could potentially alter the process of the phenomenon has occurred. Secondly, researchers may be interested in observing the phenomenon's behavior under different conditions, which can provide valuable insights and lead to a more comprehensive understanding of the underlying processes. However, the inconsistency among the datasets may lead to an inadequate inference when jointly analyzed. In these cases, a combination of sub-spatial models with the capability to filter and share consistent information across sub-models becomes essential. Though there are existing methods for joint analysis of non-geostatistical data (MÜLLER; QUINTANA; ROSNER, 2004; LOPES; MÜLLER; ROSNER, 2003), little or no attention has been given to geostatistical data. The existing methods lack the framework for a trivial extension to geostatistical data, thus creating a significant gap that requires adequate attention.

The most recent and popular `R` package for fitting a Bayesian spatial statistical model is the `INLA` (RUE; MARTINO; CHOPIN, 2009). Though the package relies on the Gaussian Markov Random Field priors, it allows the implementation of a wide range of spatial models for area/lattice and geostatistical spatial data. A major advantage of the package is the possibility of developing and documenting wrapper functions in `R`, that could be used to solve more specific problems. For example, Azevedo, Prates and Bandyopadhyay (2022) developed wrapper functions documented as a package for solving spatial confounding issues and Palmí-Perales, Gómez-Rubio and Martinez-Beneito (2021) developed a wrapper function for performing analysis with multivariate spatial prior for multiple spatial latent fields. Despite the importance of Bayesian spatial modeling in solving practical problems, there has been a lack of effort in developing and documenting wrapper functions that can generate a user-friendly interface. This interface would allow individuals to perform full Bayesian spatial modeling without requiring extensive coding expertise. This gap in development and documentation has hindered the accessibility and usability of Bayesian spatial modeling. As such, there is a need for increased attention to creating more accessible and user-friendly interfaces for Bayesian spatial modeling.

The models and methods developed in this work were motivated by data derived from

a medical experimental process on brain-spinal functionality that brought to light the existing gaps identified above. The experiment aimed at understanding human responses to Transcranial Magnetic Stimulation (TMS) pulses delivered to the motor cortex. TMS is an active neurophysiological technique that causes electrical induction in the brain. TMS has a range of applications, for example; studying brain functioning in healthy and unhealthy subjects (HALLETT, 2000), as a tool to help predict recovery outcomes after stroke (STINEAR *et al.*, 2017), and as an intervention to modulate brain activity (LEFAUCHEUR *et al.*, 2014). The TMS pulses delivered from its coil form a magnetic field that painlessly passes through the scalp to reach and depolarise neurons in the brain region under investigation (ROTENBERG; HORVATH; PASCUAL-LEONE, 2014). This stimulation causes the activation of some biological neurons, and action potentials are released in the process. Specifically for this experiment, these action potentials flow through the cortical-spinal tract to reach the right first dorsal interosseous (FDI) muscle. The stimulation effect is then recorded through an electromyograph placed over the FDI muscle. The responses are measured in microvolts and are referred to as motor-evoked potential (MEP).

1.2 Objectives

This project seeks to quantify and explain the spatial pattern of patients' responses to TMS pulses and to identify areas on the motor cortex that respond most to stimulation. Hence, this work aims to develop spatial statistical models and methods that allow the estimation of spatial patterns and quantify uncertainties in patients' responses to TMS pulses, which could give new insight into the biological process and be used to determine the accurate resting motor threshold and treatment doses. This is crucial given that a TMS therapy session heavily relies on accurately determining the true spatial response pattern of the patient undergoing treatment. This factor directly impacts the effectiveness of the therapy. Therefore, it is imperative to prioritize the precise identification of patients' spatial response patterns to ensure optimal treatment outcomes.

To achieve the goal of this project,

1 A systematic review and meta-analysis of the existing Bayesian spatial statistical methods were conducted. This helped to reveal the state-of-the-art methods in diverse fields and identified gaps in the literature.

2 Based on the state-of-the-art methods and models, a user-friendly interface formed from a wrapper function for `INLA` was developed to analyze brain TMS data. A standard `R` documentation procedure was adopted and was named `TMSBrainApp` package, which is publicly available and installable from the GitHub repository.

3 Power and commensurate prior elicitation methods from historical point-referenced data were developed for a Gaussian spatial process for modeling point-reference data to improve the estimation of spatial patterns and uncertainty in TMS data.

4 Mixture of Dirichlet processes was proposed to jointly model multiple TMS data to profile shared from subject-specific latent spatial processes. The resulting model is a non-Gaussian and non-stationary spatial process.

The remaining chapters of this dissertation are structured as follows: Chapter 2 provides an overview of the methods related to this work and presents the TMS data used in this project. Chapter 3 presents the systematic review and meta-analysis results of Bayesian spatial models. Chapter 4 introduces the developed `TMSBrainApp` package and user-friendly shiny interface, while Chapter 5 presents the proposed prior elicitation methods for geostatistical data. Chapter 6 proposes a mixture of Dirichlet process prior distributions for the joint analysis of the TMS data, and finally, Chapter 7 concludes the work and gave future considerations.

1.3 Scientific products

This dissertation has generated the following scientific articles that are either published or under review:

1 Louzada, F., Nascimento, D. C., & Egbon, O. A. (2021). Spatial Statistical Models: An Overview under the Bayesian Approach. *Axioms*.

2 Egbon O. A., Nascimento D. C. & Louzada F. (2023). Prior Elicitation for Gaussian Spatial Process: an Application to TMS Mapping Data. *Statistics in Medicine*.

3 Egbon O. A., Heumann C., Nascimento D. C. & Louzada F. Mixtures of Dirichlet processes for joint spatial modeling of TMS mapping data. Under review in *The Journal of the Royal Statistical Society, Series C (Applied Statistics)*.

4 Nascimento D. C., Egbon, O. A., Groen O., Gonzzato O., Louzada F., & Edwards D. `TMSBrainApp`: A Shiny Interface for Bayesian Spatial Statistical Hotspot Finding Task in a TMS Brain Mapping Data. Submitted to *PLOS Computational Biology*.

CHAPTER

2

REVIEW OF METHODS AND DATA SOURCE

2.1 Generalized linear mixed model

Linear regression plays a fundamental role in statistical modeling (NETER *et al.*, 1996). It is the basis through which complex models are developed. It is an approach for explaining the occurrence of a random vector $\mathbf{Y} = (Y_1, Y_2, ..., Y_m)^T$ by some observed covariate set $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_p)$, and $\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{im})^T$. Its goal is to estimate the function that links both $\mathbf{X}$ and $\mathbf{Y}$. The linear regression model is given by

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\beta}, \tau &\sim Normal(\boldsymbol{\mu}, \tau^{-1}\mathbf{I}_m), \\ \boldsymbol{\mu} &= \mathbf{X}^T\boldsymbol{\beta}, \end{aligned} \tag{2.1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^m$, $\mathbf{I}_m$ is an identity matrix of size m; parameter $\boldsymbol{\beta} = (\beta_1, \beta_2, ..., \beta_p)^T$ and β_i is the regression effects of covariate $\mathbf{x}_i$, which gives insight on how the covariate contributes to the behavior of $\mathbf{Y}$. (2.1) assumes that conditioning on τ and $\boldsymbol{\beta}$, the occurrence of $Y_i\,(i = 1, 2, ..., m)$ are independent and identical. In practical problems, this assumption is hardly true. Hence, (2.1) can be extended to a linear mixed model (NELDER; WEDDERBURN, 1972), which includes a latent variable that explains some unobserved variation in $\mathbf{Y}$. The linear mixed model is given as

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\psi}, \tau &\sim Normal(\boldsymbol{\mu}, \tau^{-1}\mathbf{I}_m), \\ \boldsymbol{\mu} &= \mathbf{X}^T\boldsymbol{\beta} + \boldsymbol{\psi}, \end{aligned} \tag{2.2}$$

where $\boldsymbol{\psi} = (\psi_1, \psi_2, ..., \psi_m)^T$, and ψ_i is a latent variable corresponding to Y_i. Nelder and Wedderburn (1972) refer to (2.2) as a linear model for systematic effect. The marginal distribution of $\mathbf{Y} \mid \boldsymbol{\beta}, \tau$ again has a normal distribution with mean $\mathbb{E}(\mathbf{Y} \mid \boldsymbol{\beta}, \tau) = \mathbf{X}^T\boldsymbol{\beta}$ and the marginal covariance is $\mathbb{C}ov(\mathbf{Y} \mid \boldsymbol{\beta}, \tau) = \tau\mathbf{I}_m + \mathbf{Q}_\psi$, where $\mathbf{Q}_\psi$ is the covariance matrix of the latent variable $\boldsymbol{\psi}$, known as the random effect. In many problems, the variable Y_i is not normally distributed, hence the generalized linear mixed model (GLMM) (BRESLOW; CLAYTON, 1993) is adopted. (2.2) can be used to investigate subject variations in longitudinal data by taking ψ_i to be a source

of variation particular to subject i. According to Nelder and Wedderburn (1972), the GLMM comprises a dependent variable $\mathbf{Y}$ whose distribution belongs to the exponential family, and a set of covariates $\mathbf{X}$ such that $g(\mathbb{E}(\mathbf{Y})) = \mathbf{X}^T\boldsymbol{\beta} + \boldsymbol{\psi}$ with a link function $g^{-1} : \mathbb{R} \to \mathscr{Y}$. Thus, the GLMM follows as

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\psi} &\sim f(\mathbf{Y}; \boldsymbol{\mu}), \\ \boldsymbol{\mu} &= g^{-1}(\mathbf{X}^T\boldsymbol{\beta} + \boldsymbol{\psi}), \end{aligned} \tag{2.3}$$

where $\mathbb{E}(\mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\psi}) = \boldsymbol{\mu}$. The linear mixed model is a special case of the GLMM. That is f is normal and $g(\boldsymbol{\mu}) = \boldsymbol{\mu}$. Different probability distributions assumed for f lead to a different type of model. For example, Yau, Lee and Ng (2002), Fabio, Paula and Castro (2012), and Bonat, Jr and Zeviani (2015) adopted gamma, poison, and beta distributions respectively. The GLMM model has been adopted in several practical applications (BOLKER *et al.*, 2009; STROUP, 2012).

2.2 Spatial statistical model

A spatial statistical model is an extension of the GLMM to accommodate geographically referenced data. It is adopted to account for spatially structured sources of variation in observed data. It has been applied in a wide range of fields (LOUZADA; NASCIMENTO; EGBON, 2021). For example, in epidemiology, investigators might wish to analyze disease rates at a county or area level; in Environmental science, investigators might wish to analyze CO_2 emission at a regional level, and in Biological science, investigators might wish to analyze neighboring cell interactions in single-cell RNA sequencing data.

Suppose Y_i is an observed random variable at a known location $s_i, i = 1, 2, ..., m$. Then a spatial statistical model can be expressed as (CRESSIE, 1993; BANERJEE; CARLIN; GELFAND, 2014)

$$\begin{aligned} \mathbf{Y} \mid \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\theta} &\sim f(\mathbf{Y}; \boldsymbol{\mu}), \\ \boldsymbol{\mu} &= g^{-1}(\mathbf{X}^T\boldsymbol{\beta} + \boldsymbol{\psi} + \boldsymbol{\theta}), \end{aligned} \tag{2.4}$$

where $\mathbf{Y} = (Y_1, Y_2, ..., Y_m)^T$; $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_m)^T$ is the spatial process and θ_i could account for the source of variation in location s_i. In area spatial data, s_i could represent a county, sub-nationals, or districts within the explorable spatial domain, and it could represent a geographical coordinate in geostatistical or point-referenced data.

The model given in (2.4) is flexible and could incorporate several interactions. For a special case, suppose f is normal, $g(\boldsymbol{\mu}) = \boldsymbol{\mu}$, and conditional variance of $\mathbf{Y}$ is $\tau^{-1}\mathbf{I}_m$, marginal covariances of $\boldsymbol{\psi}$ and $\boldsymbol{\theta}$ are $\mathbf{Q}_{\psi}^{-1}$ and $\mathbf{Q}_{\theta}^{-1}$ with zero mean. Thus, the marginal distribution of $\mathbf{Y}$ follows as

$$\mathbf{Y} \mid \boldsymbol{\beta} \sim Normal(\mathbf{X}^T\boldsymbol{\beta}, \boldsymbol{\Sigma}), \tag{2.5}$$

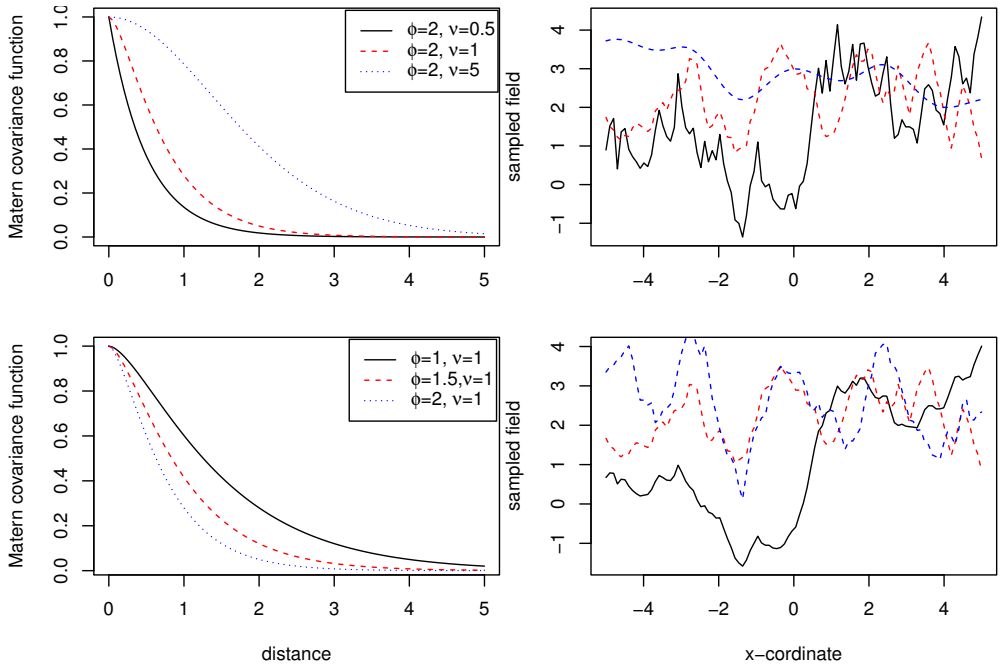


Figure 1 – An example of Matérn covariance function and the associated random field.

where $\mathbf{\Sigma} = (\tau^{-1}\mathbf{I}_n + \mathbf{Q}_\psi^{-1} + \mathbf{Q}_\theta^{-1})$, which introduces high-level interactions between observations at different locations.

In geostatistical data modeling, $\mathbf{Q}_\theta^{-1}$ can be determined through a covariance function such as exponential, Gaussian, and Matérn covariance functions. The variogram of the Matérn class is given as (BANERJEE; CARLIN; GELFAND, 2014)

$$\gamma(h) = \lambda^2 + \sigma^2\Big[1 - \frac{(2\sqrt{\nu}h\phi)^\nu}{2^{\nu-1}\Gamma(\nu)}K_\nu(2\sqrt{\nu}h\phi)\Big], h > 0. \tag{2.6}$$

Consequently, the matérn covariance function $C(h;\phi,\nu)$ is given as

$$\begin{aligned} C(h;\phi,\nu) &= lim_{u\to\infty}\gamma(u) - \gamma(h) \\ &= \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)}(2\sqrt{\nu}\phi h)^\nu K_\nu(2\sqrt{\nu}\phi h), \end{aligned} \tag{2.7}$$

where $h = d(s_i, s_{i'})$ is an euclidean distance between two points, K_ν is a Bessel function of the second kind of order ν, ν is the smoothness parameter and ϕ is the range. Figure 1 illustrates an example of a random field for model (2.5) govern by a matérn covariance function, (2.7). The spatial process generated in the figure assumed $\mathbf{X}$ is an $n \times 1$ matrix; that is, $\mathbf{X} = [x_{i1}]$, and $x_{i1} = 1, \forall i$, $\beta = 2$, $\sigma^2 = 1$, and $\mathbf{\Sigma} = \frac{1}{10}\mathbf{I}_m + [C(d(s_i, s_{i'});\phi,\nu)]$. The left panel of the figure shows the covariance function for different choices of range and smoothness parameters, and the right panel shows the corresponding sampled field for $\mathbf{Y}$ across one of the coordinates. The figure shows how ϕ and ν control the smoothness of the random field. The slower the covariance curve approaches zero with an increase in distance, the smoother the spatial process. This is also evident in the 2-dimensional form of the field shown in Figure 2.

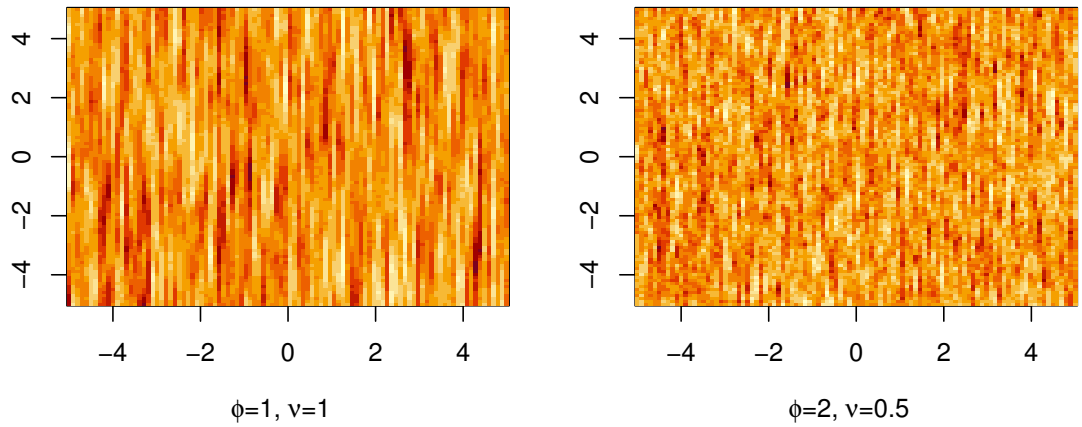


Figure 2 – An example of Matérn covariance function and the associated random field in 2D.

2.3 Bayesian analysis

Bayesian analysis is a powerful statistical modeling technique that leverages the Bayes theorem to update the probability of a hypothesis as new evidence is gathered through data (BERGER *et al.*, 1994). This approach allows for a more nuanced and accurate understanding of complex systems, as it takes into account the uncertainty inherent in any real-world scenario. It provides a flexible and adaptable framework for modeling complex systems. In Bayesian modeling, all model parameters are considered random and are represented by a probability distribution before data observation.

Consider model (2.4) for spatial modeling. Excluding the fixed and random effects for simplicity, and assuming a Gaussian random field prior distribution for the process $\boldsymbol{\theta}$, the hierarchical model follows as

$$\begin{aligned} \boldsymbol{\theta} &\sim Normal(\mathbf{0}, \mathbf{Q}_\theta^{-1}), \\ \mathbf{Y} \mid \boldsymbol{\theta} &\sim f(\mathbf{Y}; \boldsymbol{\theta}). \end{aligned} \tag{2.8}$$

Where $\mathbf{Q}_\theta$ is the precision matrix of the random field (eg. spatial effect). The first equation in (2.8) is the process model and the second is the data generation model which induces measurement errors around the process model. To make an inference about $\boldsymbol{\theta}$, the prior knowledge is updated with the available data $\mathbf{Y}$ using Bayes theorem. If f is Gaussian and $\mathbf{Q}_\theta$ is known, the conditional posterior distribution $\boldsymbol{\theta} \mid \mathbf{y}$ after some matrix algebras will again result in a multivariate Gaussian distribution (BANERJEE; CARLIN; GELFAND, 2014). In practice, these and similar assumptions may be too restrictive, and relaxing them will consequently lead to an intractable posterior probability distribution. In this case, statistical inference will rely on computational techniques.

The Markov Chain Monte Carlo (MCMC) algorithm (BROOKS *et al.*, 2011) is one of the most frequently used algorithms in the literature for estimating intractable and complex posterior distributions. However, other methods such as Variational Bayes (VB) (ATTIAS, 1999), Integrated Nexted Laplace Approximation (INLA) (RUE; MARTINO; CHOPIN, 2009), and

Expectation Maximization (EM) (MOON, 1996) are other popular options. In this project, the MCMC and INLA methods were adopted for posterior inferences.

2.4 Prior elicitation from historical data

Prior elicitation in Bayesian inference is essential for obtaining more precise estimates. It allows prior knowledge learned from experts or historical data to be constructed into a probability distribution for a Bayesian analysis. Arriving at such distribution is difficult in geostatistical modeling due to the high dimensional parameters involved. Chapter 5 developed a Spatial power and commensurate informative prior distributions to elicit knowledge from historical data. These distributions were rooted in the power and commensurate prior distributions for non-geostatistical data proposed in Ibrahim and Chen (2000) and Hobbs *et al.* (2011).

2.4.1 Power prior

In the simplest form, suppose the interest is to estimate $\boldsymbol{\theta}$ in a current data $\mathscr{D} = \{m, \mathbf{y}\}$ with available historical data $\mathscr{D}_0 = \{m_0, \mathbf{y}_0\}$, where m and $\mathbf{y}$ are the sample size and an $m \times 1$ vector of a response variable, and m_0 and $\mathbf{y}_0$ correspond to the historical quantities. As discussed in Ibrahim, Chen and Sinha (2001), the power prior for $\boldsymbol{\theta}$ is given as

$$\pi(\boldsymbol{\theta}|\omega_0, \mathscr{D}_0) \propto (L(\boldsymbol{\theta}|\mathscr{D}_0))^{\omega_0} \pi(\boldsymbol{\theta}), \tag{2.9}$$

where $L(\boldsymbol{\theta}|\mathscr{D}_0)$ is the historical likelihood, $\pi(\boldsymbol{\theta})$ is the prior distribution assumed on $\boldsymbol{\theta}$ before the historical data were observed, and ω_0 is the discounting parameter controlling the amount of the historical information incorporated into the current study and also controls the heaviness of the tail, which becomes heavier as ω_0 becomes smaller. Ibrahim, Chen and Sinha (2001) suggested restricting ω_0 to range in the interval of $[0, 1]$. The two extremes to the amount of information incorporated into the current model are $\omega_0 = 0$, which indicates that no historical information is incorporated, and $\omega_0 = 1$, which indicates that all the historical information is incorporated into the current model. Care is taken in the choice of ω_0, in that poor choice can cause the historical data to dominate the current study, which is undesirable. The main benefit of using the power prior in (2.9) is the few hyper-parameter required to be specified in model selection problems, and in many instances, it is proper and can be used to compute the Bayes factor.

It is not trivial to determine the amount of the true discount that minimizes type 1 error using the power prior given in (2.9). Thus, Ibrahim and Chen (2000) proposed a joint power prior obtained by allowing ω_0 to be random and assigned a prior distribution $\pi(\omega_0)$ as

$$\pi(\boldsymbol{\theta}, \omega_0|\mathscr{D}_0) \propto (L(\boldsymbol{\theta}|\mathscr{D}_0))^{\omega_0} \pi(\boldsymbol{\theta})\pi(\omega_0). \tag{2.10}$$

A natural choice of $\pi(\omega_0)$ is the beta distribution; however, a truncated normal and a truncated gamma distribution in the interval of $[0, 1]$ could be considered (DUAN, 2005; IBRAHIM;

CHEN, 2000). The parameter ω_0 can be interpreted as a measure of heterogeneity between the historical and the current data. It can be considered as the probability that $\mathscr{D}_0$ and $\mathscr{D}$ came from the same population. In this context, $\omega_0 = 1$ indicates that $\mathscr{D}_0$ and $\mathscr{D}$ came from the same population, and all the information on $\mathscr{D}_0$ is desired in the current study. Contrarily, $\omega_0 = 0$ indicates that the $\mathscr{D}_0$ and $\mathscr{D}$ came from a completely different population and should not be incorporated in the current study. As ω_0 becomes smaller, the heterogeneity in the data increases.

Ibrahim, Chen and Sinha (2001) discussed the propriety of (2.10) and importantly stated a sufficient, but not necessary, condition that the prior given in (2.10) is guaranteed to be proper (i.e $\int \pi(\boldsymbol{\theta}, \omega_0|\mathscr{D}_0)d\boldsymbol{\theta}d\omega_0 < \infty$) if $\pi(\boldsymbol{\theta})$ and $\pi(\omega_0)$ are themselves proper. Specifically, (2.10) exhibits the properties of a likelihood, and the propriety is useful in model selection problems.

Duan (2005) proposed a modification of the joint power prior, also known as the normalized joint power prior which follows as

$$\pi(\boldsymbol{\theta}, \omega_0|\mathscr{D}_0) \propto \frac{(L(\boldsymbol{\theta}|\mathscr{D}_0))^{\omega_0}\pi(\boldsymbol{\theta})\pi(\omega_0)}{\int_\Theta (L(\boldsymbol{\theta}|\mathscr{D}_0))^{\omega_0}\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}\mathbf{1}_{\Omega_0}(\omega_0), \tag{2.11}$$

where $\Omega_0 = \{\omega_0 : 0 < \int_\Theta (L(\boldsymbol{\theta}|\mathscr{D}_0))^{\omega_0}\pi(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty\}$, and $\mathbf{1}_B(a)$ is an indicator function that takes the value 1 if $a \in B$ and 0 otherwise. The prior on (2.11) is always proper whenever $\pi(\omega_0)$ is a proper distribution but not necessarily $\pi(\boldsymbol{\theta})$. The power prior in (2.10) specifies a joint prior distribution directly for $(\boldsymbol{\theta}, \omega_0)$, in which case its normalizing constant is independent of ω_0 and can be viewed as the normalized power prior (2.11), which first specifies a conditional prior distribution for $\boldsymbol{\theta}$ given ω_0 and then a marginal distribution for ω_0 (IBRAHIM *et al.*, 2015). As described in Duan (2005) and Ibrahim and Chen (2000) the joint prior distributions in (2.10) and (2.11) can be extended to multiple historical datasets. Given a historical dataset $\mathscr{D}_0 = \{\mathscr{D}_{01}, \mathscr{D}_{02}, ..., \mathscr{D}_{0m}\}$ with associated weights $\omega_0 = \{\omega_{01}, \omega_{02}, ..., \omega_{0m}\}$ of $\mathscr{D}_0$ adherence to the current data, the joint prior is given as

$$\pi(\boldsymbol{\theta}, \omega_0|\mathscr{D}_0) \propto \frac{\prod_{i=1}^m (L(\boldsymbol{\theta}|\mathscr{D}_{0i})^{\omega_{0i}}\pi(\omega_{0i}))\,\pi(\boldsymbol{\theta})}{\int_\Theta \prod_{i=1}^m L(\boldsymbol{\theta}|\mathscr{D}_{0i})^{\omega_{0i}}\pi(\boldsymbol{\theta})d\boldsymbol{\theta}}\mathbf{1}_{\Omega_0}(\omega_0), \tag{2.12}$$

where $\Omega_0 = \{\omega_0 = (\omega_{01}, \omega_{02}, ..., \omega_{0m}) : 0 < \int_\Theta \prod_{i=1}^m L(\boldsymbol{\theta}|\mathscr{D}_{0i})^{\omega_{0i}}\pi(\boldsymbol{\theta})d\boldsymbol{\theta} < \infty\}$. The corresponding joint posterior distribution is given as

$$\pi(\boldsymbol{\theta}, \omega_0|\mathscr{D}, \mathscr{D}_0) \propto L(\boldsymbol{\theta}|\mathscr{D})\pi(\boldsymbol{\theta}, \omega_0|\mathscr{D}_0). \tag{2.13}$$

The marginal posterior distribution of $\boldsymbol{\theta}$ is obtained by integrating out ω_0 in (2.13). The marginal distribution of ω_0 is obtained in the same manner by integrating out $\boldsymbol{\theta}$ in the joint posterior distribution.

2.4.2 Commensurate Prior

Hobbs *et al.* (2011) proposed several extensions of the power prior for location and scale parameters, referred to as commensurate power prior distribution. Here, we focused on the

commensurate prior relevant to this work. The prior assumes different location parameters for the historical and current data, $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$, respectively. The prior of $\boldsymbol{\theta}$ depends on $\boldsymbol{\theta}_0$ through the parameter $\boldsymbol{\lambda}$, which accounts for the commensurability between $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}$. Assuming independent uniform prior for $\boldsymbol{\theta}_0$, the multi-dimensional commensurate power prior $\boldsymbol{\theta}$ follows as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\theta}_0, \boldsymbol{\lambda}, \omega_0 \mid \mathscr{D}_0) \propto \frac{L(\boldsymbol{\theta}_0 \mid \mathscr{D}_0)^{\omega_0}}{\int L(\boldsymbol{\theta}_0 \mid \mathscr{D}_0)^{\omega_0} d\boldsymbol{\theta}_0} \pi(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}) \pi(\omega_0) \pi(\boldsymbol{\lambda}), \tag{2.14}$$

where $\pi(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\theta} - \boldsymbol{\theta}_0)^T \boldsymbol{\lambda} (\boldsymbol{\theta} - \boldsymbol{\theta}_0)\right)$ is the commensurate prior distribution and $\pi(\omega_0) \propto \omega_0^{g(\boldsymbol{\lambda})-1}$, $g : \mathbb{R}^q \longrightarrow \mathbb{R}$ and $g(\boldsymbol{\lambda}) > 0$.

Hobbs *et al.* (2011) proposed a commensurate prior distribution for a univariate location parameter. A multi-dimensional equivalent commensurate prior is given as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\lambda} \mid \mathscr{D}_0) \propto \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\lambda}) \int \Big[L(\boldsymbol{\theta}_0 \mid \mathscr{D}_0) \pi(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}) \Big] d\boldsymbol{\theta}_0. \tag{2.15}$$

As $\boldsymbol{\lambda}$ approaches zero, the commensurate prior distribution approaches $\pi(\boldsymbol{\theta})$, and as the eigenvalues of $\boldsymbol{\lambda}$ approaches infinity, the commensurate prior distribution is a Bayes update of $\pi(\boldsymbol{\theta})$. In the same work, Hobbs and colleagues proposed location-scale commensurate prior distributions, which account for the commensurate between the location and scale parameters of the model on the current data and historical data. That is, the extension of (2.15) is by conditioning the scale parameter of the current model on the scale parameter of the historical model in a similar way $\boldsymbol{\theta}$ is conditioned on $\boldsymbol{\theta}_0$. Let $\boldsymbol{\gamma}$ be the scale parameter involved in the model on the current data and $\boldsymbol{\gamma}_0$ be the corresponding parameter in the historical data model. Then the location-scale commensurate prior distribution follows as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2 \mid \mathscr{D}_0) \propto \pi(\boldsymbol{\theta}) \pi(\boldsymbol{\gamma}) \pi(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2) \int \Big[L(\boldsymbol{\theta}_0, \boldsymbol{\gamma}_0 \mid D_0) \pi(\boldsymbol{\theta} \mid \boldsymbol{\theta}_0, \boldsymbol{\lambda}_1) \pi(\boldsymbol{\gamma} \mid \boldsymbol{\gamma}_0, \boldsymbol{\lambda}_2) \Big] d\boldsymbol{\gamma}_0 d\boldsymbol{\theta}_0, \tag{2.16}$$

with independent uniform prior distribution assumed for the historical parameters. $(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2)$ accounts for the location and scale commensurate. Moreover, the authors proposed a commensurate prior distribution formulated from a mixture of the prior distribution specified in (2.16).

2.5 Dirichlet process

Bayesian nonparametric (BNP) models are a class of probability models on infinite dimensional probability spaces. It relaxes the assumption of using a finite number of parameters to describe a random process in the parametric counterpart. It has been considered for density and intensity estimation (KOTTAS, 2006; RABAOUI *et al.*, 2011), model validation, clustering (TEH *et al.*, 2010), and anomaly detection (KALTSA *et al.*, 2018). Suppose a spatial process $Y(s) \sim H$ with family $\{H_{\boldsymbol{\theta}} : \boldsymbol{\theta} \in \mathbb{R}^d\}$, in the Bayesian parametric framework, a fixed distribution is assumed for H, a prior distribution is assigned to $\boldsymbol{\theta} \sim P(\boldsymbol{\theta})$, and the posterior distribution

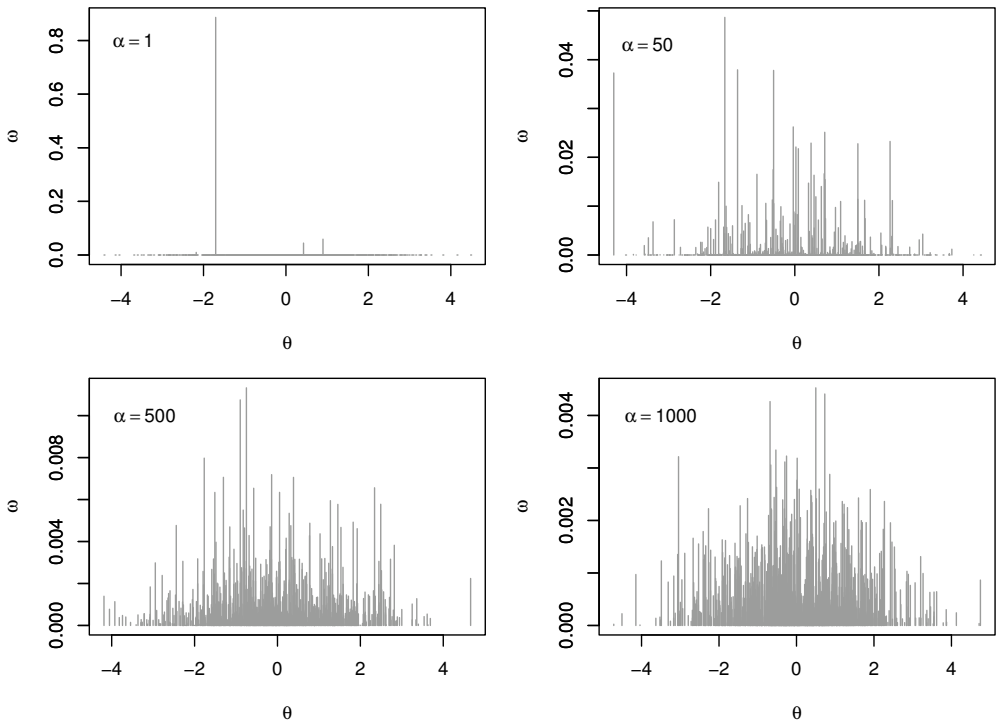


Figure 3 – A distribution of $H \sim DP(\alpha, N(0, 2^{1/2}))$ for different α simulated using the stick-breaking technique .

is obtained using Bayes theorem. In BNP, the assumption of fixed H is relaxed, and a prior distribution is assigned to $H \sim P(H)$ instead. A common choice of the prior on H is the Dirichlet process (DP) (FERGUSON, 1973; TEH *et al.*, 2010).

Suppose that $H \sim DP(\alpha, G)$, where G is the base distribution of the process over space $\mathbf{\Omega}$, and $\alpha > 0$ is the concentration parameter. For any partition $B_1, B_2, ..., B_r$ of $\mathbf{\Omega}$, the vector $(H(B_1), H(B_2), ..., H(B_r)) \sim \text{Dir}(\alpha G(B_1), \alpha G(B_2), ..., \alpha G(B_r))$, and *Dir* indicates a Dirichlet distribution. This implies that H itself is random and for any partition such that $\mathbf{\Omega} = \{B, B^c\}$, $\mathbb{E}(H(B)) = G(B)$ and $Var(H(B)) = G(B)(1 - G(B))/(\alpha + 1)$, indicating that the mean and variance are themselves random unlike in the parametric case, and thus, is able to better capture the process subtle patterns. In addition, adopting the DP prior distributions for spatial modeling can better capture nonstationary processes.

The distribution $H \sim DP(\alpha, G)$ is not simple to draw samples from. However, a single draw of H can be obtained through a stick-breaking process (SETHURAMAN, 1994). That is the distribution $DP(\alpha, G)$ can be represented as $\sum_{b=1}^{\infty} \omega_b \delta_{\theta^{(b)}}$. δ_a denotes a point mass at a and $\theta^{(b)} \sim G$. Let $\omega_1 = a_1 \sim Beta(1, \alpha)$ and $a_b \sim Beta(1, \alpha)$, $b = 2, 3, ...$ independently, then, the stick-breaking representation follows as $\omega_b = a_b \prod_{k=1}^{b-1}(1 - a_k)$. For example, suppose $G = N(0, \sqrt{2})$. Using the stick-breaking approach, Figure 3 shows the DP distribution for $\alpha = 1, 50, 500$, and 1000. The plot shows that the distribution approaches G as α increases, however, it is a discrete distribution.

The DP process can be used as prior distribution to estimate the probability distribution H that generates independent and identical variables $Y_1, ..., Y_m$. That is $Y_i \sim H$ and $H \sim DP(\alpha, G)$. The posterior distribution is again a DP process. That is $H \mid \mathbf{y} \sim DP(\alpha^\star, G^\star)$, where $\alpha^\star$ and $G^\star$ are obtained after some algebras. See Escobar and West (1995) for illustration. This idea was utilized in Chapter 6 to arrive at a nonstationary spatial process.

DP has been considered for solving spatial problems. For example, Gelfand, Kottas and MacEachern (2005) proposed a Dirichlet process mixture for spatial modeling and examined the theoretical properties, and applied it to a single dataset of precipitation measurements over the Languedoc-Roussillon region in France. Silva (2007) proposed a Dirichlet process mixture for the segmentation of single magnetic resonance images of the human brain. Moreover, Kim, Smyth and Stern (2010) proposed a hierarchical DP (hDP) for modeling multiple fMRI images to allow sharing of components across multiple images. In a similar vein, Gupta, Phung and Venkatesh (2012) proposed a hierarchical beta process for combining multiple data for transfer learning in text and image retrieval, and Liu, Wade and Bochkina (2022) proposed a similar model for clustering Single-cell RNA sequencing data. Specifically, suppose there are J related images or data with corresponding probability distribution H_j. Then, for replication t, the hierarchical DP follows as

$$\begin{aligned} Y_{jt} &\sim H_j, \\ H_j &\sim DP(\alpha_j, H_0), \\ H_0 &\sim DP(\alpha_0, G_0), t = 1, 2, ..., T_j, j = 1, 2, ..., J, \end{aligned} \tag{2.17}$$

where G_0 is the base prior distribution incorporating the shared information across multiple images. While the model is elegant and appealing for dealing with various shared effect problems, it subjects all Y_{jt} to the base prior distribution G_0, which may be inappropriate for data generated from different sources or exhibit more complex characteristics that have multiple latent sources of heterogeneity across multiple datasets. An example of such a dataset is the TMS data. This happens because the pattern of subjects' responses to TMS stimulation possesses a complex structure that has the tendency to be modulated at different unknown brain states as studies have shown intra-individual variability in a time scale of minutes (BERGMANN, 2018). This consequently leads to a higher variability across multiple subjects. In these cases, using the model 2.17 implies that all response Y_j, $j = 1, 2, ..., J$ originated from G_0. Thus, this project considered a more flexible model for the TMS dataset.

2.6 Data source

TMS is an active neurophysiological technique that causes electrical induction in the brain. The TMS pulses delivered from its coil form a magnetic field that painlessly passes through the scalp to reach and depolarise neurons in the brain region under investigation (ROTENBERG; HORVATH; PASCUAL-LEONE, 2014). This stimulation causes the activation of some biological

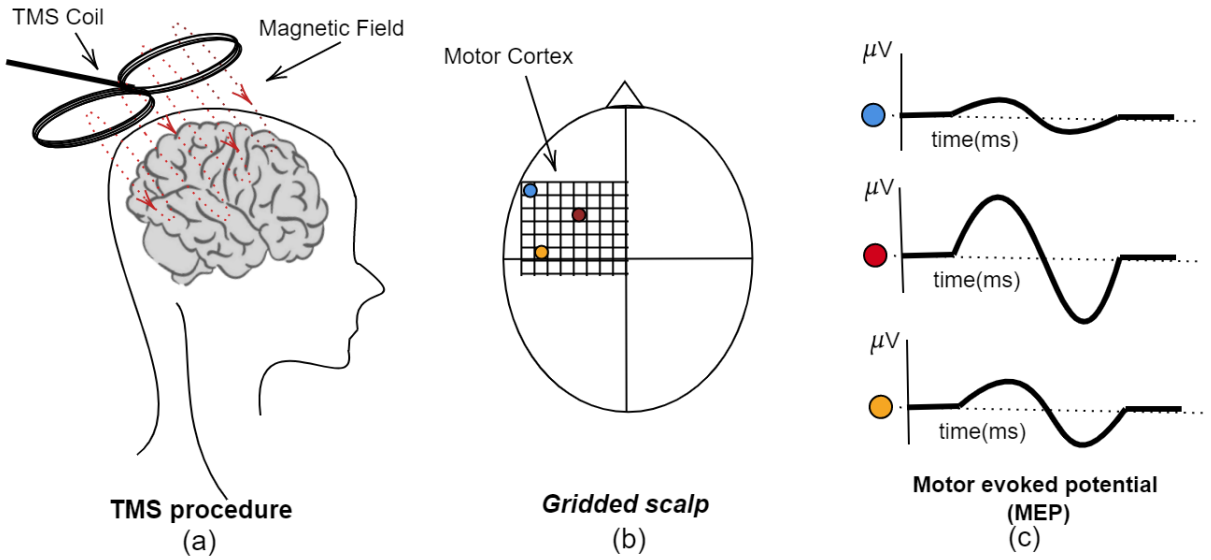


Figure 4 – Illustration of a TMS brain mapping procedure (a), a gridded motor area of the scalp where TMS pulses are delivered into the brain (b), and observed signal referred to as motor evoked potential (c).

neurons, and action potentials are released in the process. Figure 4 gives an overview of the data collection procedure used in this project. Figure 4a shows a subject being stimulated using a figure-of-eight TMS coil. The coil is placed over the scalp and TMS pulses in the form of a magnetic field are induced in the brain. In practice, a grid is formed over the subjects' scalp on the motor cortex (Figure 4b), and each grid cell is stimulated with the TMS pulse. The effect of the stimuli is recorded as a signal at the muscle of interest using an electromyograph. The recorded signal is known as motor-evoked potential (Figure 4c).

2.6.1 Experimental design and data acquisition

2.6.1.1 Subjects

The acquired data used in this study were approved by the ethics committee of St John of God Healthcare Human Research Ethics Committee (HREC, 1474) and reciprocal ethics approval has been provided by ECU HREC (2019-00023). Informed written consent was obtained from three right-handed subjects (male, aged 58, 58, and 49). For one of the subjects, data were acquired on two occasions.

2.6.1.2 Experimental setup

Each subject was seated comfortably in a chair, with the right hand pronated in a relaxed position. Subjects were instructed to keep their eyes open and to look straight ahead during stimulation. A Magstim 200^2 magnetic pulse stimulator and a 70 mm figure-of-eight coil was used. EMG was recorded with the BrainSight built-in EMG pod over the right first dorsal interosseous (FDI) muscle. Two pre-gelled, disposable $Ag/AgCl$ surface electrodes (Kendall

Medi-Trace mini electrodes) were positioned over the FDI muscle according to a belly-tendon montage, while a ground electrode was placed over the styloid process of the ulna bone. Before placing the EMG electrodes, the skin was scrubbed with an alcohol wipe to reduce impedance and improve signal quality. The data were sampled at 3 kHz and bandpass filtered with bandwidths set between 16 and 470 Hz. Data were collected on a Macintosh iMac computer running Brainsight®TMS neuronavigation software version 2.2.11 (Rogue Research, Montreal, Quebec).

Prior to the TMS experiment, the subject received a magnetic resonance image (MRI) scan (T1-weighted) anatomical image at 1 mm^3 resolution to acquire cranial MRI data (1.5T, Philips Ingenia, Philips Healthcare, Best, the Netherlands). From the MRI data, a model of the head surface was created in the neuronavigation software by co-registering the participant with their MRI scan.

2.6.2 TMS data collection

The position of the participant's head and TMS coil was constantly monitored in real-time with the Polaris Vicra Optical Tracking System (Northern Digital Inc.). This ensured that the center of the coil was kept within 2 mm of the target and that the coil orientation was consistent throughout the experiment. A grid (1cm spacing; 11x11) was centered over the hand's primary motor cortex (M1). This cortex was visually identified as it is predominantly located in the middle knee of the central sulcus (CS), formed by the posterior curvature of the precentral gyrus (PreCG) around the handknob (SILVA *et al.*, 2020). The TMS stimulator was set to 65 percent maximal stimulator output (MSO). This intensity was used to search for a motor response in the neighboring locations. If no motor response was elicited within 2 centimeters of the grid center, then the stimulator was turned up with 5 percent MSO. This procedure was repeated until a response was elicited. The MTAT 2.0 tool (AWISZUS, 2003) was used to determine the resting motor threshold (RMT) at the stimulation target. The stimulation intensity for the subsequent mapping was set to 110 percent of the RMT. The starting point for the mapping was the center of the grid (0,0). The figure-of-eight coil was held tangentially to the scalp and rotated away from the midline by 45 degrees, in order to activate the corticospinal system preferentially trans synaptically. The coil was moved forward and backward to collect 10 pulses at each location. Movement direction was changed when no motor-evoked potential (MEP) could be collected anymore, that is no responses $\geq$ 0.05mV. Online visual feedback of the coil placement, by the neuronavigation system, with its relation to the brain-enabled real-time coil adjustment.

2.6.3 Descriptive statistics

Figure 5a & b show the MEP signal recorded by EMG at different locations over the motor cortex for subject 1 at the first and second TMS sessions respectively. Figure 5c &d are the MEP signals for subjects 2 and 3. For all the subjects, the signal is flat for a period of time until it attains its lowest and highest peaks in a short time interval and levels off thereafter.

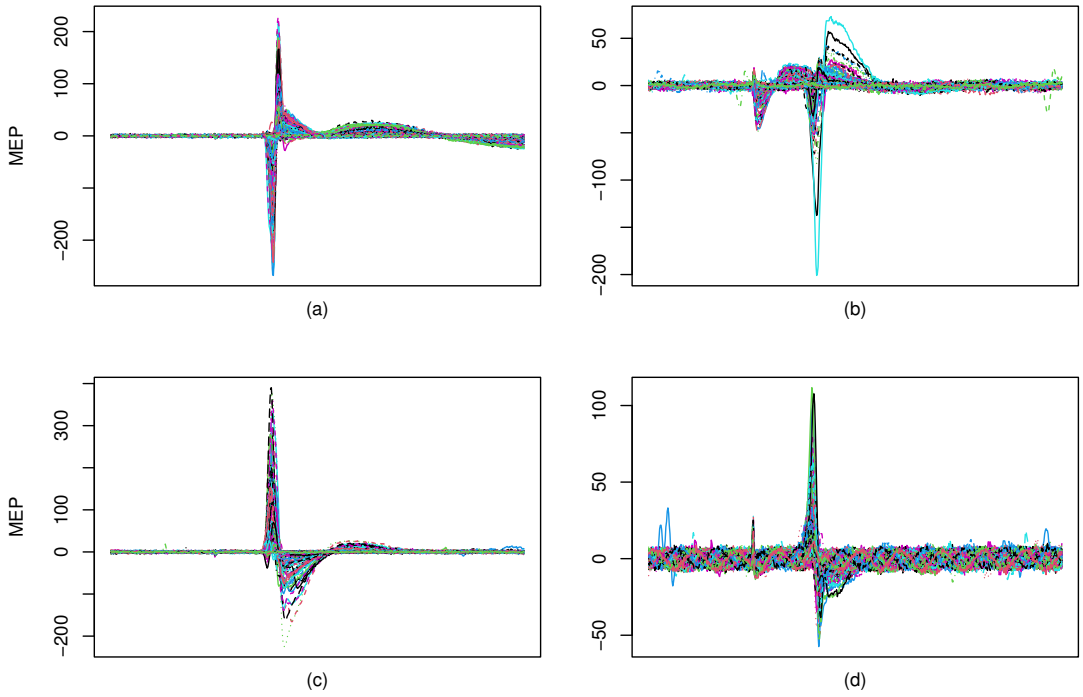


Figure 5 – Observed motor-evoked potential signal for subjects (a) 1R1, (b) 1R2, (c) 2, and (4) 3. Each curve represents a motor-evoked potential at a given spatial location within the motor cortex.

Table 1 shows the mean, median, 10th percentile, and 70th percentile of the peak-to-peak of the MEP signal shown in Figure 5. Technically, suppose the upper peak (maximum value) of a signal in question is denoted by a and the lower peak (minimum value) is denoted by b. The peak-to-peak of the signal equals the difference between a and b. Here, 1R1 and 1R2 denote subject 1 at TMS sessions 1 and 2 respectively. From the table, the acquired data show that subject 1R1 had the largest variation in the peak-to-peak MEP across different locations on the motor cortex since the mean, median, and percentiles are farthest apart.

Table 1 – Descriptive quantities of peak-to-peak MEP (μV).

Subject	Mean	Median	10%	70%
1R1	100.95	37.37	6.79	133.38
1R2	36.39	28.31	6.79	47.56
2	71.58	15.85	6.79	64.54
3	38.59	26.04	12.46	47.78

Figure 6 shows the spatial distribution of the peak-to-peak MEP for all the subjects. The figure shows how locations with high and low peak-to-peak are clustered over the motor cortex. It also shows that the coordinates where data were obtained for all the subjects are different, which could pose data integration challenges. Figure 7 shows the box plot of the log peak-to-peak MEP grouped over cells referred to as Associate targets. The box plot shows the median, 25th, and 75th percentiles of the log peak-to-peak MEP. Similarly, the plot revealed some variations across subjects and Associate targets.

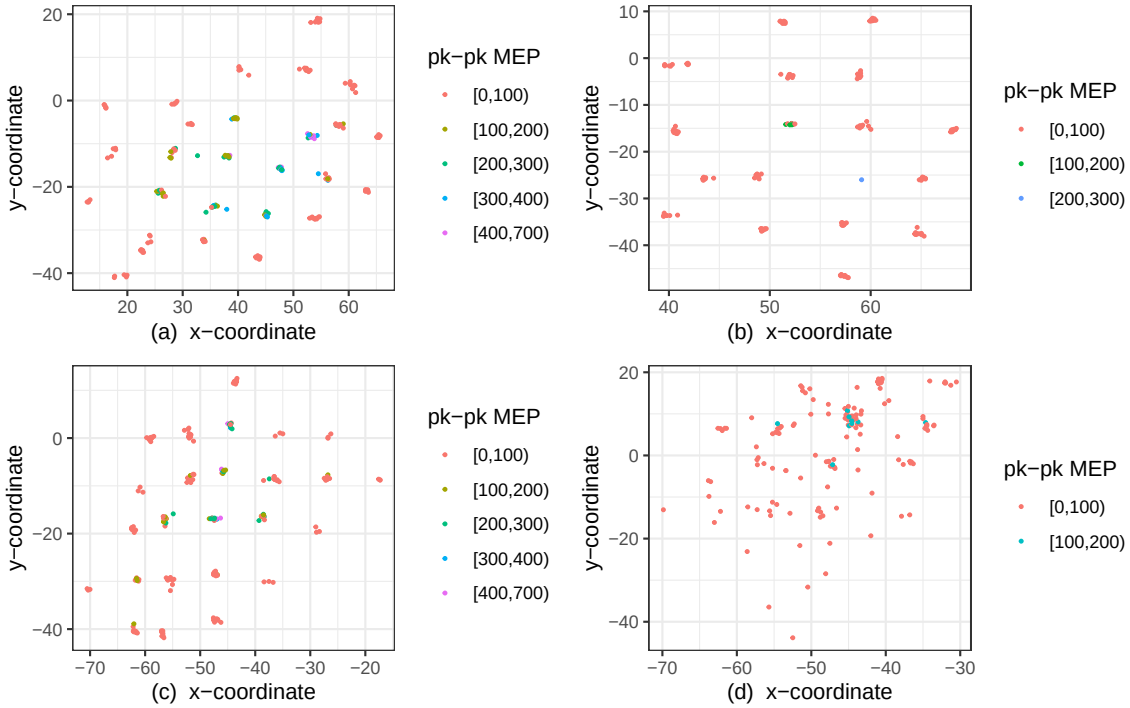


Figure 6 – The spatial distribution of the peak-to-peak motor evoked potential for subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3.

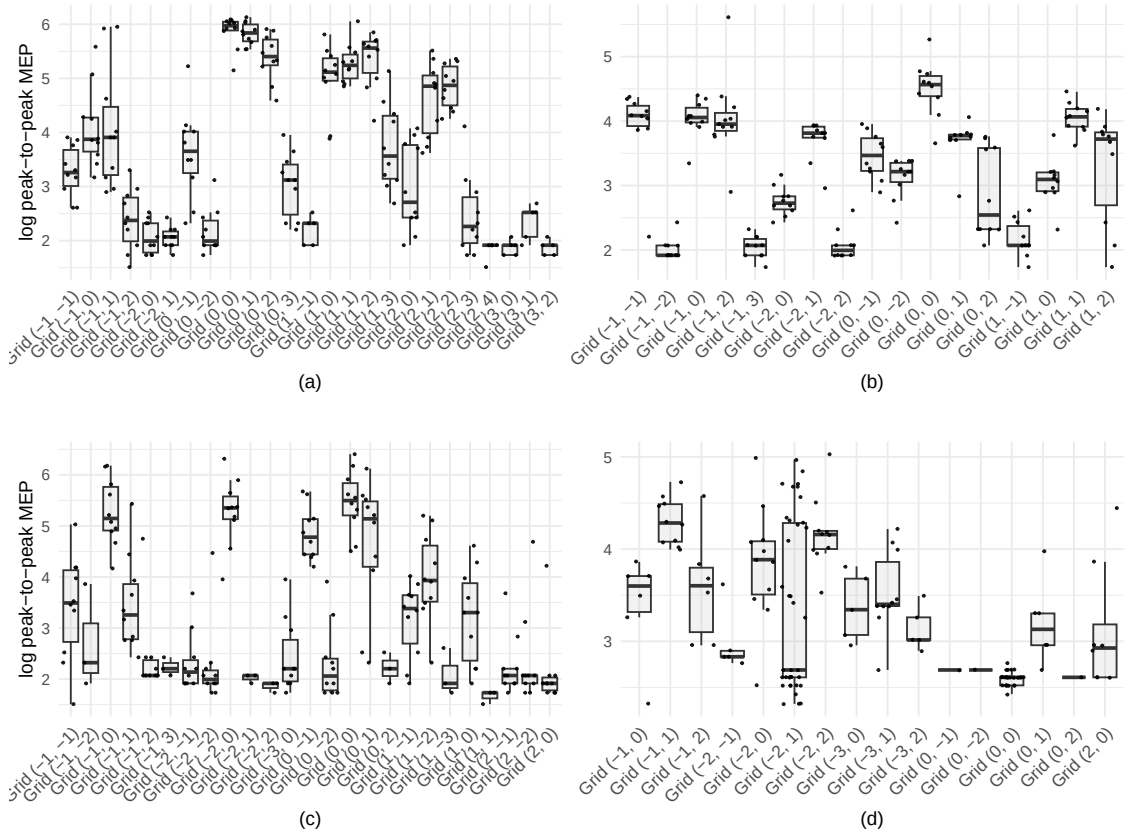


Figure 7 – The boxplot of the log peak-to-peak motor evoked potential grouped by the associate targets for subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3. Note that each point in the figure denotes an average peak-to-peak MEP for a given spatial location.

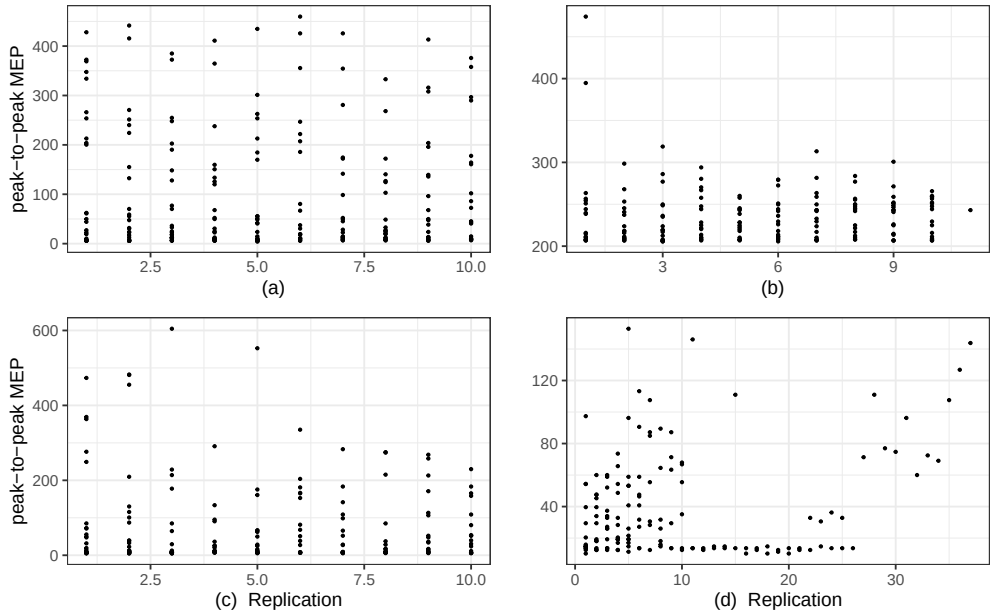


Figure 8 – The peak-to-peak motor evoked potential against the consecutive TMS pulses replication for subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3.

Figure 8 shows the peak-to-peak MEP with respect to the consecutive TMS pulses delivered over the motor cortex. The consecutive TMS pulses delivered were referred to as replication. For example, the peak-to-peak MEP at replication 3 at a given coordinate of the motor cortex is the response recorded by the EMG at the third consecutive TMS pulse induced in the brain. The figure shows variations across replications.

CHAPTER

3

BAYESIAN SPATIAL STATISTICS: A SYSTEMATIC REVIEW AND META ANALYSIS

This chapter presents the method, results, and findings of a systematic review and meta-analysis of the literature on Bayesian spatial modeling. This analysis aimed to assess the current state of development of these spatial models, identify trends, and highlight their contributions to the Bayesian spatial modeling literature. Furthermore, this chapter aims to determine the variation in the application of Bayesian spatial statistics across different fields and identify the state-of-the-art spatial models that could be useful for developing web applications for Bayesian spatial analysis of the TMS data.

3.1 Methodology

The collection and reporting methods were based on the guidelines of the Preferred Reporting Item for Systematic Review and Meta-Analysis (PRISMA) (HUTTON *et al.*, 2015; MOHER *et al.*, 2009). This procedure includes an electronic search strategy, a clear objective to define the inclusion and exclusion criterion, and an appropriate method for reporting the findings.

An online electronic search was conducted on June 10, 2020, in the following four databases: Elsevier's Scopus, Science Direct, Thompsom Reuters Web of Science, and the American Mathematical Society's MathSciNet database. Queries of the word "Bayesian Spatial" and "Bayesian spatial", using the Boolean operator "OR", throughout 2001-2020. Title, abstract, and Keywords were used in Scopus and Science Direct, the topic (which entails title, abstract, and keywords) in Web of Science, and "Anywhere" in MathSciNet.

The Mendeley Windows application was used to remove duplicated articles. The resulting set was further examined manually looking for more duplicates not identified by Mendeley's

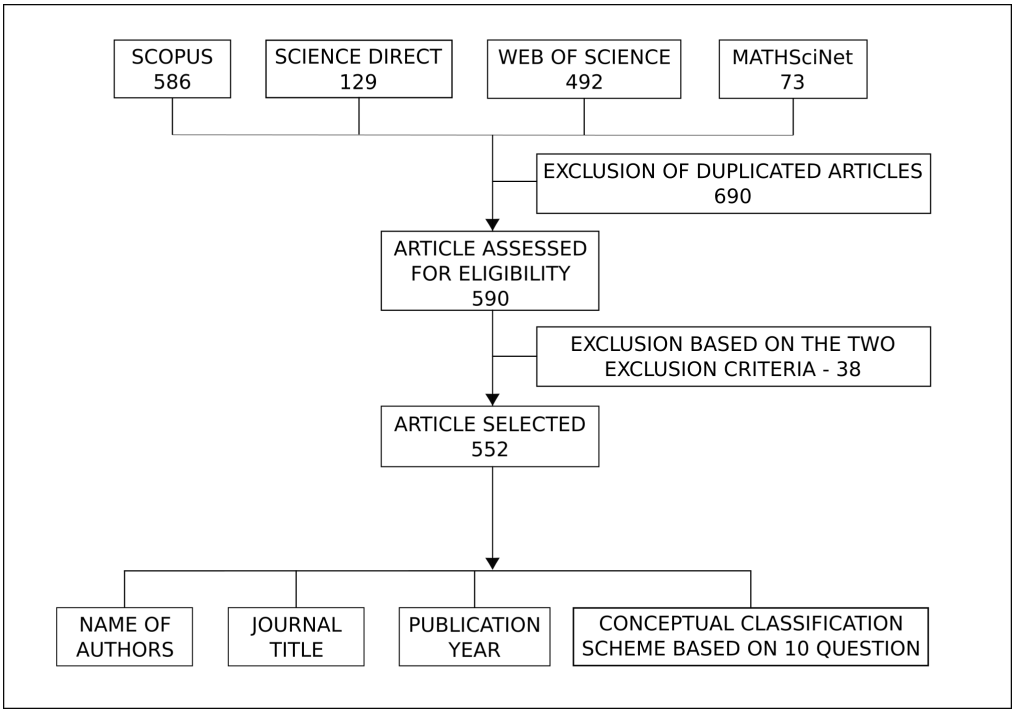


Figure 9 – Flow chart of systematic review search procedure in the Scopus, Science Direct, Web of Science, and MathSciNet databases. From 1,280 articles, based on the query words, 728 articles were removed (duplicated papers, non-English written, not peer-reviewed, nor Bayesian spatial modeling), and 552 remained to be analyzed. Then, information such as authors' names, journal titles, publication years, and the conceptual classification scheme was explored.

application. The titles and abstracts of the articles included (after removing duplicates) were first screened for Bayesian spatial methodology before applying the following inclusion criteria.

i. Search results that are written in English, and articles published in peer-reviewed journals available online. Books, dissertations/thesis, conference proceedings, and reviews (or any other form that was not an article) were excluded.

ii. Articles that specifically implement Bayesian spatial models excluding the ones that only mentioned Bayesian spatial models.

Articles that did not meet the two inclusion criteria were excluded from the review. The search flow chart is presented in Figure 9. Using the search keywords mentioned earlier, 586 articles were retrieved from Scopus, 129 from Science Direct, 492 from Web of Science, and 73 from MathSciNet. After excluding duplicated ones, 590 articles were assessed for eligibility, and 38 were further excluded based on the two exclusion criteria, leaving 552 articles selected for conceptual classification. These articles were classified into the following categories: Names of all authors, Publication year, Journal title, and response to the ten items of the conceptual classification scheme on Bayesian spatial models shown in Appendix 7.3.

3.2 Conceptual Scheme for Spatial Models

A conventional approach to content analysis (HSIEH; SHANNON, 2005) was adopted by scrutinizing samples of the articles to clearly define the characteristics that better explain the scope and richness of the literature and identify the key concepts and patterns. The first step was to determine these characteristics. New characteristics were added whenever new data that did not fit in already defined characteristics were found. This approach makes room for the literature to be classified without a priori presumption. The following subsections describe the major characteristics considered.

3.2.1 Spatial Statistics Fields of Application

Bayesian spatial statistics is a useful tool to incorporate the dependence structure and hidden patterns over space, through prior knowledge and data likelihood. In some cases, the hypotheses of interest in a random phenomenon do not directly relate to the effect of spatial dependencies. However, it is crucial to adjust for spatial variation to reduce bias (RILEY *et al.*, 2015). Adjustment for spatial patterns in modeling random occurrence has been practiced across various fields such as Agriculture, Medicine, Biology, Epidemiology, Geography, Geology, Economics, Climatology, and Ecology, among others (KARIMI; MOHAMMADZADEH *et al.*, 2012). Moreover, spatial dependence in the Agriculture experiment has long received consideration. Ronald A. Fisher identified spatial variations and used them to establish (random) blocks in the experiments to mitigate the effect of spatial dependencies in a randomized experimental design (FISHER, 1936).

In many Biological and Medical experiments, such as gene classification and brain mapping, the randomized blocking technique may not be a viable alternative. Moreover, in demography, disease mapping, image analysis, remote sensing, manufacturing engineering, and species detection, the variation due to spatial proximity cannot be neglected. It may result in bias and inconsistent estimates. Responses at close range tend to have similar behavior and variation. The homogeneity of the variation depreciates with the increased distance apart. An efficient procedure to tackle the effect of spatial proximity is to consider random field Statistical models. Random field Statistical models, known as spatial models, describe the distribution of a random phenomenon over a spatial domain.

Spatial models have long been applied in various fields. In 1949, Isard described the general theory of the spatial formation of economic activities focusing on the geographic distribution of costs, prices, and location of industries (ISARD, 1949). Spatial statistics applied to economics, often referred to as spatial econometrics, have gained more attention in recent years to analyze economic data over a wide range of spatial domains (SPARKS; SPARKS; CAMPBELL, 2013). Similarly, in 1950, DA. Krige took advantage of nearby variations to pursue the spatial prediction of gold distribution in South Africa, basing predictions practically on

lognormal-de Wijsian spatial models (KRIGE; KRIGE, 1981). In Epidemiology and Public Health, spatial statistics have gained increasing importance in predicting disease outbreaks (GRACIA *et al.*, 2015; LUAN; LAW; LYSY, 2018; MORRIS, 2015; MÜLLER; BETUELA; HIDE, 2002; SHORT; CARLIN; BUSHHOUSE, 2002). The problems that arose in these fields usually motivated the improvement of existing spatial models in the literature. Examples are models to alleviate spatial confounding and misalignment in spatial data, determine spatial risk factors, and disease surveillance (WARD, 2008; BANERJEE; CARLIN; GELFAND, 2014; AZEVEDO; PRATES; BANDYOPADHYAY, 2022; GONÇALVES; AGUILAR; PRATES, 2022).

In this review, the application fields were classified into five major groups. 1. Biological and Medicine: these include research on Biology, Medicine, Epidemiology, and Public Health, 2. Economics and Humanity: these include Economics, Demography, Criminology, and Accident Analysis. 3. Physical Science and Engineering, 4. Agricultural and Environmental Science, and 5. Sport.

Spatial statistical models play a key role in determining the spatial pattern or quantifying the relative positions of biological components, such as DNA, involved in some biological functions. Some examples include the study of the relative positioning of primordial and growing follicles in mice to identify the likely source of some regulatory muscles (SILVA-BUTTKUS *et al.*, 2009); determine the spatial patterns, relative position, and interaction of Arabidopsis thaliana heterochromatin (ARPÒN *et al.*, 2021); molecular profiling using Bayesian hierarchical Negative Binomial distribution for diagnoses and treatment procedures (LI *et al.*, 2021). Moreover, it is useful for disease mapping to determine the onset of an epidemic disease. A recent application includes the analysis of the mortality rate of COVID-19 in Spain and Italy using a Gaussian process to explain the spatial pattern (ADEGBOYE *et al.*, 2021; SAAVEDRA *et al.*, 2021); the spatial mapping of schistosomiasis in Tanzania to determine the prevalence and spatial pattern (CLEMENTS *et al.*, 2006). It is also considered to be a powerful tool for image analysis in the Medical field. An example includes the multivariate spatial model for characterizing neuroimaging data with a linear combination of multiscale basis functions to explore traits or symptoms in brain disorder (HUERTAS *et al.*, 2017).

Bayesian spatial statistics has been embraced in Economics to identify the spatial pattern of a household's share of economic distress, to understand the formation of new business, and in studies of consumer and producer behavior. It has been used to identify the impact of economic, social, and demographic factors on spatial variability of household share of economic distress (BENASSI; NACCARATO, 2017); identify the spatial structure of the calls to Portuguese health line, accounting for demographic, socio-economic information, and characteristics of the health systems (SIMÕES *et al.*, 2017); identify clustering in severe mobility crash risk, diagnosing of active transportation safety issues (OSAMA; SAYED, 2019). Moreover, it has been employed in the analysis of spatial patterns and hotspot detection of violent and property crimes at a small

spatial scale in Toronto, Canada (LAW; QUICK; JADAVJI, 2020); map main features of fertility, such as timing, pace, and scale, and to detect spatial disparity in fertility transition in Brazil (POTTER *et al.*, 2010).

Moreover, Spatial Statistics has been increasingly applied in Agricultural, Physical, and Environmental Sciences. It has been used to provide estimates for the curvature of a railway sleeper supported on compacted ballast, through the multi-output Gaussian process to guide inference in unobserved regions (GREGORY *et al.*, 2019); to quantify the uncertainty, in spatially varying material parameters, such as polycrystalline, through a Gaussian random field (RAPPEL *et al.*, 2019; KOUTSOURELAKIS, 2009), and to study material properties and spatial variability in elastostatics (KOUTSOURELAKIS, 2012). It has been used in extreme value analysis to quantify the uncertainty associated with an increased risk of flooding in Great Britain (SHARKEY; WINTER, 2019); to determine the spatial pattern of the association of socioeconomic factors to Japanese Encephalitis; to understand the seasonal effect and spatial variability in yield maps in farm precision in South-Eastern Australia (ROBINSON *et al.*, 2009); and in sport, it has been used to determine the expected number of scores in a Golf game (YOUSEFI; SWARTZ, 2013).

Spatial models have a long history in image analysis (WOODS, 1972; BESAG, 1974). Recently, it has been gaining attention in the machine learning community for image processing (EL-BAZ; FARAG, 2003; SIDÉN; LINDSTEN, 2020). For instance, Per Siden and Fredrik Lindsten established a connection between Convolutional Neural Networks (CNN) and a Gaussian Markov Random Field, which the authors applied to temperature data (SIDÉN; LINDSTEN, 2020). Vemulapalli et al. (VEMULAPALLI; TUZEL; LIU, 2016) proposed a deep network architecture based on a Gaussian conditional random field for image denoising. Moreover, Lee et al. (LEE *et al.*, 2017) developed an exact equivalence of infinitely wide neural networks and Gaussian processes and further linked the performance of these Gaussian processes to the theory of signal propagation in random neural networks.

3.2.2 Spatial data

Geographically reference data, also known as spatial data, is a collection of a realization of a stochastic process indexed by space. In other words, suppose $y(s)$ is a random process observed at location s, the set $Y(s) \equiv \{y(s), s \in \mathscr{D}\}$ is a spatial data, in which $\mathscr{D}$, a subset of $\mathbb{R}^d$, is usually (but not necessarily) fixed and represents a spatial domain. According to Blangiardo and Cameletti (2015), the spatial data are distinguished as follows:

1 Area or Lattice data: it is a simple way to represent spatial data in the domain $\mathscr{D}$. In this type of spatial domain, $y(s)$ is a random aggregated realization across an area s of distinct boundaries. For area data, the boundaries are irregular, such as administrative divisions, whereas, for lattice, the boundaries are a regular division of $\mathscr{D}$. For simplicity purposes, it

may be necessary to aggregate other types of Spatial domain realizations to form area or lattice data. This process may sometimes be referred to as a discretization of $\mathscr{D}$.

2 Geostatistical or Point-reference data: $y(s)$ is a realization at a specific location s in a continuous spatial domain $\mathscr{D}$. Location s is considered to be a coordinate made up of longitudes and latitudes, and sometimes includes altitudes. Location s could also be represented in Cartesian coordinates.

3 Spatial point pattern: Realization $y(s)$ represents the occurrence or non-occurrence of an event at location s. In this case, the location itself is considered to be random. The realization is a location indicator of the presence or absence of a phenomenon of interest in the domain $\mathscr{D}$. In Agriculture, for example, the interest may be the distribution of a specific tree species, in which each realization is the presence or absence of the tree species in domain $\mathscr{D}$. In epidemiology, the realization may be the house address of a patient that has a particular disease (BANERJEE; CARLIN; GELFAND, 2014; CRESSIE, 2015).

For instance, Sharkey and Winter (2019) adopted a Bayesian spatial model on lattice data to identify patterns for the risk of flooding in Britain; Munoz *et al.* (2013) adopted stochastic partial differential equation model on a geostatistical data to predict the spatial occurrence of fish species, and Leininger and Gelfand (2017) proposed a Bayesian technique to estimate a spatial point pattern of American sweetgum trees and Swedish pines.

3.2.3 Spatial Models

Several spatial models for spatially indexed response variable modeling or prior modeling of spatial effects were found in the literature, and most were categorized in the class of the Gaussian Markov Random Field (GMRF), defined as a Gaussian random field with Markov property (CRESSIE, 2015; RUE; TJELMELAND, 2002). The choice of spatial model is strongly informed by the study objective and available information. For modeling area spatial data, for example, the Conditional Autoregressive class of spatial model is a popular choice to account for spatially indexed county heterogeneity, whereas the Stochastic partial differential equation is popular for geostatistical data.

The classification of the spatial prior models was according to the popularity in the literature. These prior models were classified as *Conditional Autoregressive (CAR) family, Stochastic Partial Differential Equation (SPDE), GMRF (none of the above), Non-GMRF, Non-parametric, Article-specific methodology (none of the above), and Not stated*. The CAR family includes the Beseg York Mollie (BYM), Leroux's, Dean's, and Simpson's CAR models, including the multivariate versions. The GMRF class consists of GMRF priors models except for the CAR family and the SPDE. Table 2 shows a summary of the models found in the literature.

Table 2 – Summary of some Spatial Models and their variations.

		Spatial Smoothing		Gaussian Process		Non-Gaussian Process		
Spatial Model	ARTICLE	Global	Local	GMRF	Non-GMRF	Parametric	Semi-Parametric	Non-Parametric
Proper CAR	Besag (BESAG, 1974)	✓		✓				
Intrinsic CAR/BYM	Besag *et al.*(BESAG; YORK; MOLLIÉ, 1991)	✓		✓				
Spatial Partition Model	Leonhard & Raßer (KNORR-HELD; RASSER, 2000)		✓					✓
Asymmetric Laplace	Kuzobowski & Pogorski (KOZUBOWSKI; PODGÓRSKI, 2000)	✓				✓		
Leroux CAR	Leroux *et al.* (LEROUX; LEI; BRESLOW, 2000)	✓		✓				
Dean's CAR	Dean *et al* (DEAN; UGARTE; MILITINO, 2001)	✓		✓				
Hidden Markov Model	Green & Richardson (GREEN; RICHARDSON, 2002)	✓						
Dirichlet Process	Gelfand *et al.* (GELFAND; KOTTAS; MACEACHERN, 2005)	✓						✓
Skewed normal	Arellano *et al.* (ARELLANO-VALLE; AZZALINI, 2006)	✓			✓			
Globalspline	Lee & Durbán (LEE; DURBÁN, 2009)	✓					✓	
SPDE	Lindgren *et al.* (LINDGREN; RUE; LINDSTRÖM, 2011)	✓		✓				
CAR dissimilarity	Lee & Mitchell, (LEE; MITCHELL, 2012)		✓	✓				
Copula-based	Pilz *et al.* (PILZ; KAZIANKA; SPÖCK, 2012)	✓				✓		
TGMRF	Prates *et al.* (PRATES *et al.*, 2015)	✓		✓				
Simpson CAR	Simpson *et al* (SIMPSON *et al.*, 2017)	✓		✓				
Log-Gamma	Bradley *et al.* (BRADLEY; HOLAN; WIKLE, 2018)	✓				✓		
Hedonic spatial field	Oliveira & Ecker (OLIVEIRA; ECKER, 2022)	✓				✓		
POGAMP	Gonçalves *et al.* (GONÇALVES; AGUILAR; PRATES, 2022)	✓				✓		
Student-t	Ordonez *et al.* (ORDOÑEZ *et al.*, 2023)	✓				✓		

The non-GMRF is a large class that consists of non-trivial prior models that provides some special spatial correlation function to determine the dependency of a spatial process in a continuous space, including the Asymmetric Laplace, Log-Gamma, skewed normal, Student-t process, and Dirichlet process, and Poisson-Gaussian Mixture Process (POGAM). A class *Others* was created to accommodate unspecified models and those that do not belong to the aforementioned classes. For instance, Obaromi (2019) and Egbon, Somo-Aina and Gayawan (2021) adopted ICAR and BYM models to map the spatial pattern of tuberculosis in South Africa and malnutrition in Nigeria; Fontanella *et al.* (2015) adopted a Bayesian hierarchical spatial quantile regression model with asymmetric Laplace spatial component to determine the risk factors of Radon-222 noble gas, which arises naturally from uranium decays, and Munoz *et al.* (2013) adopted SPDE model to predict the spatial occurrence of fish species.

The hyperparameters of the prior distribution for stochastic processes are either elicited (WALKER; CURTIS, 2014; MOALA; O'HAGAN, 2010), drawn from the previous study, assigned weakly-informative or objective prior distribution (ORDOÑEZ *et al.*, 2023). Moreover, Simpson *et al.* (2017) developed a framework to construct informative priors called penalized complexity (PC) prior. The PC prior has parameters that are subjectively set to control the amount of flexibility allowed in the model. The construction of PC prior for the BYM spatial model and Student-t process is described in Simpson *et al.* (2017). The PC prior for hyperparameters of some GMRF models is implemented in `INLA R` package (RUE; MARTINO; CHOPIN, 2009).

3.2.4 Data model and methods of prior specification

A response variable is a quantity used to describe a random process to mathematically relate it to a deterministic process. In statistical modeling, the most frequently used response variables are the discrete (categorical), ordinal, and continuous variables. The type of variables used in modeling a random phenomenon is intuitive from the process under study. The statistical models used to describe a random phenomenon vary depending on the quantity and parameters of interest.

The Bernoulli distribution is often used for modeling the random phenomenon of two possible outcomes. The Binomial, Negative Binomial, Hypergeometric, and Poisson distributions are frequently used for modeling count cases such as disease occurrence, wildlife, signal, and more (LI *et al.*, 2021; STAUBACH *et al.*, 2002). The Poisson distribution has been used to approximate Binomial distribution for a large sample size (TEERAPABOLARN; JAIOUN, 2014; BURR, 1973). The equality of mean to variance restriction imposed by the Poisson distribution considers the Negative Binomial a better choice to model a random variable that exhibits over-dispersion (ALEXANDER; MOYEED; STANDER, 2000). The Multinomial distribution is often used to model a phenomenon naturally occurring in more than two categories usually encountered in Biological experiments (KRISZTIN; PIRIBAUER; WÖGERER, 2021).

In the continuous case, a large class of distribution of the exponential family is used, such as Gaussian, Exponential, Student t, Weibull, Gamma, and more. However, according to the Central limit theorem, the Gaussian distribution could be used to approximate continuous distributions for large sample sizes (KWAK; KIM, 2017).

In practice, analysts seek to quantify the association of a random phenomenon and a set of explanatory variables, in combination with spatial processes. In the literature, the statistical models encountered are the *Generalized Linear Mixed Model (GLMM)*, *Non-parametric*, *Survival model*, and *Spatial Econometrics models*. Additionally, the classes *Proposed*, *Unspecified* and *Others* were created to accommodate non-standard proposed and validated models, as well as unspecified models. The class of *Others* accommodates statistical models outside the above-listed classes. For instance, Fontanella *et al.* (2015) adopted a generalized hierarchical mixed model to determine the risk factors of Radon-222 noble gas; Egbon, Somo-Aina and Gayawan (2021) used a generalized hierarchical mixed model to determine the impact of Caborn (IV) oxide on the prevalence of malnutrition; Aswi *et al.* (2020) adopted survival statistical model to map the prevalence of Hospitalisation due to Dengue in Wahidin Hospital in Makassar, Indonesia and Simões *et al.* (2017) adopted spatial econometrics model (lag-model) to estimate the global spatial correlation of the calls to Portuguese national health line.

An appropriate prior distribution specification in a Bayesian inference continues to be a challenge in various fields of application. A prior distribution is associated with the representation of uncertainty of the interest parameters before data are observed. The elicitation of an appropriate prior distribution is a non-trivial task (PALACIOS; STEEL, 2006), and such challenges are accumulated in spatial models due to a large number of associated parameters. In this review, four main approaches were encountered. One of these approaches is the assumption of a *flat, vague, or noninformative priors*, which allows the data model to carry all the information. Such an approach is not always advantageous because inference on the parameters can be improved by performing prior elicitation based on identified characteristics or expert opinion. The elicitation procedure is termed elicited prior, which leads to another method, known as *prior elicitation*. Convenient prior distributions are sometimes a choice and have spread across the

literature and have been set as default priors in most simulation packages. As a result, subsequent authors use such prior distribution *verbatim*. However, several authors did not explicitly state the type of prior used and were classified as not available. For instance, Kang *et al.* (2011) elicitated prior information from expert opinion, from a meta-analysis of neuroimaging data, and parapsychologist to perform a Bayesian spatial point process to provide an interpretable model for brain imagining studies and Kibria *et al.* (2002) elicited prior knowledge from experience for predicting a particular matter.

3.2.5 Computational Techniques

Assessing the posterior distribution in Bayesian analysis to make inferences is not a trivial task, because it usually contains compound integrands with complicated and analytically intractable distributions and supports (FRAGOSO; BERTOLI; LOUZADA, 2018). Thus, the authors explore different computational approaches to make inferences. In the literature, several computation techniques were found and classified into the *Markov Chain Monte Carlo (MCMC), Integrated Nested Laplace Approximation (INLA), Expectation-Maximization (EM) and Maximum (Penalized quasi) Likelihood Method classes*. The MCMC class comprises all numerical approximation that uses the Monte Carlo method. Moreover, the unspecified class was added to accommodate articles that neither discuss nor state the approach used in the estimation procedure. The *Others* class comprises computation techniques that do not fit into the defined classes. For instance, Egbon, Somo-Aina and Gayawan (2021) utilized the INLA interface for estimating the parameters of a Gaussian latent field model; Fontanella *et al.* (2015) utilized the MCMC approach to estimate quantile regression parameters; Sun *et al.* (1998) utilized the EM algorithm to perform a Bayesian interpolation of Nitrogen dioxide, ozone, sulfur dioxide, and surface iron in public health units in Ontario, and Neill, Moore and Cooper (2005) compared the performance of maximum likelihood estimation with different computation techniques for spatial scan statistic.

3.2.6 Simulation Study and Validation

A simulation study is a systematic and scientific computer procedure that involves fixing model parameters to generate data by pseudo-random sampling (MORRIS; WHITE; CROWTHER, 2019). It comprises two main steps: data generation and estimation. In the first step, a set of parameters is fixed and used to generate pseudo-random data. In the second step, the generated data are fed back to the model to estimate the "unknown" parameters and check for adequacy. A simulation study is usually carried out for proposed models and methods. The papers reviewed were classified into two: “Yes” if the paper contains statistical simulation studies, and “No” if it does not.

In addition to the simulation studies, this review also investigated how Bayesian spatial models were validated using real data. It is a procedure to check overfitting or underfitting. A

model overfits if it performs well in the training set and badly in the test set, whereas it under fits if it performs poorly in the training set. A classical approach to cross-validation is to form a disjoint subset of whole data into training and testing sets. The model is fitted into the former and tested on the later set. Doing this process k times until all observations in the dataset participate in training and testing once is called *K-fold cross-validation*. The whole data of size n is split into k disjoint subsets, in which the combined $k-1$ sets serve as the training set, and the remaining set of size n/k serves as the test set. A particular case to the k-fold is the Leave-One-Out cross-validation, in which one observation serves as the test set, and the remainder $n-1$ serves as the training set. After going through all subsets, the validation measures are statistically combined to make a valid conclusion.

Since the spatial models are frequently modeled in a Bayesian framework, the class of *Posterior predictive check* (GELMAN; MENG; STERN, 1996) was included. In a predictive posterior check, a statistical test is chosen and computed for the observed data process. The same statistic is computed for replicated posterior predictions of the process. The model is said to present a good fit if the posterior prediction average is close to the statistic test for the observed data (FRAGOSO; BERTOLI; LOUZADA, 2018). A *None or not applicable* class was included to accommodate papers that did not conduct cross-validation and *Others* to accommodate validation methods not mentioned above. For instance, Sun *et al.* (1998) adopted leave-one-out spatial location to validate the performance of the proposed model for multivariate interpolation of air pollutant gases in the health unit of Ontario and Akseer *et al.* (2018) utilized the k-fold cross-validation method to examine the robustness of the adopted model to determine geographical inequalities and nutritional status of women and children in Afghanistan.

3.3 Result

As described in the search procedure section, a total of 552 articles were selected after applying the exclusion criteria (duplication and context). After careful analysis, the papers were categorized into applied papers, theoretical, or both, in which 4 (1%) of the papers showed no application (only theoretical with synthetic data), 188 (34%) showed an improvement in the field with real-world application and 360 (65%) only applied the existing methodologies.

The result further showed that only 28.1% of the articles validated adopted models through a simulation study. In addition, 4.5%, 3.0%, and 11.9% of the articles used the K-fold, LOOCV, and Posterior predictive check real data-driven validation methods respectively. The result shows overall low model validation rates. This is consistent since a large percentage of the reviewed articles applied existing methods that are probably already validated.

The articles were subdivided into five classes of application fields: *Agricultural and Environmental Science*, *Economics and Humanities*, *Medical Science*, *Physical Science and Engineering*, and *Other*. Three fields hold the majority of the publications, which are *Agricultural*

and Environmental Science (30.1%), *Economics and Humanities* (30.6%), and *Medical Science*, which includes epidemiology, (33.7%). Moreover, the spatial domain used was also taken into account. The *Area/lattice*, *Geostatistical*, and *Spatial Point Patterns*. The *Area/lattice* occurred at 65.6% and *Geostatistical* at 31.2% of the reviewed papers, and in combination, they hold 95.8% of the publications. It is important to note that more than one spatial domain could be used in an article, such as the 1% observed in this review. This procedure is common when a continuous spatial domain is discretized to lower the computational burden.

Table 3 presents the frequency distribution of the Spatial Models adopted in the literature versus the Statistical modeling, showing that the statistical model most often adopted is the Generalized Linear Mixed Model (GLMM), which appears in 75.7% of the total articles reviewed. For the spatial models, the CARs variation was utilized in 44.4% of the total articles reviewed and it is generally used for spatial area/network data. From this family, the percentage of CAR and the BYM spatial model appeared at 96.3%. Examples of these papers are Sharkey and Winter (2019) and Fairley *et al.* (2008). Among the other categories, *SPDE* (3.6%), commonly used for smooth projection of geostatistical data with INLA; *GMRF* except CAR family and SPDE (4.0%), commonly developed for more specific problems; non-parametric models (1.3%) often used for nonstationary stochastic processes, especially for expensive medical spatial data; Article-specific models (7.2%) include those with author-defined covariance structure, which is common in solving problems in econometrics. *Unspecified models* and those that did not fit into any of these groups appeared in 8.2%. The result shows that the combination of CAR models and GLMM accounts for nearly half of the articles reviewed. The CAR models have been validated in numerous simulation experiments and hence it is suitable for modeling area/lattice spatial data in GLMM.

Table 3 – Crosstab spatial priors used versus statistical model adopted. The GLMM with a CAR spatial prior family for the spatial component is the most frequently used modeling structure in the literature. Though some alternatives have been growing in the past decade such as the GLMM framework combined with non-GMRF, GLMM with SPDE, and spatial autoregressive model define dependence matrices.

	GLMM	Non-parametric	Spatial Econ.	Survival models	Proposed	Unspecified	Other	TOTAL
CAR Family	227	1	1	9	-	4	3	245 (44.4%)
SPDE	20	-	-	-	-	-	-	20 (3.6%)
GMRF (excluding above)	19	-	-	-	-	-	3	22 (4.0%)
non-GMRF	101	-	49	1	5	1	16	173 (31.3%)
Non-Parametric	4	-	-	-	-	1	2	7 (1.3%)
Article-specific method (none of the above)	17	-	-	-	-	-	23	40 (7.2%)
Unspecified	30	-	2	-	-	9	4	45 (8.2%)
TOTAL	418 (75.7%)	1 (0.2%)	52 (9.4%)	10 (1.8%)	5 (0.9%)	15 (2.7%)	51 (9.2%)	552

The nature of the target variable dictates the adopted class of statistical model. In the search, 14.1% of the articles modeled a dichotomous target variable, 38.6% modeled a continuous variable, 43.1% modeled a discrete count variable, 1.8% modeled combined discrete and continuous variables, and 1.8% modeled an ordinal variable. Due to the integrated complexity of the posterior marginal distribution, the MCMC estimation method is the most frequently adopted numerical integration, as depicted in Figure 10. However, the use of the INLA technique

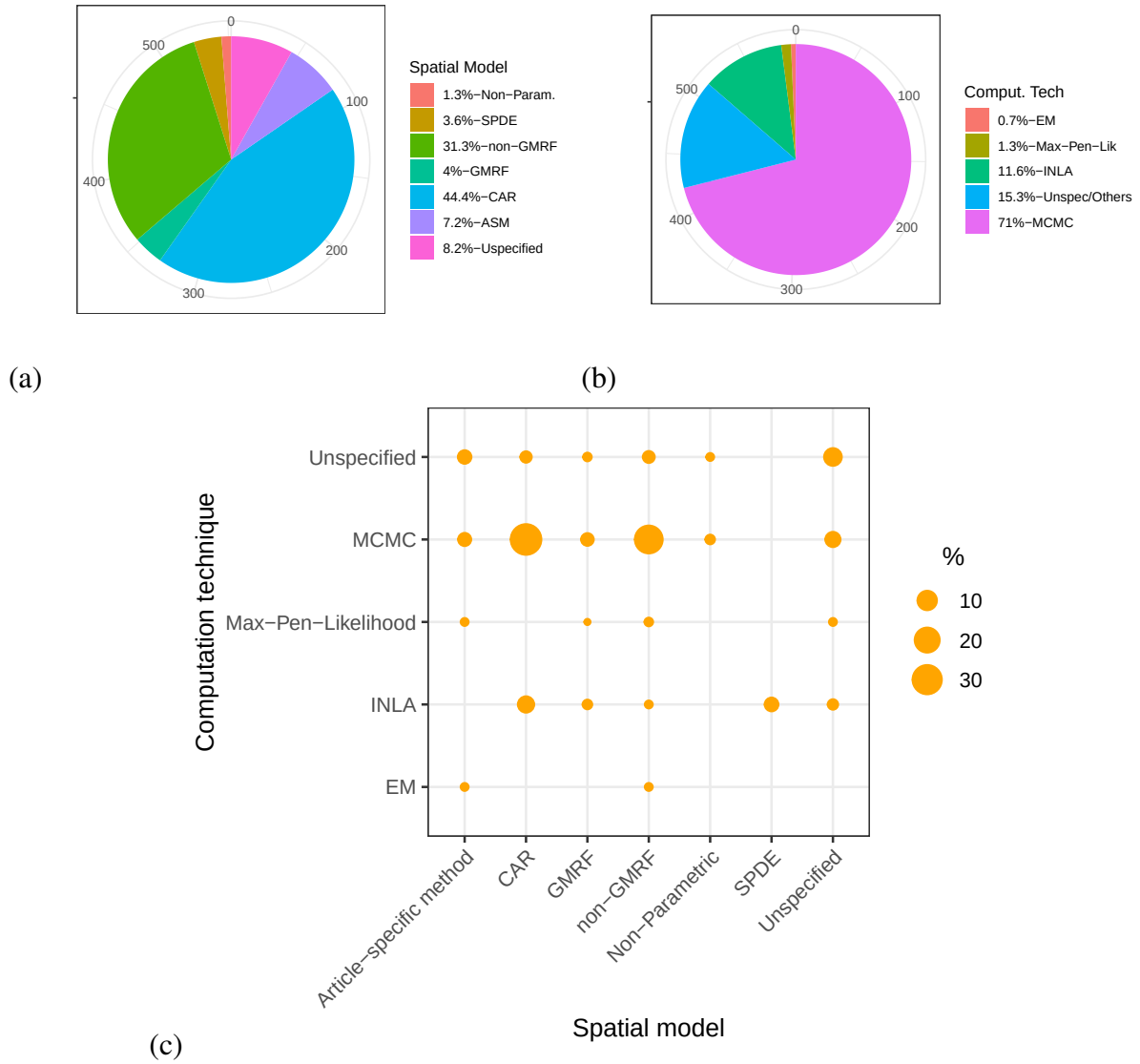


Figure 10 – A pie chart illustrating the percentage of (a) the spatial models and (b) computational techniques found in the literature. (c) The interaction of the type of spatial models and computational method. The result indicates that the CAR family with the MCMC technique is the most frequent combination in the literature.

is fast growing as studies have shown that INLA is significantly more robust and faster than the MCMC technique (TAYLOR; DIGGLE, 2014). The CAR family in combination with MCMC is the most popular combination relative to the others (Figure 10c).

The results obtained in this systematic review show that the expert's knowledge was used in conjunction with the data information (30.43%), as shown in Table 4, although this can be better explored.

The top 5 journals that published the most reviewed papers related to the subject under study (combining theoretical methodology with publications of real-world applications) were: Spatial and Spatio-temporal Epidemiology (# 15), Accident Analysis and Prevention (# 14), PLoS ONE (# 14), Spatial Statistics (# 11), and Environmentrics (# 10).

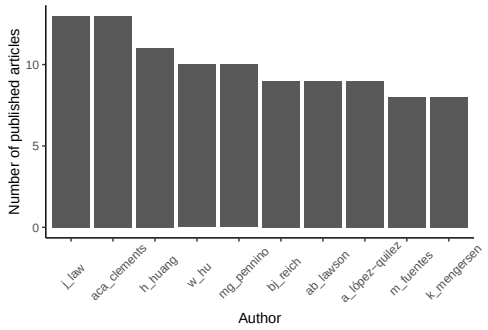


Figure 11 – Top ten authors with the highest number of published articles based on the query keywords and inclusion/exclusion criteria.

Table 4 – Model Prior specified.

Prior specified	**Percentage(%)**
Elicited from experts or the problem	30.3
No explicit use or reference/not applicable	18.2
Used verbatim from the literature	30.0
Vague prior (Non-informative)	21.5

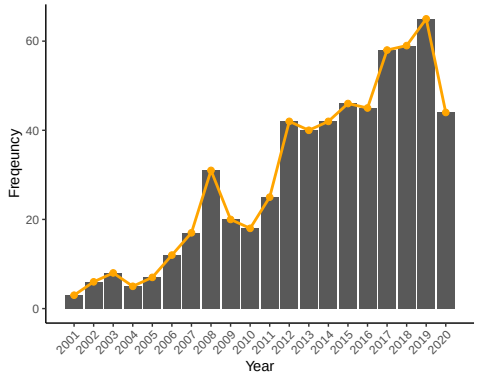


Figure 12 – Growth in scientific publications related to topics in Bayesian Spatial Models from 2001 to June 2020. There was a positive growth over the 20 years considered. The growth could be associated with improvement in computational tools and data collection.

The most frequently appeared authors are Jane Law (University of Waterloo, Canada), Archie C. A. Clements (University of Queensland, Australia), and Helai Huang (Central South University, China) as shown in figure 11. These authors are followed, orderly, by Wenbiao Hu (University of Queensland, Australia), Maria Grazia Pennino (Universitat de València, Spain), Brian J. Reich (North Carolina State University, USA), Andrew B. Lawson (Medical University of South Carolina, USA), Antonio López-Quílez (Universitat de València, Spain), Montserrat Fuentes (North Carolina State University, USA), and Kerrie Mengersen (Queensland University of Technology, Australia). Across time, the spatial modeling publication rate using the Bayesian approach proliferated, as shown in Figure 12. The year 2020 covers only the first half of the year.

3.4 Concluding remarks

Spatial statistics has gained tremendous attention in recent years due to the efficiency in collecting spatial dependence data. Neglecting such dependencies may result in bias and, consequently, lead to inadequate inferences. The Bayesian approach for analyzing spatial data often outperforms the frequentist approach, given that the prior information is taken into account, and it is able to estimate the distribution of the latent variables. In a Bayesian framework, spatial priors play a significant role in accounting for space dependencies. With the consistency in the improvement of data collection and computational tools in analyzing large spatial data, Bayesian spatial statistics will further penetrate numerous fields and become one of the leading tools for analyzing data.

Many authors account for spatial dependencies by assuming a Gaussian random field. In many real data applications, the Gaussian random field may be inappropriate, especially in extreme data, skewed data, and data with spikes and heavy tails. Examining a different random field such as the Laplace, Student-t, and more, may improve inference. A significant issue in assuming these distributions is the computational cost and difficulty in the choice of the prior model for the hyper-parameters. For instance, setting the prior distribution for the degrees of freedom of the student-t distribution. Beyond the Gaussian assumption, several articles reviewed also assumed the Markov theorem (conditional independence), which may be too restrictive as several phenomena exhibit a wide range of spatial autocorrelation. The Markov property is commonly assumed to lower the computational cost in Bayesian inferential statistics. For instance; immediate neighbor spatial dependence. However, the occurrence of some random phenomenon exhibits strong spatial dependencies beyond the immediate neighbors, and truncating such dependencies structure would result in bias and incorrect inferences. Moreover, there is an insufficient standard approach to determine the covariance matrix structure of spatial dependence with Markov property. Thus, it is difficult to compare different models.

The Bayesian spatial literature lacks sufficient information on objective priors, such as Jeffery's prior, reference priors, matching priors, and more. These priors stand out to elicit ideas from data that could improve inference. To derive an objective prior distribution for a spatial random field's model parameters is currently an unaddressed problem that needs attention. Though the objective prior and the penalized complexity (PC) prior have been developed in Ordoñez *et al.* (2023) and Simpson *et al.* (2016) for student t process and log-Gaussian cox processes respectively, their performances have not been adequately explored. Regardless of the prior distribution, eliciting priors for the parameters is critical, and when wrongly assumed, it could lead to misleading results and inference. To circumvent them, which is also not trivial, it is essential to consider objective priors for the random field model parameters and hyper-parameters to improve inference.

Many spatial phenomena are influenced by covariates that vary across space. While Bayesian spatial models can incorporate covariates, there is a need for more advanced techniques

that allow for spatially varying covariates, capturing their heterogeneity and interactions with the underlying spatial processes more accurately. Bayesian spatial statistics inherently provides a framework for uncertainty quantification, but more attention is required in the development of robust techniques for assessing and communicating uncertainty in spatial models.

In Neuroscience and Bioinformatics, the application of Bayesian spatial statistics to brains and single-cell spatial transcriptomic experiments, though limited, has been gaining interest (DERADO *et al.*, 2013; SONG; NATHOO; BABUL, 2019; ALLEN *et al.*, 2022). The complexity of the cell structure, biological interaction, and misalignment induced by the data acquisition procedure have prevented the application of classical spatial models to address intrinsic problems in the field. In other words, the primary assumption of spatial contiguity may be inappropriate. That is, a response received at one location may be due to an activity in the opposite distant location. Hence, models capable of estimating these types of interactions and correcting for data misalignment would be highly relevant and informative.

Therefore, it is imperative to give sufficient attention to the field of spatial statistics since researchers are constantly confronted with the task of analyzing spatial data, which is becoming increasingly complex as data collection methods improve.

CHAPTER

4

TMSBRAINAPP: A SHINY INTERFACE FOR TMS MAPPING DATA

This chapter presents the TMSBrainApp R package that contains a user-friendly Shiny interface for analyzing Transcranial Magnetic Stimulation (TMS) mapping data. This interface allows users access to an application that utilizes a Bayesian spatial statistical modeling technique, making the interface a powerful tool for performing Bayesian spatial analysis in R without the need for coding expertise. The interface was documented using standard R documentation. It is publicly available and can be downloaded or installed directly from a GitHub repository through https://github.com/eosafu/TMSBrainApp1.1.0. A demonstration of the interface is available here. The user interface was used to analyze the TMS datasets in this project, and the findings are presented in this chapter.

4.1 TMSBrainApp R package

The TMSBrainApp is an R-package that features a shiny application designed for descriptive visualization, Bayesian analysis, and inference-making regarding the spatial patterns of a subject's response to TMS pulses. With this tool, users can perform full Bayesian spatial statistical modeling without the need for programming skills. Though the TMSBrainApp was specifically developed to uncover new insights into the spatial patterns of patient responses to TMS pulses, it can also be used to perform Bayesian modeling of any dataset. It is an ideal tool for practitioners and researchers who are interested in estimating spatial patterns and quantifying the effect of explanatory variables on the phenomenon under study.

TMSBrainApp is designed to allow users to fit Bayesian spatial statistical models and determine their uncertainty about the hotspot location and make predictions on unsampled locations based on the posterior distribution. To carry out an analysis, users only need to click buttons and the application processes the inputs and generates tables and figures. The output

Table 5 – The R functions of the TMSBrainApp package that are available to users.

Function call	Parameters	Description
ExceedanceProb	INLAoutput, dist, quantile	Computes exceedance probability at the hotspot location.
GetAdjMatrixTMS	ID	Computes neighborhood matrix
QuantResidexp	INLAoutput, y	Quantile residual for exponential data model fitted with inla().
QuantResidGamma	INLAoutput, y	Computes quantile residuals for gamma data model fitted with inla().
QuantResidGauss	INLAoutput, y	Computes quantile residuals for Gaussian data model fitted with inla().
QuantResidLNorm	INLAoutput, y	Computes quantile residuals for Log normal data model fitted with inla().
QuantResidWeibull	INLAoutput, y	Computes quantile residual for Weibull data model fitted with inla().
QuantResidTMS	INLAoutput, y,dist	Computes the quantile residual of fitted model.
TMSapp.check	TMSdata	Determines whether a user dataset is suitable for the application.
TMSapp.run	-	Launches the shiny Web interface for hotspot finding.
TMSdata	-	TMS sample data.

INLAoutput is an INLA object obtained using inla(), dist is the chosen likelihood in inla(), y is the response variable, quantile is the quantile to which exceedance probability is to be calculated, ID is a coordinate-like indicator to form counties from geostatistical data. TMSdata is the inbuilt sample data.

comprises both one and two dimension graphs for clearer visualization and interpretation. The processing time is fast as it only takes a few minutes to complete a single analysis, and almost immediately to perform a post-modeling analysis. It allows users to generate an HTML report of the analysis executed. The application performs posterior integration and obtains the posterior densities using the INLA package (RUE; MARTINO; CHOPIN, 2009) in R.

4.2 TMSBrainApp Package Installation

The TMSBrainApp R source package is available on the GitHub repository and users can install directly using utilities available in devtools R package. An alternative is to install from a local drive, which holds the source package. The installation process of TMSBrainApp is easy and requires only internet access. The following R codes can be used to install a working version of the TMSBrainApp from the GitHub repository:

```
1 R> if (!require(devtools)) {
2 install.packages("devtools",dependencies=TRUE)
3 }
4 R>devtools::install_github("eosafu/TMSBrainApp1.1.0")
```

Additional details on how to install directly from GitHub are available in this link. After successful installation, executing the function TMSBrainApp::TMSapp.run() in the R console launches the application and a screenshot of the interface is shown in Figure 13. Table 5 shows the descriptions of the functions available in the package.

TMSBrainApp depends on and suggests other R packages, which are available on different repositories. All dependence packages, except INLA, are available on the CRAN repository. Those available on CRAN are automatically installed along with the TMSBrainApp installation process. Moreover, users are expected to independently install INLA from its repository, available on INLA web page.

4.3 TMSBrainApp interface

TMSBrainApp interface consists of two panels: the sidebar panel and the main panel. The sidebar panel contains the instruction tabs and the main panel contains the output tabs. A screenshot of the interface is shown in Figure 13.

Sidebar Panel

The sidebar panel is the panel that appears on the left side of the application window when TMSBrainApp is launched. It can be sectioned into top and bottom, where the top contains the pre-modeling menu and the bottom contains the post-modeling menu. The top of the panel holds two action button widgets. The first button, ***Browse***, is to upload user data into memory for analysis. The second button, ***Sample Data***, makes available an inbuilt dataset for users to practice and accustom to navigating through the application window. Right below the Browse button is the ***Summary for Signal*** dropdown menu. The dropdown menu includes *Range, Standard deviation, Median*, and *Mean*, which determines the type of summary statistics to adopt for the motor-evoked potential (MEP) signal data. The default value is set to the Range which measures the peak-to-peak MEP signals. Right below is the ***Replication*** slider which allows users to select the position of the MEP replication to display on a descriptive plot. Below the Replication are the ***Response Model*** and ***Spatial Model*** dropdown menus. The Response Model menu includes five likelihood models: *gaussian, exponential, gamma, lognormal*, and *weibull* models. The Spatial model menu includes four types of spatial models: *SPDE, ICAR, BYM,* and *Leroux CAR*. The default values are set to *gaussian* and *SPDE*. These spatial models are described in Section 4.5. Based on the choice of the spatial model selected, additional numerical inputs appear and allow the users to set hyperparameter values for the chosen model. However, a user can continue with the analysis using the default values. Right below these menus are ***Fixed Effect*** and ***Random Effect*** menus whose menus are updated based on the data in the memory. The menus are dropdown, and allow users to select explanatory variables to be included in the model. Right below is the action button, ***Fit Model***. It allows the user to execute all the inputs to perform statistical analysis. Turning attention to the post-analysis menus, the ***MEP*** numerical box and ***Exceedance Probability*** action button allow users to calculate the probability of a random draw from the posterior predictive distribution of the summary MEP at the hotspot exceeds the user entered MEP value. See Section 4.6.2 for details. Finally, the ***Generate Report*** action button allows a user to generate a report of the analysis executed.

Main panel

The main panel holds the result of the analysis, plots, and maps. It contains three tabs: **ADJUST, MODEL SUMMARY**, and **VIEW DATA**. The **ADJUST** tab has seven display windows: the ***Model Diagnostic***, ***Hotspot Coordinate***, ***Data Location***, ***Hotspot Map***, ***Hotspot Credible Interval***, ***Testing Orders***, and ***Replication Effect***. The ***Model Diagnostic*** displays a quantiles residual of the fitted model. The more the residuals overlap the straight line the better

the model. The ***Hotspot Coordinate*** displays the X-Y coordinate with the highest posterior mean. The ***Hotspot Map*** displays a 2D map of the posterior predictive mean of the spatial effect of the whole stimulated region if SPDE is selected, and displays the posterior predictive mean of the MEP if other spatial models are selected. The ***Hotspot Credible Interval*** displays the highest posterior density credible interval, which is also suitable for testing a hypothesis about the hotspot region (see Section 4.6.2). The ***Data Location*** and ***Testing Orders*** display a descriptive plot of the observed MEP. This allows the user to understand the distribution of the current data on memory. Lastly, the ***Replication Effect*** displays the line plot of the posterior mean of the nonlinear effect (random effect) variable included in the analysis. The **MODEL SUMMARY** tab shows the model results, which include fixed effects, hyper-parameters, and marginal likelihood estimates. The **VIEW DATA** tab shows the current data frame in memory.

4.4 Data preparation

The `TMSBrainApp R` package contains several functions, and one of them is the `TMSapp.check()`. The function checks the suitability of user data for the shiny application. If the data check is passed, the function returns a data frame with renamed variables suitable for `TMSapp.run()` and also creates an additional variable `"Rep"` which indicates the number of replication pulses of TMS that gave rise to the MEP signals. It is recommended that users check and extract data using `TMSapp.check(user-data)$data` prior to loading it into the interface for analysis. Additional details about the function are available in the function's documentation in `R`.

`TMSBrainApp` interface allows users to upload personal data and also has an example dataset for users to practice with. It gives users an opportunity to learn how to interpret results before performing real analysis. `TMSBrainApp` only accepts TMS datasets in ***xls*** and ***xlsx*** file extensions. Users are expected to prepare data of interest in ***xls*** or ***xlsx*** format prior to uploading it into memory.

A dataset with a structure different from the one generated by a TMS-EMG device should contain at least fives columns, where the first three are the x, y, and z numerical grid coordinate variables indicating the location where MEP was observed. Column z can be set as missing values if unavailable. The fourth column should be the *Associate target* as *ID* that would be used to aggregate the point reference data, which is required in the *CAR*, *BYM*, *Leroux CAR* models (see Section 4.5.1) and the descriptive plots. The aggregation indicator should be a centroid coordinate, necessary to determine spatial contingency. The last column should be *EMG_Data* variable; an example could be a summarized MEP signal data. Moreover, explanatory variables can be placed in between these columns. In a case where users wish to include EMG signals in the dataset, these signals should be placed on columns on the right side of the *EMG_Data* column already mentioned. It is recommended that users check for minimum data requirements before uploading into memory. Users can view the sample data for a guide.

4.5 Adopted statistical Models

Let the spatial domain $\mathscr{G}$ represent a TMS explorable brain region. By construction, each stimulation location is referred to by its Cartesian coordinate. The spatial domain $\mathscr{G}$ is a 2D representation of the subject explorable motor cortex. That is, $\mathscr{G} \subset \mathbb{R}^2$. Let $Y \in \mathbb{R}$ be a random variable representing the summarised value of the MEP signal, $\mathbf{x}$ be a $1 \times p$ design vector of fixed effects of covariates detailing the biological and socio-economical characteristics of the included subject or TMS technical noise control variables, and w be a metrical covariate, and s be a spatial covariate. Then, the hierarchical model is given as

$$\begin{aligned} Y_i &\sim f(\mu_i), \\ \mu_i &= \mathbf{x}_i^T \boldsymbol{\beta} + g(w_i) + h(s_i), \\ \boldsymbol{\beta} &\sim Normal(\mathbf{0}, \boldsymbol{\Sigma}), \\ (g,h) &\sim \pi(g,h) \\ \boldsymbol{\Lambda} &\sim \pi(\boldsymbol{\Lambda}), i = 1,...,m, \end{aligned} \tag{4.1}$$

where m is the number of locations where data were observed, $\boldsymbol{\beta}$ represents a $p \times 1$ vector of unknown linear effect, $\pi(g,h)$ is a joint probability distribution of the latent smooth nonlinear function g and spatial function h. $\pi(\boldsymbol{\Lambda})$ is an arbitrary joint prior probability distribution of the hyperparameter vector $\boldsymbol{\Lambda}$. f is the probability distribution of $Y_i (i = 1,2,...,m)$ such that $\mathbb{E}(Y_i) = \mu_i$. The variable $\mathbf{x}$ allows the inclusion of explanatory variables and w allows the inclusion of metrical covariates to be modeled non-linearly. f allows the user to specify their own probability distribution. These response distributions include Exponential, Gamma, Gaussian, Lognormal, and Weibull density functions.

The mean of the process is described by the linear predictor (μ_i). The model assumed that conditioning on the latent parameters, the process Y_i is independent and identically distributed. In this application, w was modeled using an order two random walk model. That is $g(w) = \mathbf{z}\boldsymbol{\psi}$ where $\boldsymbol{\psi} = (\psi_1, \psi_2, ..., \psi_r)^T$ and $\mathbf{z}$ is a binary row vector that is only one at the entry which corresponds to the metrical value of w. Thus, the latent parameter $\boldsymbol{\psi}$ is modeled as

$$\begin{aligned} e_i &= \psi_i - 2\psi_{i+1} + \psi_{i+2}, \\ e_i &\sim Normal(0, \tau_\psi^{-1}). \end{aligned} \tag{4.2}$$

Its probability density function is derived from its $n-2$ second-order increment as

$$\pi(\boldsymbol{\psi}) \propto \tau_\psi^{r-2} \exp\Big\{ -\frac{1}{2} \boldsymbol{\psi}^T \tau_\psi \mathbf{Q}_\psi \boldsymbol{\psi} \Big\} \tag{4.3}$$

where $\mathbf{Q}_\psi$ is a structured precision matrix reflecting the neighborhood of the random walk model (see Section 5.2).

Four spatial models were adopted for the spatial function h. These spatial models include the Intrinsic conditional autoregressive model (ICAR), Besag, York, and Mollié (BYM), Leroux

conditional autoregressive mode (Leroux CAR), and the Stochastic Partial Differential Equation (SPDE). Let $\mathbf{a}$ be a row vector such that $h(s_i) = \mathbf{a}_i\boldsymbol{\theta}$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_q)^T$. Thus, the latent spatial field $\boldsymbol{\theta}$ is modeled using the earlier mentioned spatial models. These models are described in the following subsections.

4.5.1 Intrinsic conditional autoregressive model (ICAR)

The ICAR (BESAG, 1974) model is one of the most popular GMRF models to account for spatial dependence in spatially aggregated data mainly due to the availability of computational tools such as the WinBUGS.

Assuming the spatial domain $\mathscr{G}$ can be partitioned into q disjoint subsets, $G_1, G_2, ..., G_q$, called the "Associate Targets", such that $s_i \in G_l$ and s_i is at a closer distance from the centroid of G_l than the centroid of $G_{l'}$, $l' \neq l$. That is the spatial effect corresponding to the spatial partition G_l is θ_l. Then the conditional distribution of the fields is given as

$$\theta_l|\boldsymbol{\theta}_{-l}, \tau_\theta \sim Normal\left(\frac{1}{c_l}\sum_{k\in\delta_l}\theta_k, \frac{\tau_\theta^{-1}}{c_l}\right), \; l = 1, , ..., q, \tag{4.4}$$

where $\boldsymbol{\theta}_{-l} = (\theta_1, \theta_2, ..., \theta_{l-1}, \theta_{l+1}, ..., \theta_q)^T$, c_l is the number of partitions that share a border with D_l and δ_l is an index set of those partitions having a border with partition D_l. In this model, $\mathbf{a}_i$ is a $1 \times q$ binary vector that has a single non-zero entry in index l. This entry corresponds to the partition D_l where Y_i belongs.

4.5.2 Besag, York, and Mollié (BYM) -

BYM model (BESAG; YORK; MOLLIÉ, 1991) is an extension of the ICAR model to include an unstructured component that accounts for over-dispersion exhibited in spatial data. Suppose that the effect θ_l is partitioned such that $\theta_l = \theta_{ul} + \theta_{vl}$, where θ_{ui} is the structured component of the spatial effect and the θ_{vl} is the unstructured component. Thus, θ_{vl} is modeled as

$$\theta_{vl} \mid \sigma_v^2 \sim Normal(0, \sigma_v^2) \; \forall l, \tag{4.5}$$

and the structured component is modeled as $\theta_{ui} \sim ICAR$. Thus,

$$\begin{aligned} E(\theta_l|\boldsymbol{\theta}_{-l}, \tau_\theta, \sigma_v^2) &= \frac{1}{c_l}\sum_{k\in\delta_l}\theta_k, \\ Var(\theta_l|\boldsymbol{\theta}_{-l}, \tau_\theta, \sigma_v^2) &= \sigma_v^2 + \frac{\tau_\theta^{-1}}{c_l}. \end{aligned} \tag{4.6}$$

Similarly, vector $\mathbf{a}_i$ is defined in the same way as for the ICAR model. A drawback of the BYM model is the identifiability problem in which each observation is linked to θ_{ui} and θ_{vi}. However, this is overcome by adding a sum-to-zero constraint on $\boldsymbol{\theta}$ (LEE, 2011).

4.5.3 Leroux conditional autoregressive model (Leroux CAR)

Leroux CAR is an alternative way to circumvent the identifiability problem incurred in the BYM model. Leroux, Lei and Breslow (2000) proposed a variant of the CAR model with full conditional distribution given as

$$\theta_l|\boldsymbol{\theta}_{-l},\tau_\theta \sim Normal\left(\frac{\rho\sum_{j=1}^{q} m_{lj}\theta_j}{\rho\sum_{j=1}^{q} m_{lj}+(1-\rho)},\frac{\tau_\theta^{-1}}{\rho\sum_{j=1}^{q} m_{lj}+(1-\rho)}\right), \tag{4.7}$$

where m_{lj} is the $l,j\,th$ entry of a proximity matrix and $\rho \in (0,1)$ is a correlation parameter to be estimated. The vector $\mathbf{a}_i$ is defined in the same way as for the ICAR model.

4.5.4 Stochastic Partial Differential Equation (SPDE)

Lindgren, Rue and Lindström (2011) proposed an SPDE model whose solution is a Gaussian random field with a Matérn covariance function. Further, they proposed a different representation of the solution to the SPDE with a GMRF through a finite element method mesh triangulation, therefore providing a possible sparse precision matrix representation for the Gaussian random field $\boldsymbol{\theta}$. The sparse representation of the precision matrix is to enable fast numeric computation, as in the case of the lattice models described in previous subsections. Moreover, the SPDE model is given as

$$(\kappa^2-\Delta)^{\alpha/2}h(s)=\mathbf{W}(s),\;\; \alpha=v+d/2,\;\; \phi>0,v>0, \tag{4.8}$$

where $\mathbf{W}(s)$ is a Gaussian white noise stochastic process with unit variance, v is the smoothing parameter of the Matern covariance function, d is the dimension of $\mathscr{G}$, and Δ is a Laplace operator. The Matern covariance function to the solution of the SPDE in 4.8 is then given by

$$\mathbf{C}ov(s_i,s_j;v,\phi)=\frac{\sigma^2}{\Gamma(v)2^{v-1}}\Big(\phi d(s_i,s_j)\Big)^{v}K_v\Big(\phi d(s_i,s_j)\Big), \tag{4.9}$$

where $d(s_i,s_j)$ is a metrical distance between location s_i and s_j, and σ^2 is the marginal variance given as

$$\sigma^2=\frac{\Gamma(v)}{\Gamma(v+d/2)(4\pi)^{d/2}\phi^{2v}}.$$

Lindgren, Rue and Lindström (2011) showed that, on a regular grid, the GMRF representation when $v=2$ is a convolution of the processes with a precision matrix of $v=1$. That is, the GMRF representation becomes denser for increased values of v, leading to a wider neighborhood structure, thereby losing its sparsity, increasing smoothness, and incurring expensive computation. Further, they proposed a set of solutions on a triangular mesh representation for irregular points in the spatial domain using the finite element method. The representation is given as

$$h(s)=\sum_{l=1}^{q} a_l(s)\theta_l, \tag{4.10}$$

where θ_l is a normally distributed weight, q is the total number of triangulation nodes, and $a_l(.)$ is a deterministic basis function at each node. The value of the basis function centered at triangle vertices is formed and evaluated at a given coordinate. It is a function of the area of the opposite triangle formed by the other two vertices. The basis function values at the vertices sum to one. If a coordinate, however, falls on the edge of the triangulation, the basis function at this point will only have two non-zero values, whereas, when the point falls on the node, the basis function has a single non-zero value, which is one.

The basis function values for each data point can be coded into a row vector $\mathbf{a}$, such that (4.10) can be written as $h(s_i) = \mathbf{a}_i\boldsymbol{\theta}$ and $\mathbf{a}_i = (a_1(s_i), a_2(s_i), ..., a_q(s_i))$. The spatial field $\boldsymbol{\theta}$ is then modeled using a Gaussian Markov Random Field (GMRF) with a precision matrix governed by the Matérn covariance function. Details of the covariance matrix are given in Krainski *et al.* (2018).

An intuitive parameterization is to control the Matérn covariance parameters through the marginal standard deviation σ and range $\rho_v = \sqrt{8v}/\phi$. The Penalized Complexity (PC) prior is assigned for these parameters. A choice of the pair (U_σ, p_0) and (U_ρ, q_0) are chosen such that $P(\sigma > U_\sigma) = p_0$ and $P(\rho_v < U_\rho) = q_0$ (SIMPSON *et al.*, 2017; GÓMEZ-RUBIO, 2020) so that the hyperparameter of the prior distribution is determined. In `TMSBrainApp`, the spatial models were implemented using `INLA` package in R (R Core Team, 2019). A major concern in spatial modeling is the spatial confounding problem, where the fixed effects are correlated with the spatial effect. However, it is straightforward to address this problem using the restricted spatial regression technique. This is described in detail in Appendix 7.6.

4.6 Data analysis with `TMSBrainApp`

The first step is to import data through the browse button and select the type of signal summaries to adopt. A box plot according to the associate targets is displayed in the main window. Use the Replication slider to vary the box plot, which shows how MEP's first replication compares with other replications. Select the response and spatial models and explanatory variables. Input the hyperparameter values and click on Fit Model to analyze the data. The results show the location of the hotspot and the highest posterior density credible interval of the MEP at the hotspot. For post-modeling analysis, type a numerical value of a summarized MEP signal in the MEP text box and click on exceedance probability. Figure 14 shows a screenshot of an example of modeling results using the `TMSBrainApp` .

4.6.1 Choice of hyperparameters

The SPDE model requires setting hyperparameter values, including σ^2 and ρ_v. Thus, `TMSBrainApp` gives room for users to select hyperparameters values for U_σ, U_ρ, p_0, and q_0, with default values derived from the analysis of multiple TMS datasets. Specifically, the default

values are $U_\sigma = 1, U_\rho = 0.05, p_0 = 0.01$, and $q_0 = 0.01$. Similarly, the default values of the mesh parameters were obtained in the same procedure. The lattice spatial models allow users to select the scale and shape hyperparameters for the prior distributions on $\log \tau_\theta$ and $\log 1/\sigma_v^2$, where both are assigned log gamma prior distribution with parameters **shape** $= 1$ and **scale** $= 0.001$ in the application.

4.6.2 Highest posterior density, Tail probability, and Quantile residual

Let $U(R_k(Y), Y) = \mathbb{I}(Y \in R_k(Y)) - k \int_{Y \in R_k(Y)} dY$ be a utility function we wish to maximize, then the best credible region for Y_o is such that $R_k(Y_o) = \{Y \in \mathscr{Y} : f(Y_o \mid \mathbf{y}) > k\}$, then $R_k(Y_o)$ is the Highest Posterior Density (HPD) credible region for Y_o with credible probability $(1-\alpha_k) = \int_{R_k(Y_o)} f(Y_o|\mathbf{y}) dY_o$, where $f(Y_o|\mathbf{y})$ is the marginal posterior predictive distribution of Y_o, and α_k is the level of significance. Particularly to the `TMSBrainAPP`, Y_o was taken as the highest posterior predictive mean of the MEP among all locations in the study region, and $R_k(Y_o)$ is the credible interval at 95% displayed on the ***Higher posterior density*** window on the main panel. Pereira and Stern (1999) proposed a Full Bayesian Significant Test (FBST) using the HPD credibility interval. The null hypothesis (H_0) is tested as follows. Construct an HPD credibility interval with $1-\alpha_k$ credibility, then reject H_0 if and only if none of the points of H_0 lies within the credible interval. Thus, users can perform hypothesis tests of 95% credibility.

It may be intuitive to understand the posterior distribution of the hotspot to compare the hotspot across subjects or the same subject but different TMS sessions. `TMSBrainApp` allows the users to calculate tail probabilities of the posterior distribution of the MEP at the hotspot. For example, assuming the interest is to determine

$$p_t = P(Y_o > y_o|\mathbf{y}_t) = \int \mathbb{I}(Y_o > y_o) f(Y_o \mid \mathbf{y}_t) dY_o \tag{4.11}$$

for TMS session $t = 1, 2$, where y_0 is a user specified MEP value. If $p_2 > p_1$, then we could conclude that the subject at session 2 is more likely to respond higher to TMS pulses at the hotspot compared with the subject at session 1. Equation 4.11 is computed using Monte Carlo simulation from the marginal posterior predictive distribution of the MEP at the hotspot. $10,000$ samples are drawn at each click of the ***Exceedance probability*** button.

Pearson and deviance residuals are not often suitable for evaluating the adequacy of a generalized linear mixed model (FENG; LI; SADEGHPOUR, 2020), and thus, in this work, the quantile residual alternative was adopted to evaluate goodness-of-fit. Suppose $\boldsymbol{\nabla}$ is a vector of all the parameters of the adopted model, and $F(y; \boldsymbol{\nabla})$ is the distribution function with density f given in (4.1). Further, let $F^\star(Y; \hat{\boldsymbol{\nabla}})$ be the estimated distribution function obtained by replacing $\boldsymbol{\nabla}$ with the posterior Bayes estimate, $\hat{\boldsymbol{\nabla}}$. The quantile residual of Y follows as

$$z^Q = \Phi^{-1}(F^\star(Y; \hat{\boldsymbol{\nabla}})), \tag{4.12}$$

where Φ^{-1} is the quantile function of a known probability distribution. In this work, Φ^{-1} was taken as the normal quantile due to its well-known characteristics. For Y_i, $i = 1, 2, ..., m$,

z_i^Q is sometimes standardized as $r^Q = (z_i^Q - \bar{z}^Q)/s_{z^Q}$, where $\bar{z}^Q = m^{-1}\sum_{i=1}^m z_i^Q$ and $s_{z^Q} = (m-1)^{-1}\sum_{i=1}^m (z_i^Q - \bar{z}^Q)^2$. Then, r^Q is compared with the theoretical standard normal distribution shown in the ***Model Diagnostic*** plot in the output window.

4.7 Application to TMS data

The subject's data were first checked for minimum requirement using the function `TMSapp.check()` and the output data frame was fed into the shiny application for modeling. The log peak-to-peak (log range) MEP was adopted and served as the response variable. The adjusted model included the indicators of associates target, stimulation depth, replications, and spatial location. Based on the DIC and WAIC model selection criteria, the SPDE with the Gaussian response model outperformed the other response and spatial model combinations for subjects 1R1, 1R2, and 2, and the log-normal response model was optimal for subject 3. Therefore results from these models were adopted for inference.

4.7.1 Result

The quantile residual plot is shown in Figure 15. The coordinate of the spatial domain for each subject and the corresponding estimated hotspot location is presented in Table 6. In addition, the table shows the posterior mean and 95% credible interval in log scale, and the exceedance probability of obtaining more than $300\mu V$ peak-to-peak at the hotspot.

Table 6 – Posterior estimate about the subjects' hotspot.

Subject	Spatial Domain	Hotspot	Posterior Mean	95% credible Interval (CI)	$P(\text{MEP} > 300\mu V)$
1R1	$[12.7, 65.6] \times [-41.0, 19.0]$	(46.282,-11.989)	4.838	(3.716,6.156)	0.110
1R2	$[39.4, 68.5] \times [-46.9, 8.5]$	(59.621,-25.223)	3.774	(2.796,4.788)	0.000
2	$[-70.5, -17.3] \times [-41.8, 12.5]$	(-45.125,-16.466)	4.470	(3.301,5.586)	0.020
3	$[-69.8, -30.5] \times [-43.8, 18.5]$	(-42.686, 6.059)	3.315	(3.315, 4.152)	0.000

In Table 6, the spatial domain represents the explored rectangular spatial region, and it is represented as a rectangular coordinate $[a,b] \times [c,d]$, where $[a,b]$ is the interval limit of the x coordinate and $[c,d]$ is for the y coordinate. The adopted coordinates vary depending on the head structure and positioning of the participants. The "hotspot" in the form (a,b) is the coordinate of the estimated hotspot, where a is the x coordinate, and b is the y coordinate. The 95% credible interval of the posterior predictive at the hotspot shows that subject 1R1 has the highest response rate, with a probability of 0.110 that the peak-to-peak MEP at the hotspot will exceed $300\mu V$. Subject 2 shows a similar response pattern with a probability of 0.020 that the peak-to-peak MEP will exceed $300\mu V$. In a similar way, subjects 1R2 and 3 have similar response patterns with 0 probability that the peak-to-peak MEP response will exceed $300\mu V$.

Figure 16 shows the spatial pattern of the mean of the posterior predictive distribution of the peak-to-peak MEP for all the subjects. The result shows a wide peak that spans over the hotspot coordinates. Figure 17 shows the locations on the grid over the motor cortex where

data were obtained for all the subjects. The plots differentiate these locations according to the posterior mean of the random intercepts. In addition, the cross sign denotes the estimated hotspot location, and the box point types are the Associate target location where stimulation with TMS at the motor cortex begins. This location is commonly referred to as the primary motor cortex. The result shows that the posterior means of the random intercepts for locations around the hotspot have the highest positive effect on the log peak-to-peak MEP. Figure 18 shows the posterior density of the spatial range for all the subjects. The vertical dashed line indicates the posterior mean. The distribution shows that the spatial range is highest for subject 1R1 and lowest for subject 3. A higher spatial range indicates that the MEP signal exhibits spatial correlation across a wider spatial region over the motor cortex. The result of the nonlinear effect of replication for all the subjects (result not shown here) showed that there is no significant effect on the peak-to-peak MEP.

4.7.2 Discussion of result

Estimating the hotspot in a TMS treatment session is essential to determining its effectiveness. As shown in the result, the hotspot's peak-to-peak MEP and spatial pattern vary between participants. While a certain quantity of magnetic pulses may cause the MEP signal to reach a high peak for one participant, the same quantity may lead to a flat MEP signal for a different participant at the same cross-sectional location in the motor cortex. The finding from the result also showed that the location of the hotspot is not often among the stimulated locations, and could be found in unstimulated locations within the motor cortex. Moreover, the hotspot location can be regarded as a region rather than a single point. This is evident in the plots in Figure 16. That is, the population of neurons surrounding the hotspot may produce a response that is similar to the one at the hotspot. This finding corroborates the result of Rao (2013). These findings could be useful for producing a consistent motor response for the improvement of patients' experiences.

4.8 Conclusion

This chapter presented a user interface for performing Bayesian spatial modeling to give new insight into the spatial pattern of subjects' responses to TMS brain mapping, which could help in improving the treatment experience for patients with motor impairment. The interface allows users to detect a patient's response hotspot to TMS pulses, thus giving insight that could be useful for investigating several hypothetical questions. The interface is documented using standard R documentation procedure and executes smoothly on the R interface. It is available publicly and could be used by practitioners and experimenters. The interface is capable of being deployed and hosted online with the right resources. This will enable its usage without the need for R programming software.

The analysis conducted showed that there is a variation in the spatial pattern of response to TMS pulses. Moreover, the result showed a non-significance in the effect of replication on the MEP and strong evidence of unexplained variation in the residual. The next chapter investigates these characteristics further by proposing a way to perform prior elicitation from historical TMS data. The Chapter focused on how data obtained from subjects $1R1$ and $1R2$ could be statistically integrated to explain additional variation in the MEP signal.

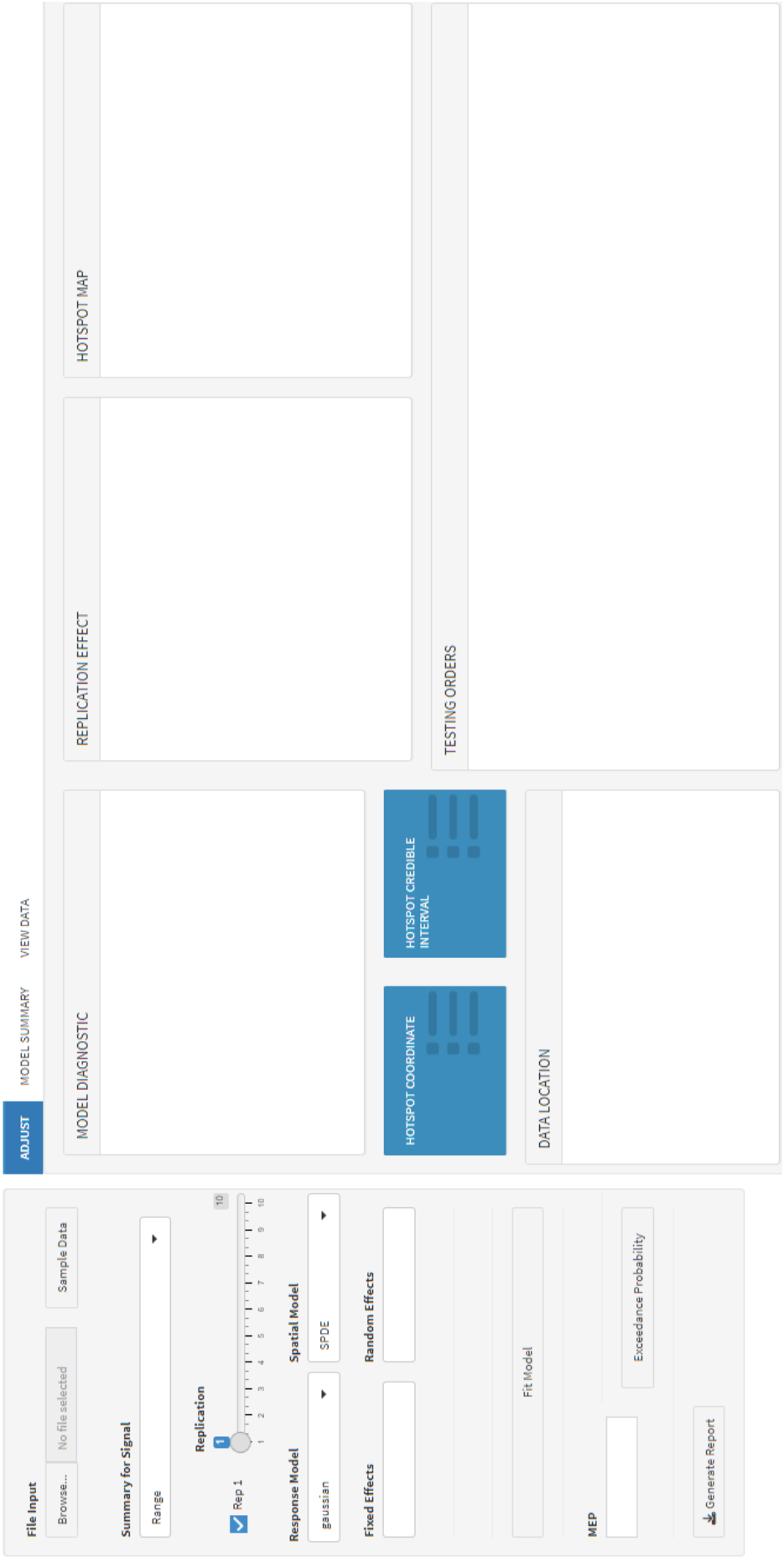


Figure 13 – A screenshot of the sidebar and main panels of `TMSBrainApp`.

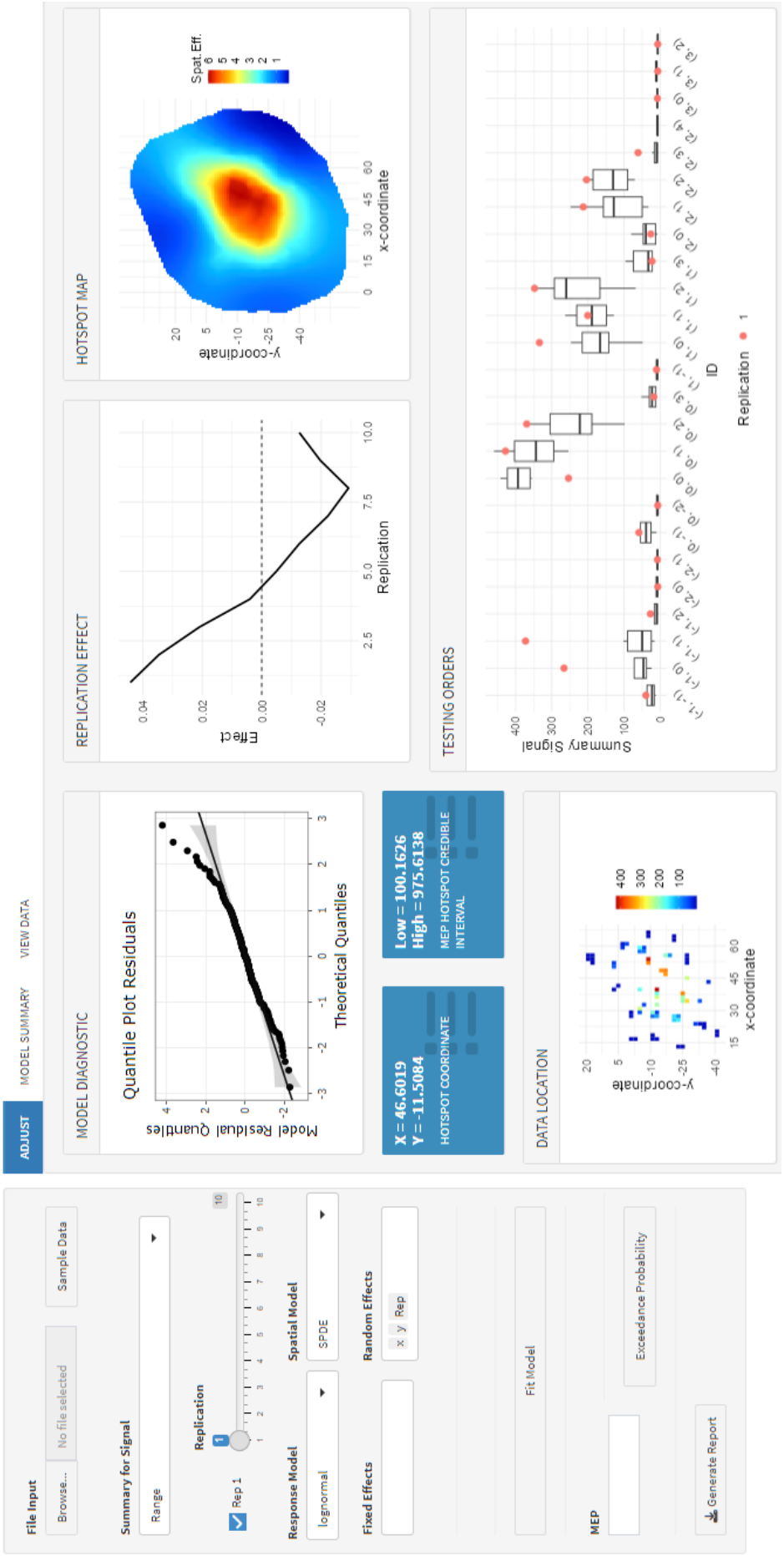


Figure 14 – A screenshot of the analysis of sample data.

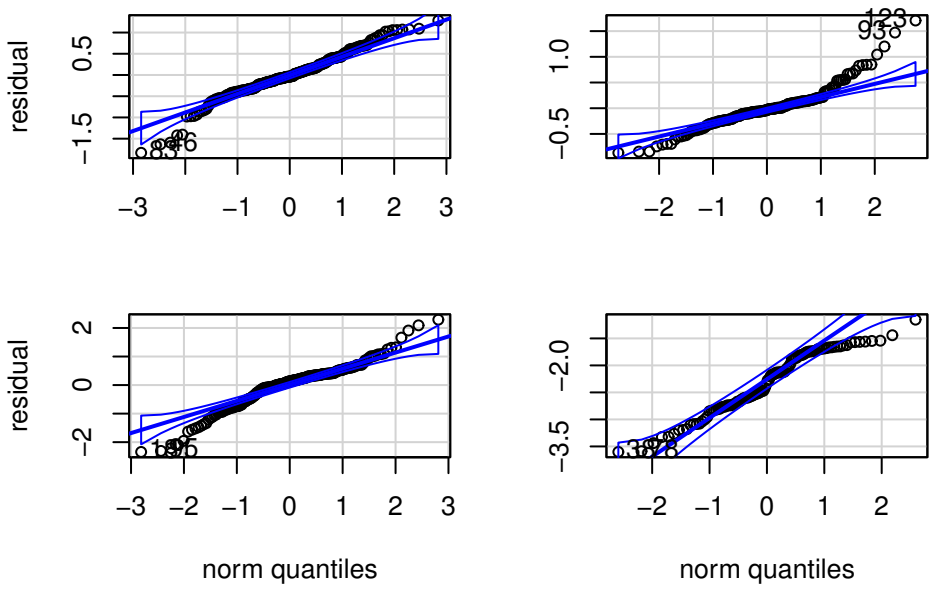


Figure 15 – Residual plot associated with subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3.

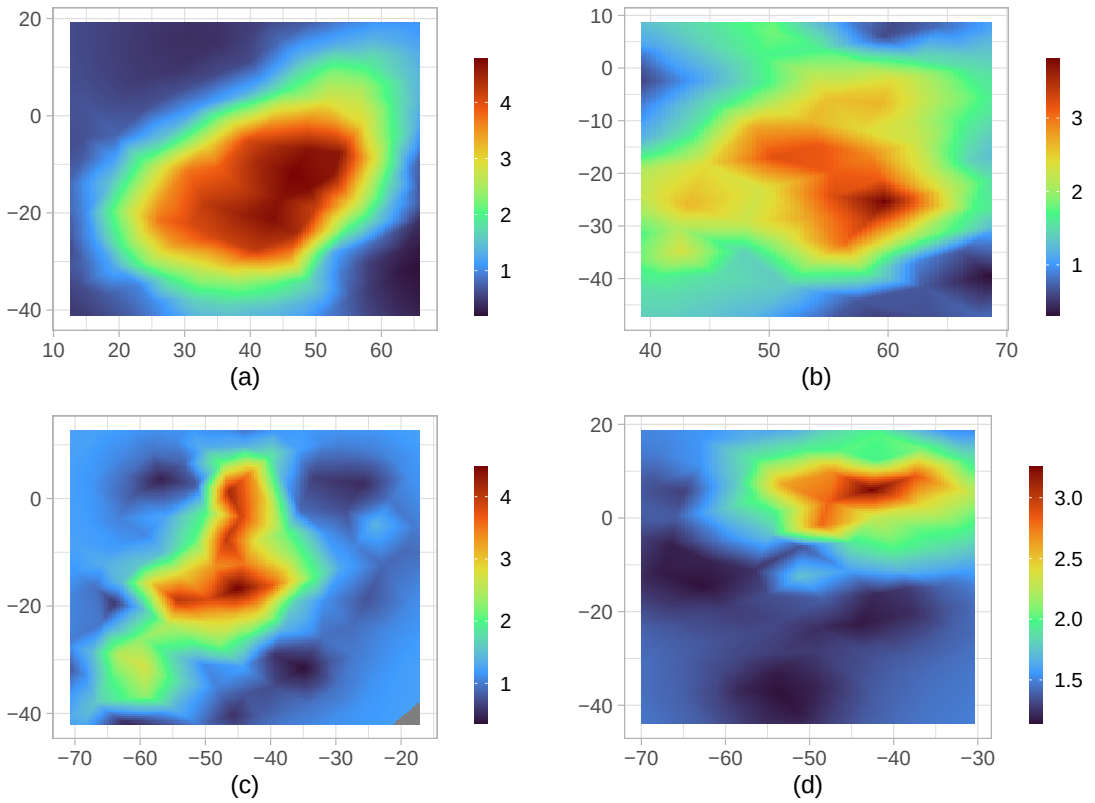


Figure 16 – Mean of the posterior predictive distribution for subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3.

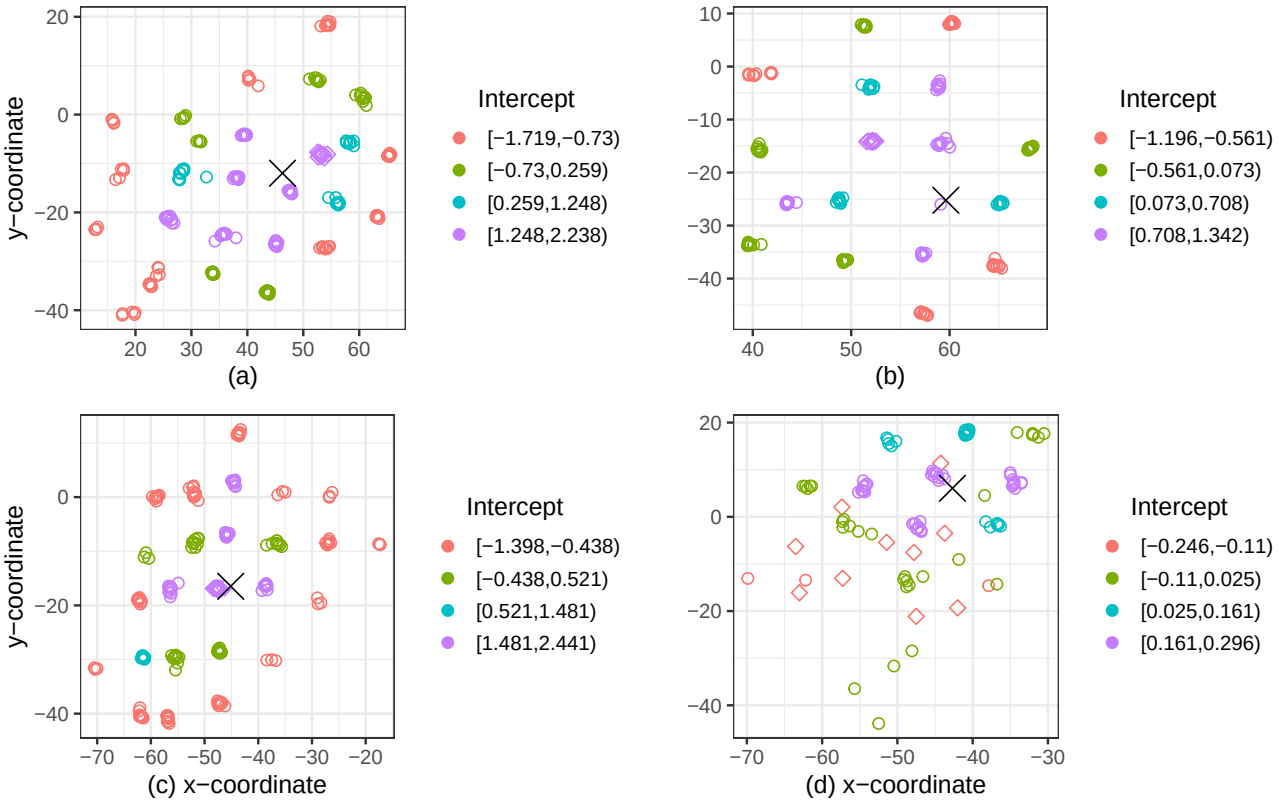


Figure 17 – The data location for subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3 according to the estimated random intercept. The cross sign denotes the predicted hotspot location and the blue box denotes the primary motor cortex where stimulation begins.

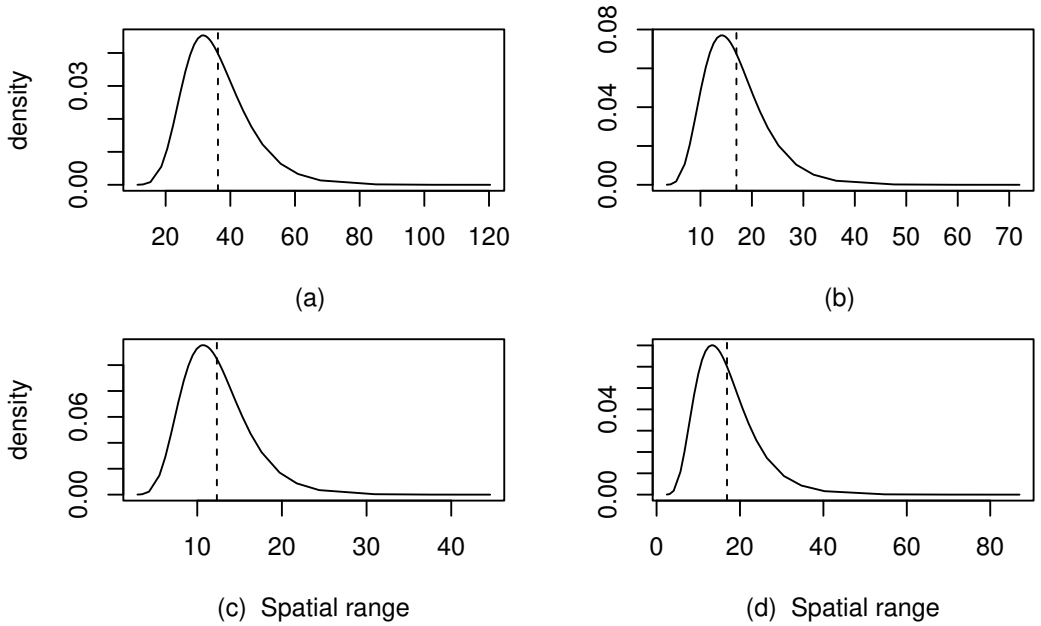


Figure 18 – Posterior density of the spatial range parameter for subjects (a) 1R1, (b) 1R2, (c) 2, and (d) 3. The vertical dashed line indicates the posterior mean.

CHAPTER

5

PRIOR ELICITATION FOR GAUSSIAN SPATIAL PROCESS

The power and commensurate prior distributions are informative prior distributions that incorporate historical data as prior knowledge in Bayesian analysis to improve inference about a phenomenon under study. Although these distributions have been developed for analyzing non-spatial data (see Section 2.4 in Chapter 2), little or no attention has been given to spatial geostatistical data. In this chapter, these informative prior distributions were extended to a Gaussian spatial process, which enables the elicitation of prior knowledge from historical geostatistical data for Bayesian analysis. Three informative prior distributions were proposed for spatial modeling, and an efficient Markov Chain Monte Carlo algorithm was developed for inference. A simulation study was used to assess the adequacy of the informative prior distributions. Hierarchical models combined with the developed informative prior distributions were used to analyze subjects 1R1 and 1R2 TMS datasets.

5.1 Model formulation

In many spatial modeling problems, the interest is to estimate the latent spatial process generating a random vector $\mathbf{Y}$ through the model

$$\begin{aligned} \mathbf{Y}_j &= \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_j, \\ \boldsymbol{\mu}_j &= \mathbf{X}_j^T \boldsymbol{\beta} + w(\mathbf{z}_j) + f(\mathbf{s}),\ j = 1,2...,n, \end{aligned} \tag{5.1}$$

where $\mathbf{Y}_j = (Y_{j1}, Y_{j2}, ..., Y_{jm})^T$ is an $m \times 1$ vector of response variable at jth replication, and each entry, Y_{ji}, corresponds to the response at spatial location $s_i\,(i = 1,2,...,m)$ in a given spatial domain $\mathscr{G}$ at replication j. $\boldsymbol{\mu}_j \in \mathbb{R}^m$, $\boldsymbol{\varepsilon}_j$ has a multivariate normal distribution. That is, $\boldsymbol{\varepsilon}_j \sim N(\mathbf{0}, \tau^{-1}\mathbf{I})$, and $\mathbf{I}$ is an identity matrix of size m. $\boldsymbol{\beta} = (\beta_1, ..., \beta_p)^T$ is a regression parameter for a $p \times m$ design matrix $\mathbf{X}_j$, and the terms $w(.)$ and $f(.)$ are smooth functions, where $w(.)$ corresponds to smooth function of metrical covariate $\mathbf{z}_j = (z_{j1}, z_{j2}, ..., z_{jm})$ and $f(.)$ corresponds

to spatial smooth function for location covariate $\mathbf{s} = (s_1, s_2, ..., s_m)$. The implication of (5.1) is that the realizations of the response variable at each spatial location are independent conditioning on the latent variables in the linear predictor. The functions $w(.)$ and $f(.)$ are approximated using some basis functions, such as cubic regression spline and Gaussian Markov Random Field respectively.

Let the unknown smooth function $w(.)$ be approximated by a spline function such that

$$w(z_{ji}) = \sum_{k=1}^{r} \psi_k B_k^d(z_{ji}), \tag{5.2}$$

where B_k^d is the kth basis function and $r = K + d + 1$. K is the total number of chosen knots and d is the spline's degree of freedom. Particularly in this work, $d = 3$ and K is the number of chosen knots, giving rise to a cubic B-spline (PERPEROGLOU *et al.*, 2019). $B_k^d(z_{ji})$ represents the evaluation of the kth basis function at the observed covariate. The value of the basis function at z is obtained as

$$\begin{aligned} B_k^d(z) &= \frac{z - \xi_k}{\xi_{k+d} - \xi_k} B_k^{d-1}(z) - \frac{\xi_{k+d+1} - z}{\xi_{k+d+1} - \xi_{k+1}} B_{k+1}^{d-1}(z), \; k = 1, 2, ..., r, \\ B_k^0(z) &= \begin{cases} 1, \; \xi_k \leq z \leq \xi_{k+1} \\ 0, \text{ otherwise,} \end{cases} \end{aligned} \tag{5.3}$$

and $B_k^0(z) = 0$ if $\xi_k = \xi_{k+1}$, where ξ_ks are the chosen equidistant knots. Hence (5.2) can be rewritten as $w(z_{ji}) = \mathbf{c}_{ji}^T \boldsymbol{\psi}$, where $\mathbf{c}_{ji} = (B_1^d(z_{ji}), B_2^d(z_{ji}), ..., B_r^d(z_{ji}))^T$ and $\boldsymbol{\psi} = (\psi_1, ..., \psi_r)^T$. Thus, for all z_{ji} in $\mathbf{z}_j$,

$$w(\mathbf{z}_j) = \mathbf{Z}_j^T \boldsymbol{\psi}, \tag{5.4}$$

where each row of $\mathbf{Z}_j^T$ is the $\mathbf{c}_{ji}^T$ corresponding to the evaluation of the basis functions in (5.2) at z_{ji}.

The stochastic spatial process $f()$ at location s_i was modeled as

$$f(s_i) = \sum_{l=1}^{q} \theta_l a_l(s_i), \tag{5.5}$$

where a_l is obtained from some spatial kernel function. Equation 5.5 shares similarities with the Karhunen-Loeve decomposition of stochastic processes. This decomposition is founded on the eigendecomposition of the process's covariance function (SIMPSON *et al.*, 2016). In this work, a_l is defined as

$$a_l(s_i) = \exp\left(-\frac{d(s_i, v_l)^2}{\varphi^2} \right), \varphi > 0, \tag{5.6}$$

where v_l is the lth location of the chosen spatial knot, $v_l \in \mathscr{D}$ and $d(s_i, v_l)$ is an euclidean distance between data location s_i and knot location v_l. Reich, Bandyopadhyay and Bondell

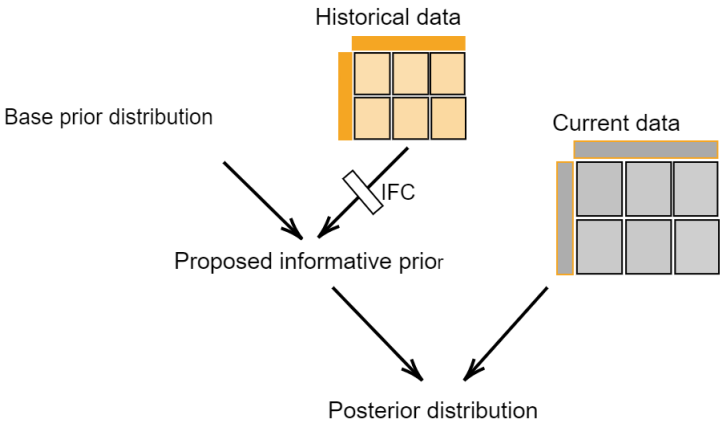


Figure 19 – Flow chat illustrating the borrowing/data integration scheme. The Information Flow Control (IFC) varies the amount of information passed down from the historical data to the Bayesian modeling and it is a function of the consistency between the two datasets.

(2013) showed that for a fixed number of spatial knots q and bandwidth φ, and any continuous marginal distribution for $\mathbf{Y}_j$, there exists a density function for θ_l that leads to that marginal distribution of $\mathbf{Y}_j$. Hence, this representation leads to a valid probability distribution. Given (5.6), (5.5) can be written as $\mathbf{a}_i\boldsymbol{\theta}$, where $\mathbf{a}_i = (a_1(s_i), a_2(s_i), ..., a_q(s_i))$ and $\boldsymbol{\theta} = (\theta_1, \theta_2, ..., \theta_q)^T$. Therefore, for all s_i in $\mathbf{s}$,

$$f(\mathbf{s}) = \mathbf{A}\boldsymbol{\theta}, \tag{5.7}$$

where each row of $\mathbf{A}$ is $\mathbf{a}_i$ at the corresponding location s_i. The choices of spatial knot location are chosen to obtain an optimal model. A higher number of spatial knots over $\mathscr{D}$ improves the performance of the model but increases complexity. (5.1) becomes

$$\begin{aligned} \mathbf{Y}_j &= \boldsymbol{\mu}_j + \boldsymbol{\varepsilon}_j, \\ \boldsymbol{\mu}_j &= \mathbf{X}_j^T\boldsymbol{\beta} + \mathbf{Z}_j^T\boldsymbol{\psi} + \mathbf{A}\boldsymbol{\theta},\ j = 1, 2..., n. \end{aligned} \tag{5.8}$$

This framework presents a platform to facilitate information borrowing from geostatistical historical data. Therefore, the main model parameters we require elicitation are $\boldsymbol{\beta}, \boldsymbol{\psi}$, and $\boldsymbol{\theta}$.

Figure 19 illustrates the proposed borrowing scheme. The first step is to determine the base prior distribution to incorporate the knowledge about the parameters before the historical data observation. The second step involves developing informative prior distribution that integrates the historical data and the base prior distribution. The Information Flow Control (IFC) regulates the amount of information from the historical data allowed into the modeling. The regulation depends on the consistency between historical data and current data. IFC allows more information to go through if both data are consistent; otherwise, less information is allowed if both data are inconsistent. The last step is to combine the informative prior distribution with the current data in a Bayesian analysis for inference.

5.2 Base prior distributions

In the frequentist approach, $\boldsymbol{\beta}$, $\boldsymbol{\psi}$, and $\boldsymbol{\theta}$ are commonly penalized through some penalty terms to avoid overfitting. However, in the Bayesian framework, prior distributions are assigned to all the parameters in the model. This subsection presents the base prior distributions that incorporate prior knowledge about these parameters before the observation of historical data. Particularly, the spatial parameter $\boldsymbol{\theta}$ assumes a Gaussian Random Field (GRF) to capture the spatial dependencies in the data. That is, $\boldsymbol{\theta} \sim N(\mathbf{0}, \tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))$, where $\mathbf{Q}_\theta(\boldsymbol{\kappa})$ is a structured variance-covariance matrix derived from a Matérn covariance function such that $\mathbf{Q}_\theta(\boldsymbol{\kappa}) = \sigma_\theta^2[q_{l,l'}]$, and $q_{l,l'} = C(v_l, v_{l'}|\phi, \nu)$, which follows as

$$C(v_l, v_{l'}; \phi, \nu) = \frac{1}{2^{\nu-1}\Gamma(\nu)}(d(v_l, v_{l'})\phi)^\nu \mathscr{K}_\nu(d(v_l, v_{l'})\phi); \phi > 0, \nu > 0, l, l' = 1, 2, ...q, \quad (5.9)$$

where $\mathscr{K}_\nu$ denotes the Bassel function of the second kind. ϕ controls the range of the covariance function and ν controls its smoothness. Notice that v_l is the knot location of the spatial fields.

The computation of $\mathbf{Q}_\theta(\boldsymbol{\kappa})$ in the prior distributions is expensive for large values of q. To circumvent this issue, a Nearest Neighbor Gaussian Process (NNGP) approximation for $\boldsymbol{\theta}$ (FINLEY *et al.*, 2019) was considered. First let $\boldsymbol{\theta} = \mathbf{H}\boldsymbol{\theta} + \mathbf{e}$, where $\mathbf{H} = [h_{ll'}]$ is a $q \times q$ strictly lower-triangular matrix, $h_{ll'} = 0$ if $l' \geq l$ and $\mathbf{e} \sim N(\mathbf{0}, \tau^{-1}\sigma_\theta^2 \mathscr{V})$ and $\mathscr{V} = [v_{ll'}]$ is a diagonal matrix. It is straightforward to show that the covariance of $\boldsymbol{\theta}$, $\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}) = (\mathbf{I} - \mathbf{H})^{-1}\tau^{-1}\sigma_\theta^2 \mathscr{V}(\mathbf{I} - \mathbf{H})^{-T}$ where for any matrix K, K^{-T} is the inverse of its transpose. For the expression to be valid, $\mathbf{H}$ and $\mathscr{V}$ are obtained as follows. Let $N(v_l)$ be a collection of the indexes less than l and $C = \mathbf{Q}_\theta(\boldsymbol{\kappa})$ be the covariance matrix obtained using the matérn covariance function earlier defined. Then the matrices $\mathbf{H}$ and $\mathscr{V}$ are computed as follows:

$$\begin{aligned} \mathbf{H}[l, N(v_l)] &= C(v_l, N(v_l))(C(N(v_l), N(v_l)))^{-1}, \\ \mathscr{V}[l, l] &= C(v_l, v_l) - \mathbf{H}[l, N(v_l)]C(N(v_l), v_l), l > 1, \\ \mathbf{H}[1, :] &= \mathbf{0} \text{ and } V[1, 1] = C[1, 1], \end{aligned} \quad (5.10)$$

where $\mathbf{H}[1, :]$ indicates the first row of matrix $\mathbf{H}$. Now, to make $\mathbf{Q}_\theta(\boldsymbol{\kappa})$ a sparse matrix, only the indexes in $N(v_l)$ that are nearest neighbors of knot v_l are considered to obtain $\tilde{\mathbf{Q}}_\theta$. Thus, the inverse $\tilde{\mathbf{Q}}_\theta^{-1}(\boldsymbol{\kappa}) = (\mathbf{I} - \tilde{\mathbf{H}})^T \frac{1}{\sigma_\theta^2} \tilde{\mathscr{V}}^{-1}(\mathbf{I} - \tilde{\mathbf{H}})$ is a sparse matrix, where $\tilde{\mathbf{H}}$ and $\tilde{\mathscr{V}}$ are computed using only the nearest neighbors. The determinant $| \tau^{-1}\tilde{\mathbf{Q}}_\theta(\boldsymbol{\kappa}) | = (\tau^{-1}\sigma_\theta^2)^q \prod_{l=1}^q v_{ll}$. Note that from (5.10), $C(1:2, 5:6)$ for example is a 2×2 matrix.

A weakly informative prior is assumed for the regression coefficients $\boldsymbol{\beta}$. Here, $\boldsymbol{\beta}$ is assigned a base Gaussian prior distribution with zero mean and a $p \times p$ variance-covariance matrix $\tau^{-1}\mathbf{Q}_\beta = \tau^{-1}\sigma_\beta^2\mathbf{I}$ with a large value for σ_β^2. A base random walk (RW) prior model (LANG; BREZGER, 2004) according to a metrical basis matrix $\mathbf{Z}$, with Gaussian error, was assumed for $\boldsymbol{\psi}$. Here, a random walk 1 (RW1) model with independent increments is demonstrated. The RW1 model follows as

$$\psi_t = \psi_{t-1} + e_t, \psi_1 = e_1, t = 2..., r, e_t \sim N(0, \sigma_\psi^2/\tau), t = 1, ..., r.$$

Solving recursively,

$$\begin{aligned}
&\psi_1 = e_1,\\
&\psi_2 = \psi_1 + e_2 = e_1 + e_2,\\
&\psi_3 = \psi_2 + e_3 = e_1 + e_2 + e_2 + e_3,\\
&\dots\\
&\psi_r = e_1 + \dots + e_r.
\end{aligned}$$

Therefore, $\boldsymbol{\psi} = \mathbf{Re}$, where $\mathbf{e} = (e_1, \dots, e_r)^T$ and matrix $\mathbf{R} = [I_{tt'}]$, and $I_{tt'} = 1$ if $t \leq t'$ and 0 otherwise. Then the precision matrix becomes $\tau \mathbf{Q}_\psi^{-1}$, where

$$\mathbf{Q}_\psi^{-1} = \frac{1}{\sigma_\psi^2}(\mathbf{R}\mathbf{R}^T)^{-1} = \frac{1}{\sigma_\psi^2}\begin{pmatrix} 2 & -1 & & & \dots & 0 \\ -1 & 2 & -1 & & \dots & 0 \\ & -1 & 2 & -1 & \dots & 0 \\ \vdots & \dots & \dots & \dots & \dots & \dots \\ \vdots & \dots & \dots & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 1 \end{pmatrix}.$$

Therefore, $\boldsymbol{\psi}$ is a multivariate normal with zero mean and covariance matrix $\mathbf{Q}_\psi$. The precision matrix revealed that ψ_t only depends on ψ_{t-1} and ψ_{t+1} conditioning on the rest.

Since $\boldsymbol{\beta}$, $\boldsymbol{\psi}$, and $\boldsymbol{\theta}$ have multivariate Gaussian prior distributions centered at zero with covariance matrices $\tau^{-1}\mathbf{Q}_\beta$, $\tau^{-1}\mathbf{Q}_\psi$, and $\tau^{-1}\tilde{\mathbf{Q}}_\theta(\boldsymbol{\kappa})$, then the field $\mathscr{X} = (\boldsymbol{\beta}^T, \boldsymbol{\psi}^T, \boldsymbol{\theta}^T)^T$ has a Gaussian distribution with mean $\mathbf{0}$ and a block diagonal variance-covariance matrix $\tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa})$, where $\boldsymbol{\Sigma}(\boldsymbol{\kappa})$ is obtained by diagonally stacking the corresponding covariance matrices, and each block corresponds to the vectors in $\mathscr{X}$. Then $\mathbf{Y}_j \mid \mathscr{X} \sim N(\mathbf{V}_j\mathscr{X}, \tau^{-1}\mathbf{I})$, where $\mathbf{V}_j$ is obtained by horizontally stacking $\mathbf{X}^T, \mathbf{Z}^T$ and $\mathbf{A}$ such that $\mathbf{V}_j\mathscr{X} = \mathbf{X}_j^T\boldsymbol{\beta} + \mathbf{Z}_j^T\boldsymbol{\psi} + \mathbf{A}_j\boldsymbol{\theta}$. The covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{\kappa})$ is sparse since $\mathbf{Q}_\beta, \mathbf{Q}_\psi$, and $\tilde{\mathbf{Q}}_\theta(\boldsymbol{\kappa})$ are themselves sparse matrices. Usually, $p, r < q$. To further lower the computational burden, a fill-reducing permutation matrix technique described in Appendix 7.4.1 was adopted.

5.3 Prior elicitation from point-referenced historical data

Let $\mathscr{D}_0 = \{m_0, n_0, \mathbf{Y}_0, \mathbf{C}_0\}$ denote the historical geostatistical dataset, where $\mathbf{Y}_0$ is the response indexed by m_0 spatial locations in the spatial domain $\mathscr{G}_0$; m_0 indicates the number of locations data were observed; n_0 denotes numbers of independent replications in the historical data and $\mathbf{C}_0 = \{\mathbf{X}_0, \mathbf{Z}_0, \mathbf{A}_0\}$ is the set of observed matrices. It is important to state that $\mathscr{G}_0$ need not be identical to the spatial domain of the current data ($\mathscr{G}$), which jeopardizes the trivial extension of the power and commensurate prior distributions to geostatistical data. Figure 20 illustrates an example of the availability of one historical spatial data in 2 and 3 denominational view. The spatial window of the historical data were represented by $\mathscr{G}_0$ with observed values denoted by $\{y_{01}, y_{02}, y_{03}\}$. The current data we wish to make inferences from is represented by

spatial window $\mathscr{G}$ with observed response y. Pooling or integrating the historical and current data is only appropriate if both data are highly consistent; otherwise, the historical data may dominate the current data especially when the historical data are inconsistent with the current data, which results in inadequate inferences. Thus, in this section, three prior elicitation distributions were developed for borrowing information from point-referenced historical data to improve model performance and prevent historical data dominance in Bayesian analysis. These informative priors adopt different techniques to arrive at the IFC.

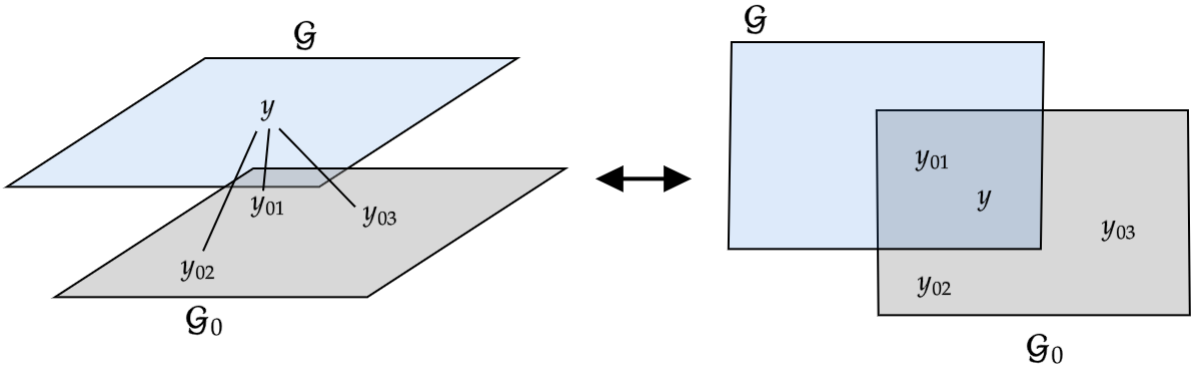


Figure 20 – A two and three-dimensional illustration of historical and current spatial data observed at different spatial windows. The historical data sets were observed in window $\mathscr{G}_{01}$, and are denoted by $\{y_{01}, y_{02}, y_{03}\}$. The response y was observed in the current spatial window $\mathscr{G}$.

5.3.1 Joint spatial power prior

This section proposed a spatial power prior elicitation method for $\mathscr{X}$ based on available historical data. For notation convenience, $s_{0i} \in \mathscr{G}_0$ is used to denote the spatial location where historical data were observed, and $\mathbf{A}_0$ is the corresponding projection matrix.

Particularly for a Gaussian process, suppose $\mathbf{Y}_{0j} \mid \mathscr{X} \sim N(\boldsymbol{\mu}_{0j}, \tau_y^{-1}\mathbf{I}), \boldsymbol{\mu}_{0j} = \mathbf{V}_{0j}\mathscr{X}$ are observed at locations $\{s_{0i}\}_i^{m_0}$ on $j = 1, ..., n_0$ replications, where $\mathbf{V}_{0j} = rowstactk(\mathbf{X}_{0j}^T, \mathbf{Z}_{0j}^T, \mathbf{A}_0)$, $\mathbf{X}_{0j}, \mathbf{Z}_{0j}$, and $\mathbf{A}_0$ were as previously defined for the historical data $\mathscr{D}_0$. Let $\tau \sim Gamma(a_\tau, b_\tau)$, $\omega_0 \sim \pi(\omega_0)$, and a Gaussian random field prior $\mathscr{X} \sim N(\mathbf{0}, \tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa}))$ and $\boldsymbol{\Sigma}(\boldsymbol{\kappa})$ is constructed from the chosen spatial knots with respect to the union of graphs $\mathscr{G}_0$ and $\mathscr{G}$. For example, Figure 21 illustrates spatial knot positioning across $\mathscr{G}_0$ and $\mathscr{G}$. In the example, $\mathbf{Q}_\theta(\boldsymbol{\kappa})$ is a 5×5 covariance matrix derived using the Matérn covariance function (5.9). Similarly, $\mathbf{A}$ and $\mathbf{A}_0$ are respectively 1×5 and 2×5 projection matrices derived using the kernel function (5.6).

The proposed joint spatial power prior with hyperparameter vector $\boldsymbol{\Lambda}$ follows as

$$\pi^p(\mathscr{X}, \tau, \omega_0, \boldsymbol{\kappa} \mid \tau_x, \boldsymbol{\Lambda}, \mathscr{D}_0) \propto \pi(\omega_0)\pi(\boldsymbol{\kappa}) \frac{2^{\frac{v+q^\star}{2}} k^{v/2} \tau^{(m_0 n_0 \omega_0 + q^\star)/2 + a - 1}}{\Gamma(v/2)|\mathbf{D}|^{1/2}} \times \exp\left\{ -\frac{1}{2}\Big(\tau(\mathscr{X} - \mathbf{DB})^T \mathbf{D}^{-1}(\mathscr{X} - \mathbf{DB}) + \tau k\Big)\right\}, \; v > 2, \tag{5.11}$$

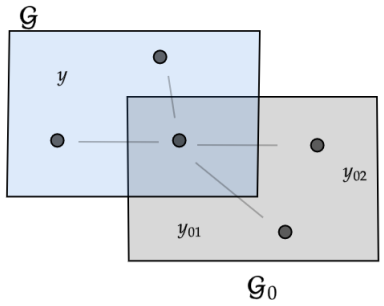


Figure 21 – Five spatial knot positions across historical and current spatial windows.

where $\mathbf{D}^{-1} = (\omega_0\tau_x\sum_j^{n_0}\mathbf{V}_{0j}^T\mathbf{V}_{0j} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\kappa}))$, $\mathbf{B} = \omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{Y}_{0j}$, $k = \omega_0(\tau_x\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\mathbf{Y}_{0j}) - \mathbf{B}^{\mathbf{T}}\mathbf{D}\mathbf{B} + 2b_\tau$, $v = 2a_\tau + m_0n_0\omega_0$, $q^\star = p+r+q$, $\tau_y = \tau\tau_x$ and $\boldsymbol{\kappa} = \{\sigma_\beta^2, \sigma_\psi^2, \sigma_\theta^2, \phi, \nu\}$. A concern about the prior distribution is whether it is proper. A sufficient condition, but not necessary, is that for proper distribution $\pi(\omega_0)$ and $\pi(\boldsymbol{\kappa})$ the joint prior distribution is proper. Notice importantly that setting $\omega_0 = 0$, the historical data vanishes and the prior (5.11) reduces to the base prior distribution, indicating that no information is borrowed from the historical data. It is straightforward to extend to K historical data by substituting $\mathbf{D}^{-1} = \left(\sum_{k=1}^K \omega_{0k}\tau_{xk}\sum_j^{n_{0k}}\mathbf{V}_{0jk}^T\mathbf{V}_{0jk} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\kappa})\right)$, $\mathbf{B} = \sum_{k=1}^K \omega_{0k}\tau_{xk}\sum_{j=1}^{n_{0k}}\mathbf{V}_{0jk}^T\mathbf{Y}_{0jk}$, $k = \sum_{k'=1}^K \omega_{0k'}\tau_{xk'}\sum_{j=1}^{n_{0k'}}\mathbf{Y}_{0k'j}^T\mathbf{Y}_{0k'j} - \mathbf{B}^{\mathbf{T}}\mathbf{D}\mathbf{B} + 2b_\tau$, $m_0n_0\omega_0 = \sum_{k=1}^K m_{0k}n_{0k}\omega_{0k}$. If the interest is only on the spatial effects, then the computation burden can be lessened by integrating out the fixed and the nonlinear effect latent parameters, which results in a simpler spatial power prior distribution followed as

$$\pi^p(\boldsymbol{\theta}, \tau, \omega_0, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}_0) \propto \pi(\omega_0)\pi(\boldsymbol{\kappa})\frac{2^{\frac{v+q}{2}}k^{\frac{v}{2}}\tau^{\frac{n_0m_0\omega_0+q+2a}{2}-1}}{\Gamma(v/2)\mid\mathbf{E}\mid^{1/2}}\exp\left\{-\frac{\tau}{2}\Big((\boldsymbol{\theta}-\mathbf{EF})^T\mathbf{E}^{-1}(\boldsymbol{\theta}-\mathbf{EF})+k\Big)\right\}, \tag{5.12}$$

where $k = 2b - \mathbf{F}^T\mathbf{E}\mathbf{F} + \omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\mathbf{Y}_{0j} - (\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{P}_{0j}\mathbf{Y}_{0j})^T\mathbf{D}(\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{P}_{0j}\mathbf{Y}_{0j})$, $\mathbf{E}^{-1} = (\mathbf{Q}_\theta^{-1} + n_0\omega_0\tau_x\mathbf{A}_0^T\mathbf{A}_0 - (\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{A}_0^T\mathbf{P}_{0j}^T)\mathbf{D}(\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{P}_{0j}\mathbf{A}_0)$, $v = n_0m_0\omega_0 + 2a$, $\mathbf{F} = \Big(\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{A}_0^T\mathbf{Y}_{0j} - (\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{A}_0^T\mathbf{P}_j^T)\mathbf{D}(\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{P}_{0j}\mathbf{Y}_{0j})\Big)$, $\mathbf{D}^{-1} = \tau\Big(\omega_0\tau_x\sum_{j=1}^{n_0}\mathbf{P}_{0j}\mathbf{P}_{0j}^T + \mathbf{L}^{-1}\Big)$, and $\mathbf{P}_{0j}^T = rowstack(\mathbf{X}_{0j}^T, \mathbf{Z}_{0j}^T)$, $\mathbf{L} = \text{diag}(\mathbf{Q}_\beta, \mathbf{Q}_\psi)$. ω_0's linked to the current data in $\mathscr{D}$ is strengthened by the precision parameter τ. The derivations are shown in Appendix 7.4.3.1.

A competing choice of distribution for ω_0 is the beta probability distribution. Suppose ω_0 follows a beta distribution, with parameters γ_0 and λ_0, then the density function is given as

$$\pi(\omega_0|\gamma_{0_b}, \lambda_{0_b}) = \frac{\omega_0^{\gamma_{0_b}-1}(1-\omega_0)^{\lambda_{0_b}-1}}{B(\gamma_{0_b}, \lambda_{0_b})},\ \omega_0 \in [0,1],\ \gamma_{0_b}, \lambda_{0_b} > 0,$$

where $B(\gamma_{0_b}, \lambda_{0_b})$ is the beta function. The hyperparameters for the prior distributions on ω_0 play a crucial role in determining the amount of information passed down to the current data. The choice of γ_0 and λ_0 could be challenging, in that, different choices could lead to different inferential

estimates for ω_0. Ibrahim, Chen and Sinha (2001) suggested working with the reparameterization $\mu_0 = \gamma_{0_b}/(\gamma_{0_b} + \lambda_{0_b})$ and $\sigma_0 = \frac{\mu_0(1-\mu_0)}{\gamma_{0_b}+\lambda_{0_b}+1}$ under beta prior distribution on ω_0, which represent its mean and variance. The reparameterization enables direct control over the location and dispersion of the distribution. The adopted parameterization is $\gamma_{0_b} = \mu_0\left(\frac{\mu_0(1-\mu_0)}{\sigma_0} - 1\right)$, $\lambda_{0_b} = \left(\frac{1-\mu_0}{\mu_0}\right)\gamma_{0_b}$, $\mu_0 \in (0,1)$ and $0 < \sigma_0 < \mu_0(1-\mu_0)$. Thus, the prior distribution for ω_0 follows as

$$\pi(\omega_0|\mu_0,\sigma_0) = \omega_0^{\mu_0\left(\frac{\mu_0(1-\mu_0)}{\sigma_0}-1\right)-1}(1-\omega_0)^{\mu_0\left(\frac{1-\mu_0}{\mu_0}\right)\left(\frac{\mu_0(1-\mu_0)}{\sigma_0}-1\right)-1}. \tag{5.13}$$

The hyper-prior distributions follows as $\mu_0 \sim Uniform(0,1)$ and $\sigma_0 \sim uniform(0, \mu_0(1-\mu_0))$. A different prior distribution choice that can be considered for ω_0 includes simplex probability distribution and truncated normal at [0, 1] interval.

5.3.2 Joint spatial commensurate prior

In the previous subsection, the developed prior elicitation technique act as an ancillary for borrowing information about the parameters of interest from historical data. However, for the Gaussian process, the posterior variance of these parameters is inflated by the power parameter ω_0. In this section, a framework for incorporating commensurability between historical and current data as a primary tool to measure the concordance between these data is developed. Unlike the joint spatial power prior, the spatial commensurate prior separates the fields associated with the current and historical data. Let $\mathscr{X}, \mathscr{X}_0 \in \mathbb{R}^{q^\star}$, where $\mathscr{X}$ is the target spatial field. $\mathscr{X}_0$ can be referred to as a copy of $\mathscr{X}$, which learn directly from the historical data. The joint location commensurate prior is given as

$$\pi(\mathscr{X}, \boldsymbol{\lambda}|\mathscr{D}_0) \propto \int L(\mathscr{X}_0|\mathscr{D}_0)\pi(\mathscr{X}|\mathscr{X}_0,\boldsymbol{\lambda})\pi(\mathscr{X})\pi(\boldsymbol{\lambda})d\mathscr{X}_0, \tag{5.14}$$

where $L(\mathscr{X}_0|\mathscr{D}_0)$ is the likelihood of the historical data with parameter $\mathscr{X}_0$ learned from the historical data. $\pi(\mathscr{X}|\mathscr{X}_0,\boldsymbol{\lambda})$ is chosen such that the probability distribution of $\mathscr{X}$ is centered at $\mathscr{X}_0$, where the positive definite matrix $\boldsymbol{\lambda} \in \mathbb{R}^{q^\star \times q^\star}$ measures the commensurate between $\mathscr{X}_0$ and $\mathscr{X}$. $\pi(\mathscr{X})$ is the base prior knowledge on $\mathscr{X}$ and $\pi(\boldsymbol{\lambda})$ denotes the prior knowledge on the commensurate.

To develop the prior for a Gaussian spatial process, start by supposing that the historical response variable $\mathbf{Y}_{0j} \in \mathbb{R}^{m_0}$ is $\mathbf{Y}_{0j} \mid \mathscr{X}_0 \sim N(\mathbf{V}_{0j}\mathscr{X}_0, \mathbf{H}_0^{-1})$, $\mathscr{X} \sim N(\mathscr{X}_0, \boldsymbol{\lambda}^{-1})$, and $\mathscr{X} \sim N(\mathbf{0}, \tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa}))$, where $\mathscr{X}_0 = (\boldsymbol{\beta}_0^{\boldsymbol{T}}, \boldsymbol{\psi}_0^{\boldsymbol{T}}, \boldsymbol{\theta}_0^{\boldsymbol{T}})^T$. A similar construction of the spatial covariance and projection matrices for the joint spatial power prior is adopted for $\mathbf{Q}_\theta(\boldsymbol{\kappa})$. In this setup, Equation 5.14 is reasonable since as $\mathscr{X} \to \mathscr{X}_0$, $\pi(\mathscr{X} \mid \boldsymbol{\lambda}, \mathscr{X}_0) \to L(\mathscr{X}|\mathscr{D}_0)\pi(\mathscr{X})$. Integrating out $\mathscr{X}_0$ according to the relation of equation 5.14, the commensurate prior for a historical dataset

with hyperparameter vector $\boldsymbol{\Lambda}$ follows as

$$\pi^c(\mathscr{X},\boldsymbol{\lambda},\boldsymbol{\kappa}\mid\boldsymbol{\Lambda},\mathscr{D}_0)\propto\frac{\mid 2\pi\mathbf{D}\mid^{1/2}\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})}{\mid 2\pi\mathbf{H}\mid^{n_0/2}\mid 2\pi\boldsymbol{\lambda}^{-1}\mid^{1/2}\mid 2\pi\tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa})\mid^{1/2}}\times$$
$$\exp\Big(-\frac{1}{2}\Big((\mathscr{X}-\mathbf{EF})^T\mathbf{E}^{-1}(\mathscr{X}-\mathbf{EF})-\mathbf{F}^{\mathbf{T}}\mathbf{EF}-(\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{Y}_{0j})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{Y}_{0j})\Big)\Big), \tag{5.15}$$

where $E^{-1}=(\tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1}+\boldsymbol{\lambda}-\boldsymbol{\lambda}^T\mathbf{D}\boldsymbol{\lambda})$, $\mathbf{F}=\boldsymbol{\lambda}^T\mathbf{D}(\sum_j^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{Y}_{0j})$, $\mathbf{B}=(\sum_j^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{Y}_{0j}+\boldsymbol{\lambda}\mathscr{X})$, $\mathbf{D}^{-1}=(\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{V}_{0j}+\boldsymbol{\lambda})$, and $\boldsymbol{\kappa}=\{\sigma_\beta^2,\sigma_\psi^2,\sigma_\theta^2,\phi,\nu\}$. It is easy to see that matrix $\mathbf{E}$ is positive definite. That is

$$\begin{aligned}\boldsymbol{\lambda}^T\mathbf{D}\boldsymbol{\lambda}&=\boldsymbol{\lambda}\boldsymbol{\lambda}^{-1}(\boldsymbol{\lambda}^{-1}\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{V}_{0j}+\mathbf{I})^{-1}\boldsymbol{\lambda}\\&=(\boldsymbol{\lambda}^{-1}\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{V}_{0j}+\mathbf{I})^{-1}\boldsymbol{\lambda}.\end{aligned} \tag{5.16}$$

Since $\boldsymbol{\lambda}^{-1}\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{V}_{0j}$ is positive definite, the eigenvalues of $(\boldsymbol{\lambda}^{-1}\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{V}_{0j}+\mathbf{I})$ are all greater than 1, which implies that the eigenvalues of the inverse $(\boldsymbol{\lambda}^{-1}\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}^{-1}\mathbf{V}_{0j}+\mathbf{I})^{-1}$ lie between 0 and 1. Thus, all the eigenvalues of $(\boldsymbol{\lambda}^{-1}\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{H}_0^{-1}\mathbf{V}_{0j}+\mathbf{I})^{-1}\boldsymbol{\lambda}$ are less than eigenvalues of $\boldsymbol{\lambda}$.

A prior choice for the commensurate parameter $\boldsymbol{\lambda}$ is the Wishart distribution. That is, $\boldsymbol{\lambda}\sim Wishart(\mathscr{M},\upsilon)$, where $\mathscr{M}$ is a positive scale matrix and $\upsilon>q^\star-1$ is the degree of freedom. A different choice that is more computationally efficient is to consider a commensurate parameter for the spatial component and non-spatial components. For instance, let $\lambda_\beta,\lambda_\psi,\lambda_\theta\in\mathbb{R}^+$ be the commensurate parameters. Thus, $\boldsymbol{\lambda}=[\lambda_{ll'}]$, where $\lambda_{ll}=\lambda_\beta$ for $l=1,...,p$, $\lambda_{ll}=\lambda_\psi$ for $l=p+1,...,p+r$, and $\lambda_{ll}=\lambda_\theta$ for $l=p+r+1,...,q^\star$, and 0 otherwise. A gamma prior distribution can then be assumed for λ_β, λ_ψ and λ_θ. The joint commensurate prior given in 5.15 can be easily extended to accommodate multiple observed historical data. The derivations are shown in Appendix 7.4.3.2.

5.3.3 Joint scalable spatial commensurate prior

The joint spatial power prior and commensurate prior distributions developed so far use spatial fields (knots) that span over the spatial windows of both the historical data and the current data. That is the matrices $\mathbf{A}$ and $\mathbf{A}_0$ are constructed from (5.6) for fixed knots $\{v_1,...,v_q\}$ which span over $\mathscr{G}$ and $\mathscr{G}_0$. For several available historical data, q is large, thus high computational power is required to invert $\boldsymbol{\Sigma}(\boldsymbol{\kappa})$. Instead, this section proposed a technique that uses spatial knots constructed for each historical data. For example, the spatial knots $\{v_1,...,v_q\}$ and $\{v_{01},...,v_{0q_0}\}$ are used to construct projection matrices $\mathbf{A}$ and $\mathbf{A}_0$ respectively, such that $q_0,q_1<<q$.

Figure 22 illustrates knot positions for the historical and current spatial windows. Let $N(s_0)$ be the set of q locations of the current knots and $\mathscr{X}_0\in\mathbb{R}^q$ denotes the corresponding spatial

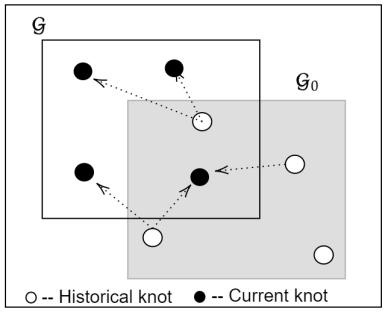


Figure 22 – An illustration of spatial knot positions corresponding to the historical and current spatial windows. The spatial fields at the historical knot locations learn from the historical data and share with the spatial fields at the knots of the current data.

field (parameter) at those locations. Further let $\mathscr{X}^{'} = (\mathscr{X}^{'}[N(s_0)]^T, \mathscr{X}_1^{'T})^T \in \mathbb{R}^{q_c+q_0}, q_c \leq q$ denotes the spatial field at the historical knots, where $q_c + q_0$ is the total number of historical knots. Here $\mathscr{X}^{'}[N(s_0)] \in \mathbb{R}^{q_c}$ is used to denote the spatial field at the historical knot considered to be in close proximity to $\mathscr{X}_0$. In the figure, $\mathscr{X}^{'}[N(s_0)]$ is the spatial fields in the locations of historical knots with arrows, and $\mathscr{X}_1^{'} \in \mathbb{R}^{q_0}$ is the spatial field in the location of the historical knots without arrows, and are considered not in close proximity with $\mathscr{X}_0$. The idea is that the information learned by the spatial field from the historical data at the historical knots' positions is borrowed by the spatial field at the positions of the current knots.

To begin, suppose that $\mathbf{Y}_0 \mid \mathscr{X}^{'}$ is conditionally independent and identically distributed with Gaussian. Thus, given the posterior distribution of $\mathscr{X}^{'} \mid \mathbf{Y}_0$, we can derive the posterior predictive distribution $\pi(\mathscr{X}_0 \mid \mathscr{X}^{'}[N(s_0)], \mathscr{D}_0)$ of the spatial field $\mathscr{X}_0$ at locations in $N(s_0)$. Let $\mathscr{X} \in \mathbb{R}^q$ be the spatial fields at locations of interest in the current spatial window, the joint prior distribution is given as

$$\begin{aligned}\pi^{sc}(\mathscr{X}, \mathscr{X}_0, \mathscr{X}^{'}[N(s_0)], \boldsymbol{\lambda} \mid \mathscr{X}_1^{'}, \mathscr{D}_0) \propto \pi(\mathscr{X} \mid \mathscr{X}_0, \boldsymbol{\lambda}) \times \\ \pi(\mathscr{X}_0 \mid \mathscr{X}^{'}[N(s_0)]) \pi(\mathscr{X}^{'}[N(s_0)] \mid \mathscr{X}_1^{'}, \mathscr{D}_0) \pi(\mathscr{X}) \pi(\boldsymbol{\lambda}),\end{aligned} \tag{5.17}$$

A major difference between (5.14) and (5.17) is that the former assumes a uniform prior distribution for the historical field, whereas, the latter assumes an arbitrary prior distribution. Integrate out $\mathscr{X}_0$ and $\mathscr{X}^{'}[N(s_0)]$ leads to the proposed joint scalable spatial commensurate prior given as

$$\begin{aligned}\pi^{sc}(\mathscr{X}, \boldsymbol{\lambda} | \mathscr{X}_1^{'}, \mathscr{D}_0) \propto \pi(\mathscr{X}) \pi(\boldsymbol{\lambda}) \int \int \Big[\pi(\mathscr{X} | \mathscr{X}_0, \boldsymbol{\lambda}) \times \\ \pi(\mathscr{X}_0 \mid \mathscr{X}^{'}[N(s)]) \pi(\mathscr{X}^{'}[N(s)] \mid \mathscr{X}_1^{'}, D_0) \Big] d\mathscr{X}_0 d\mathscr{X}^{'}[N(s_0)]).\end{aligned} \tag{5.18}$$

Notice that the largest covariance matrix in (5.18) has size $max(q, q_0)$ unlike (5.14). A natural benefit of the prior distribution is that it is hardly affected by noise in the historical data.

Particularly for Gaussian process, let $\mathbf{Y}_{0j} \mid \mathscr{X}^{'} \sim N(\mathbf{V}_{0j}\mathscr{X}^{'}, \tau_1^{-1}\mathbf{I})$ and $\mathscr{X}^{'} \sim N(\mathbf{0}, \tau_1^{-1}\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1))$, $\tau_1 \sim Gamma(a_{\tau_1}, b_{\tau_1})$, where $\mathscr{X}^{'}$ is the spatial field of the historical data and $\boldsymbol{\kappa}_1 = \{\sigma_\beta^2, \sigma_\psi^2, \sigma_\theta^2, \phi_1, \nu_1\}$.

Let $\mathscr{X}^{'} = (\mathscr{X}^{'}[N(s_0)]^T, \mathscr{X}_1^{'T})^T$, then the Bayes update is $(\mathscr{X}^{'}[N(s_0)], \mathscr{X}_1^{'T})^T \mid \tau_1, D_0 \sim N\left(\boldsymbol{\mu}^{'}, \boldsymbol{\Sigma}^{'}\right)$ where

$$\begin{aligned} \boldsymbol{\mu}^{'} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} = (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{V}_{0j} + \boldsymbol{\Sigma}_1^{-1})^{-1} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{Y}_{0j}) \text{ and} \\ \boldsymbol{\Sigma}^{'} = \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix} = \tau_1^{-1} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{V}_{0j} + \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\kappa}_1))^{-1}. \end{aligned} \tag{5.19}$$

Here $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are associated with $\mathscr{X}^{'}[N(s_0)]$ and $\mathscr{X}_1^{'}$ respectively and the corresponding matrices are denoted by the block entries in $\boldsymbol{\Sigma}^{'}$. Therefore, $\mathscr{X}^{'}[N(s_0)] \mid \mathscr{X}_1^{'}, \tau_1 \sim N(\mathscr{A}, \mathscr{B})$, where $\mathscr{A} = \boldsymbol{\mu}_1 + \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}(\mathscr{X}_1^{'} - \boldsymbol{\mu}_2)$ and $\mathscr{B} = \mathbf{Q}_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}$. Let $\mathscr{X}^{'}[N(s_0)]$ be the spatial field in $\mathscr{G}_0$ having close proximity to spatial field $\mathscr{X}_0$ of locations in $\mathscr{G}$. Adopting Markov property, the conditional predictive distribution $\mathscr{X}_0 \mid \mathscr{X}^{'}$ equals $\mathscr{X}_0 \mid \mathscr{X}^{'}[N(s_0)] \sim N\left(\mathbf{c}^T \mathscr{X}^{'}[N(s_0)], \mathbf{D}\right)$, where $\mathbf{c}^T = \boldsymbol{\Sigma}_{01}\mathbf{Q}_{11}^{-1}$ and $\mathbf{D} = \boldsymbol{\Sigma}_{00} - \boldsymbol{\Sigma}_{01}\mathbf{Q}_{11}^{-1}\boldsymbol{\Sigma}_{10}$. Here, $\boldsymbol{\Sigma}_{01}$ is the covariance matrix between the spatial field at locations in $\mathscr{X}_0$ and $\mathscr{X}^{'}[N(s_0)]$, $\boldsymbol{\Sigma}_{00}$ is the covariance matrix of the spatial field associated only to $\mathscr{X}_0$, and $\boldsymbol{\Sigma}_{01} = \boldsymbol{\Sigma}_{01}^T$.

Given this background, the proposed commensurate prior distribution for Gaussian spatial process with hyperparameter vector $\boldsymbol{\Lambda}$ follows as

$$\begin{aligned} \pi^{sc}(\mathscr{X}, \boldsymbol{\lambda}, \boldsymbol{\kappa}, \boldsymbol{\kappa}_1 \mid \mathscr{X}_1^{'}, \boldsymbol{\Lambda}, \mathscr{D}_0) = \frac{\mid 2\pi\mathbf{E} \mid^{1/2} \mid 2\pi\mathbf{K} \mid^{1/2}}{\mid 2\pi\boldsymbol{\lambda}^{-1} \mid^{1/2} \mid 2\pi\mathbf{D} \mid^{1/2} \mid 2\pi\mathscr{B} \mid^{1/2}} \times \\ \exp\left\{ -\frac{1}{2}\left(\mathscr{X}^T\mathbf{M}^{-1}\mathscr{X} - \mathbf{W}^T\mathscr{X} - \mathscr{X}^T\mathbf{W} - \mathscr{A}^T\mathscr{B}^{-1}(\mathbf{K}\mathscr{B}^{-1} - \mathbf{I})\mathscr{A} \right) \right\} \times \\ \pi(\mathscr{X})\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\boldsymbol{\kappa}_1), \end{aligned} \tag{5.20}$$

where $\mathbf{M}^{-1} = (\boldsymbol{\lambda} - \boldsymbol{\lambda}\mathbf{E}\boldsymbol{\lambda} - \boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathbf{K}\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda})$, $\mathbf{W} = (\boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathbf{K}\mathscr{B}^{-1}\mathscr{A})$, $\mathbf{E}^{-1} = (\boldsymbol{\lambda} + \mathbf{D}^{-1})$, and $\mathbf{K}^{-1} = (\mathbf{c}\mathbf{D}^{-1}\mathbf{c}^T + \mathscr{B}^{-1} - \mathbf{c}\mathbf{D}^{-1}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T)$. $\pi(\mathscr{X}) = N(\mathbf{0}, \tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa}))$, and $\boldsymbol{\kappa}$ is as defined in Section 5.3.2. The derivations are shown in Appendix 7.4.3.3. It is straightforward to extend to independent multiple historical data by simply deriving (5.20) for each historical data and obtaining the grand product.

5.4 Hierarchical model

Let $\mathscr{D} = \{m, n, \mathbf{Y}, \mathbf{C}\}$ denote the current geostatistical data, where $\mathbf{Y}$ is the response indexed by m spatial locations in the spatial domain $\mathscr{G}$; m indicates the number of locations data were observed; n denotes numbers of independent replications in the current data; $\mathbf{C} = \{\mathbf{X}, \mathbf{U}, \mathbf{Z}, \mathbf{A}\}$ is the set of observed matrices. Consider $\mathbf{Y}_j$ to be a realization assuming multivariate normal distribution ($\mathbf{Y}_j \sim N(\boldsymbol{\eta}_j, \tau^{-1}\mathbf{I})$) of the current data ($\mathscr{D}$) with $j = 1, 2, ..., n$ replications at spatial location $s_i = 1, ..., m$, and $\boldsymbol{\eta}_j \in \mathbb{R}^m$ is a $m \times 1$ vector of the linear predictor on replication

j. Let $\mathbf{V}_j\mathscr{X} = \mathbf{X}_j^T\boldsymbol{\beta} + \mathbf{Z}_j^T\boldsymbol{\psi} + \mathbf{A}\boldsymbol{\theta}$, $\mathbf{V}_j = rowstack(\mathbf{X}_j^T, \mathbf{Z}_j^T, \mathbf{A})$ and $\mathscr{X} = (\boldsymbol{\beta}^T, \boldsymbol{\psi}^T, \boldsymbol{\theta}^T)^T$, the hierarchical structure is given as

$$\begin{aligned}
\mathbf{Y}_j|\mathscr{X}, \boldsymbol{\vartheta} &\sim N(\boldsymbol{\eta}_j, \tau^{-1}\mathbf{I}),\\
\boldsymbol{\eta}_j &= \mathbf{U}_j^T\boldsymbol{\vartheta} + \mathbf{V}_j\mathscr{X},\\
\mathscr{X} &\sim \pi^o, o \in \{p, c, sc\},\\
\boldsymbol{\vartheta} &\sim \pi(\boldsymbol{\vartheta}),\\
\tau &\sim Gamma(a_\tau, b_\tau).
\end{aligned} \tag{5.21}$$

Here, $\mathbf{U}$ is used to introduce a new set of covariates only available in the current data, and $\boldsymbol{\vartheta}$ is the corresponding effect with prior knowledge $\pi(\boldsymbol{\vartheta}) = N(\mathbf{0}, \mathbf{R})$. To make inferences from the marginal posterior distributions, the Metropolis-Hasting within Gibbs algorithm (MH-G) of the Markov Chain Monte Carlo was implemented to draw posterior samples from the joint posterior distribution. The Gibbs sampling algorithm allows drawing posterior samples from standard posterior conditional distributions, and the Metropolis-Hasting algorithm permits drawing posterior samples from non-standard posterior conditional distributions. The MH-G algorithm is implemented for all the models considered and the detailed algorithms and links to `R` codes are shown in Appendix 7.4.2.

5.5 Simulation study

Forty-five simulation scenarios were used to investigate the performance of the three competing prior distributions for the recovery of the simulated spatial pattern. Current and historical spatial responses were simulated from different distributions with varied expected values and variances, which collectively form the scenarios considered. The marginal expected values of the historical response $\mathbf{Y}_0$ that give rise to the historical data were classified into three scenarios as: "lower", "equal", and "higher" than the marginal expected values of the current response $\mathbf{Y}$ forming the current data. Similarly, with the precision of the historical response, the scenarios were classified into: "lower", "equal" and "higher" than the precision of the response forming the current data. The general form of the sampling distribution scheme is given in the following equations.

For sampling the current data,

$$\begin{aligned}
\boldsymbol{\theta} \sim N(\mathbf{0}, \tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa})), \boldsymbol{\beta} &= (0.5, 3, -1, -2)^T,\\
\mathscr{X} &= (\boldsymbol{\beta}^T, \boldsymbol{\theta}^T)^T,\\
\mathbf{Y} &\sim N(\mathbf{V}\mathscr{X}, \tau^{-1}\mathbf{I}),\\
\text{where } \mathbf{V} &= rowstack(\mathbf{X}^T, \mathbf{I}_\theta),
\end{aligned} \tag{5.22}$$

and for sampling the historical data,

$$\begin{aligned}
\boldsymbol{\beta}_0 &= c_1\boldsymbol{\beta},\ \tau_0^{-1} = c_2\tau^{-1},\\
\boldsymbol{\theta}_0 &\sim N(\mathbf{0}, \tau_0^{-1}\mathbf{Q}_{\theta_0}(\boldsymbol{\kappa})),\\
\mathscr{X}_0 &= (\boldsymbol{\beta}_0^T, \boldsymbol{\theta}_0^T)^T,\\
\mathbf{Y}_0 &\sim N(\mathbf{V}_0\mathscr{X}_0, \tau_0^{-1}\mathbf{I}),\\
\text{where } \mathbf{V}_0 &= \mathit{rowstack}(\mathbf{X}_0^T, \mathbf{I}_{\theta_0}).
\end{aligned} \tag{5.23}$$

Here $\mathbf{X}_0$ and $\mathbf{I}_{\theta_0}$ are the covariates of the fixed effect and an identity projection matrix of the historical data respectively, and $\mathbf{X}$ and $\mathbf{I}_\theta$ are the corresponding matrices in the current data; c_1 and c_2 are constant factors. The sample locations which gave rise to the matrices $\mathbf{Q}_\theta(\boldsymbol{\kappa})$ and $\mathbf{Q}_{\theta_0}(\boldsymbol{\kappa})$ are different and were selected randomly. Specifically, the spatial window of the historical data was fixed at $(0.4,0.9)\times(0.4,0.9)$ and for the current data, $(0.2,0.8)\times(0.2,0.8)$, while $(0.0,1.0)\times(0.0,1.0)$ for the test window in the form $(xcord_1, xcord_2)\times(ycord_1, ycord_2)$. Figure 23 shows the train locations for historical and current spatial data used in the modeling stage and the test locations used to evaluate the out-of-sample predictive performance of the models. A bimodal spatial pattern was simulated, with peaks toward locations $(0.3,0.3)$ and $(0.8,0.8)$.

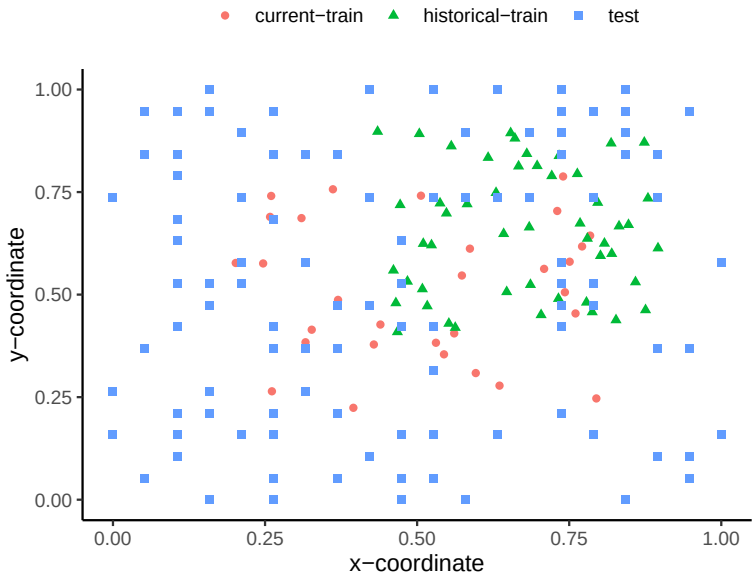


Figure 23 – Simulated data location of current and historical data for training and testing.

Marginalizing over the spatial effect, the resulting marginal expectation $\mathbb{E}(\mathbf{Y}) = \mathbf{X}^T\boldsymbol{\beta}$ and $\mathbb{E}(\mathbf{Y}_0) = \mathbf{X}_0^T\boldsymbol{\beta}_0 = c_1\mathbf{X}_0^T\boldsymbol{\beta}$, and the corresponding marginal variances are $\text{Var}(\mathbf{Y}) = \tau^{-1}(\mathbf{A}\mathbf{Q}_\theta(\boldsymbol{\kappa})\mathbf{A}^T + \mathbf{I}_m)$ and $\text{Var}(\mathbf{Y}_0) = c_2\tau^{-1}(\mathbf{A}_0\mathbf{Q}_{\theta_0}(\boldsymbol{\kappa})\mathbf{A}_0^T + \mathbf{I}_{m_0})$, where $\mathbf{I}_{m_0}$ and $\mathbf{I}_m$ are corresponding identity matrices, and $m = 30$ and $m_0 = 50$ are the numbers of sampled train locations (Figure 23). c_1 and c_2 were varied over a grid $c_1 \in \{0.2, 1, 2\}$ and $c_2 \in \{0.5, 1, 100\}$ leading to nine possible combinations. For $c_1 = c_2 = 1$, the marginal expectation and variance of the historical response are

most consistent with the response of the current data compared to other combinations. Moreover, in other cases, the marginal properties of the historical data are either discounted or inflated, causing both data to lose consistency.

The columns of the design matrix $\mathbf{X}$ were formed from the covariates x_0, x_1, x_2, x_3, each drawn from a standard normal distribution, except $x_0 = 1$, acting as the intercept. The design matrix $\mathbf{X}_0$ was also obtained in a similar manner. The hyper-parameters $\nu = 1$ and $\phi = 5$ were fixed for both data and the true precision parameter $\tau = 1$. To investigate the performance of the models over various sample sizes, n and n_0 were varied over the set $\{1, 5, 10, 20, 50\}$ for each combination of c_1 and c_2. In addition to the borrowing schemes using spatial power prior (π^p), spatial commensurate prior (π^c), and scalable spatial commensurate prior (π^{sc}), this work investigated the "No-borrowing" (π^{no}) and "Full-borrowing" (π^{full}) models. The π^{full} and π^{no} act as baselines for the other models: π^{no} disregards the existence of historical data and π^{full} uses all the information in the historical data regardless of whether it is consistent with the current data or not. It is expected that models depending on the developed priors perform equivalently as π^{full} if the current and historical data are consistent. In addition, it is expected that models depending on the developed priors perform equivalently as π^{no} if the historical and current data were inconsistent. In other situations, it is expected that the borrowing techniques developed should outperform the extreme models.

In the estimation procedure, the chosen hyperparameter values were $a_\tau = 2, b_\tau = 1, a_{\tau_1} = 2, b_{\tau_1} = 1, \sigma_\beta^2 = 1000, \tau_x = 1, \sigma_\theta^2 = 10$, and the diagonal representation form of the commensurate matrix was adopted and the prior distribution for $\lambda_\beta \sim gamma(a_{\lambda_1}, b_{\lambda_1})$, $\lambda_\theta \sim gamma(a_{\lambda_2}, b_{\lambda_2})$ and uniform priors for the hyperparameters $a_{\lambda_1}, b_{\lambda_1}, a_{\lambda_2}$, and b_{λ_2}, which were updated using Metropolis Hasting algorithm. The model performances were evaluated using Conditional Predictive Ordinate (CPO) and the out-of-sample root mean square error (rMSE). The CPO for the current observation $\mathbf{Y}_j$ was approximated as

$$CPO_j = f(\mathbf{y}_j \mid \mathscr{D}_{-j}) \approx \left[\frac{1}{M} \sum_{k=1}^{M} \frac{1}{L(\mathbf{y}_j \mid \mathscr{D}_{-j}, \mathscr{X}_{(k)}, \tau_{(k)})} \right]^{-1}, \tag{5.24}$$

where $\mathscr{X}_{(k)}, \tau_{(k)}, k = 1, 2, ..., M$ are draws from the posterior distribution, $\mathscr{D}_{-j}$ is the training data excluding jth replication, and $L()$ is the likelihood of the current data. The higher the CPO_j the better the model fits the data. The summary $mlCPO = \frac{1}{n} \sum_{j=1}^{n} \log CPO_j$ was used to quantify the overall model fit performance in the training samples. Moreover, the out-of-sample rMSE was evaluated as

$$rMSE = \sqrt{\sum_{l=1}^{N} \frac{(\hat{\mathbf{Y}}_l - \mathbf{Y}_{out})^2}{N}}, \tag{5.25}$$

where $\hat{\mathbf{Y}}_l$ is the out-of-sample prediction on test locations, $\mathbf{Y}_{out}$ is the true sampled response at the test location. The prediction was repeated $N = 200$ to evaluate (5.25), where for each l, different data were simulated from the true model to evaluate the performance of predicting $\mathbf{Y}_{out}$.

It is expected that the rMSE equals zero, indicating that the closer to zero it is, the better the estimation model.

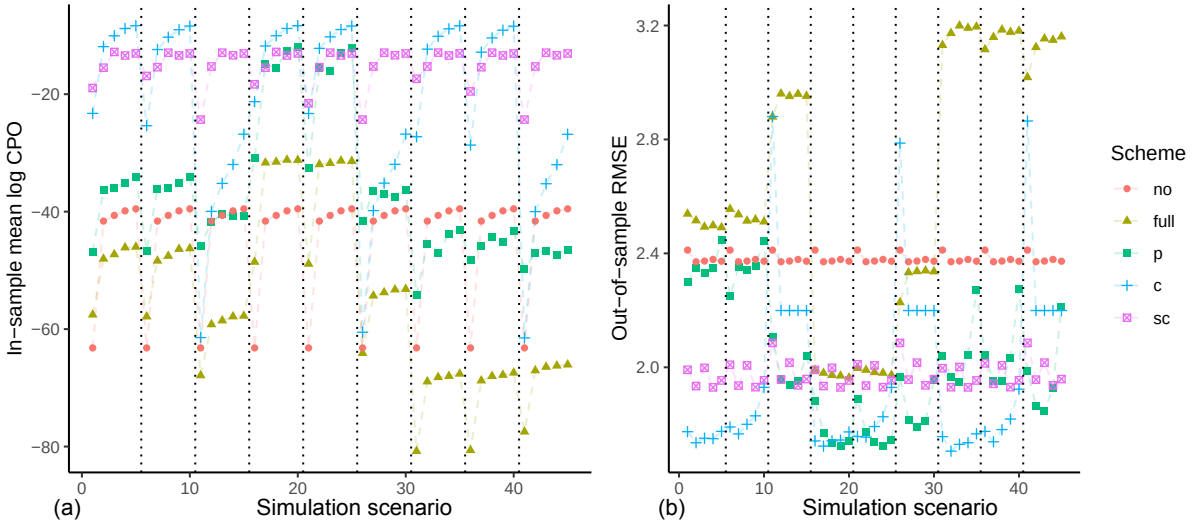


Figure 24 – (a) The in-sample mlCPO and (b) the out-of-sample rMSE for nine possible combinations and varying sample sizes. Each block, demarcated by a vertical dashed line, is a single combination of c_1 and c_2 with varying replication sizes $n_0 = n \in \{1,5,10,20,50\}$ in increasing order. Starting from the left, the combinations of c_1 and c_2 are (0.2,0.5), (0.2,1.0), (0.2,100.0), (1.0,0.5), (1.0, 1.0), (1.0,100.0), (2.0,0.5), (2.0,1.0), and (2.0,100.0) in the form (c_1, c_2).

Figure 24a & b showed the *mlCPO* of the samples used in the training data and the *rMSE* on the test locations respectively. The figure shows nine blocks of scenarios demarcated by dashed vertical lines, each representing a combination of (c_1, c_2). Within each block, $n_0 = n = 1, 5, 10, 20$ and 50. It can be seen that the performance of all the models improves as the replication at each training location increases, indicating consistency. Focusing on the π^{no} and π^{full}, the result shows that when both historical data and current data are inconsistent, π^{no} outperforms π^{full} and when both data are consistent π^{full} outperformed π^{no} in both in-sample and out-of-sample examples. Moreover, models involving π^{c} and π^{sc} prior distributions uniformly outperformed the baseline models (Figure 24a). The models involving π^{p} uniformly outperformed baseline models, except for scenarios where (c_1, c_2) were $(2.0, 0.5), (2.0, 1.0)$, and $(2.0, 100.0)$. This was directly linked to the sensitivity of the power parameter ω_0. A poor choice of the assumption of the prior distribution for ω_0 could lead to too much borrowing when it should not. Thus, a sensitivity analysis is recommended in real data applications when using π^{p}. Another possible way is to assume a simplex prior distribution for ω_0 due to its stability as compared to the beta distribution (LÓPEZ, 2013). In addition, Figure 24b shows that the models involving the proposed priors performed better than the baseline models in the out-of-sample prediction, except in a few cases for the power prior distribution which occurred in three scenarios, one each from the first three blocks. As expected the power and commensurate parameters increase as c_1 and c_2 draw closer to 1 in a finer grid (result not shown).

Figure 25 showed a posterior predictive of the models involving the proposed prior

distributions and baseline models at the scenario where $c_1 = 0.2$ and $c_2 = 1.0$ at $n_0 = n = 1$ in each train location. The true pattern of the current data exhibits bimodal peaks towards coordinates (0.3,0.3) and (0.8,0.8). The posterior predictive showed that models that involved the proposed prior distributions better captured the peaks and troughs in the spatial pattern of the response. While π^{full} could somewhat indicate both peaks, the π^{no} showed no indication of peak at (0.8.0.8).

This simulation showed the potential risk associated with pooling multiple spatial datasets, which is a commonly used technique for data integration. The historical dataset may dominate the current dataset, as shown by the "Full borrowing" baseline model (Figure 24), particularly when both datasets are inconsistent. However, the simulation study has shown that the developed models have the capability to prevent this dominance and filter the information provided by the historical data to improve model performance at a little additional computational cost. Thus the proposed models give the assurance that the worst-case scenario when combining geostatistical datasets is to attain baseline model performance.

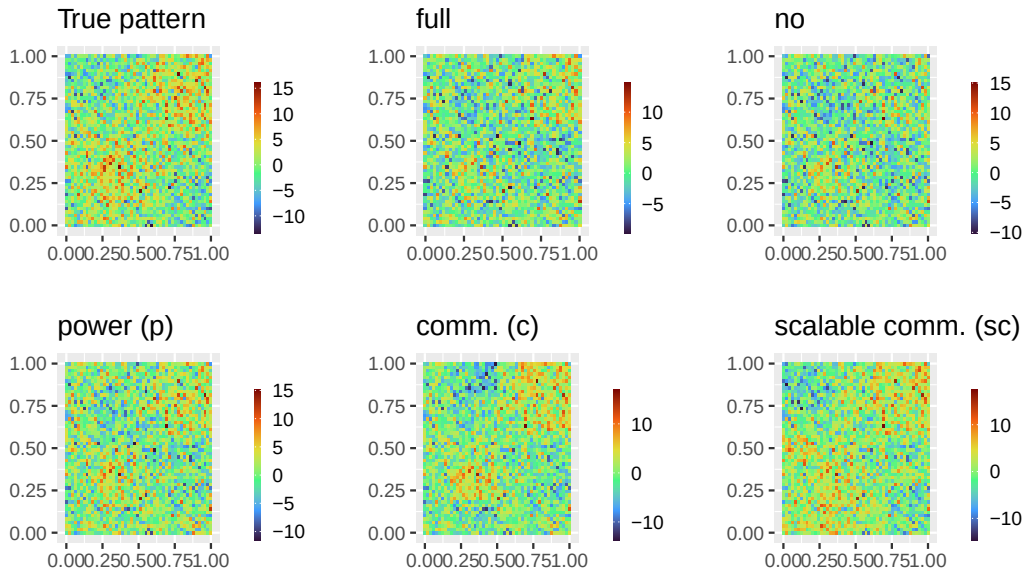


Figure 25 – The posterior predictive of the simulated pattern based on the developed priors and baseline models at a scenario ($c_1 = 0.2, c_2 = 1.0$) for $n_0 = n = 1$. The true pattern exhibits a bimodal characteristic, which was captured better by the models involving the developed priors.

5.6 Application

5.6.1 Data

This application aims to understand the spatial pattern of the subject's brain response to TMS pulses over the motor area after adjusting for possible confounding variables such as stimulation depth and the number of repetitive simulations. This pattern will help to identify the

Table 7 – Variables included in the model.

Variable	Type	Description
`Log peak-to-peak MEP`	Response (continuous)	The range summary of the output signal generated after TMS are delivered at each predefined location on the patient's motor cortex.
`Associate target`	Covariate (categorical)	The label differentiating the clusters of locations on the motor cortex.
`Replication`	Covariate (metrical)	Indicate at what TMS replicated pulse an MEP response was obtained. It ranges from 1-10. Replication equal to 3 means that the MEP response was obtained at the third TMS pulse in a given associate target. The time interval between the pulses were kept constant.
`Location`	Covariate (Coordinate)	The exact Cartesian coordinate where the response was obtained.

areas in the motor cortex that have optimal control over the movement of the right FDI muscle. In addition, this work seeks to make inferences on the suitable number of repetitive pulses necessary to produce desired responses. This study is crucial for advancing the understanding of the brain's motor function and has the potential to inform the development of more treatment options for motor disorders among patients.

Two TMS datasets of subject 1 (1R1 and 1R2) were available. Dataset 1R2 was derived from a formal TMS therapy session of subject 1, and dataset 1R2 is the current TMS dataset to be analyzed. It is believed that the 1R1 dataset contains useful information that could assist in estimating the spatial pattern of subject 1 response to TMS pulses. This work wishes to improve the estimation of the spatial pattern by incorporating 1R1 dataset in the Bayesian analysis of 1R2 dataset. Figure 26 shows the geographical locations where the current data (1R1) and historical data (1R2) were obtained. Each location indicates the coordinate where TMS pulses were delivered in the motor cortex. In total, $m = 225$ and $m_0 = 171$ locations were stimulated in the current and historical data respectively. These locations are clustered over the "Associate Targets" (AT). The target variable is `peak-to-peak motor-evoked potential (MEP)` recorded by EMG in microVolt (μV) and the response variable $\mathbf{Y} = \log$ `peak-to-peak MEP`. The linear covariate is the "`Associate target`" indicator (X_1), indicated as the square boxes in Figure 26. The metrical covariate is the "`replication`" (X_2), which indicates what TMS replicated pulse generated a given response at a given AT, and the spatial covariate is the location (X_3) on the motor cortex where the TMS pulses were delivered. Further descriptions of the covariates are shown in Table 7. Variable X_1 was introduced to account for the varying intercept across the associate targets, X_2 was introduced to understand the pattern of multiple TMS pulses on MEP at controlled intervals, and X_3 was introduced to understand the spatial pattern of MEP.

Figure 27 shows the box plots of $\mathbf{Y}$ for historical data (left) and the current data (right). The overall mean, median, and standard deviation of the MEP for the current data are 100.95μV, 37.37μV, and 123.96μV, and for the historical data are 36.39μV, 38.309μV, and 33.37μV. The prior models developed in this work were applied to understand the spatial patterns in the current data. The covariate X_1 was modeled linearly, X_2 was modeled non-linearly, and X_3 was modeled spatially. The hyper-parameters are $\sigma_\theta^2 = \sigma_\phi^2 = 10$, $\sigma_\beta^2 = 1000$, $a_\tau = a_{\tau_1} = 10, b_\tau = b_{\tau_1} = 2, \mu_\phi = \mu_\nu = 0, \sigma_\phi^2 = \sigma_\nu^2 = 10$, and $\tau_x = 10$. The chosen hyperparameter set produced the best model

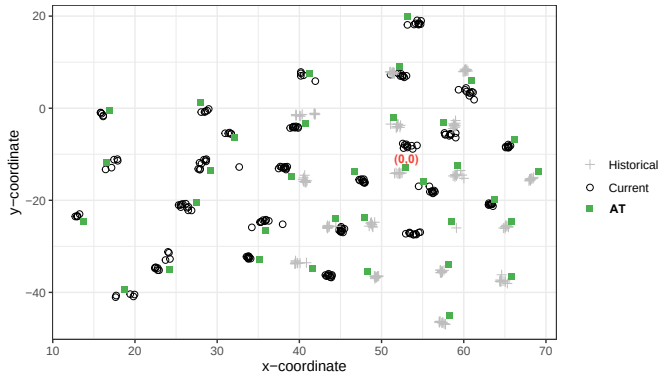


Figure 26 – Geographical locations where the current and historical data were observed. The "AT" indicates the associate target. The point (0,0) indicates the location of the primary motor cortex, M1, of the subject.

performance among various combinations. 10,000 iterations were conducted, and 2000 was dropped as burn-in samples. The remaining samples were used for inference. The acceptance rates of all the Metropolis-Hasting steps for all the models lie in the interval $(0.21, 0.57)$. The trace plot of the marginal log-likelihood was used to verify the chain convergence, and the Geweke test (GEWEKE *et al.*, 1991) was used to numerically verify the convergence of the chains for all the models. In addition, the spatial pattern of the exceedance probability was obtained as

$$P(\mathbf{Y} > y\mathbf{I} \mid \mathscr{D}) = \int P(\mathbf{Y} > y\mathbf{I} \mid \mathscr{X}, \tau, \mathscr{D}) \pi(\mathscr{X}, \tau \mid \mathscr{D}, \mathscr{D}_0) d(\mathscr{X}, \tau). \tag{5.26}$$

Here, $\pi(.)$ is the posterior distribution, $P(.)$ is the distribution of the current data, and $y = \log(400)$, chosen relative to the subject's motor threshold.

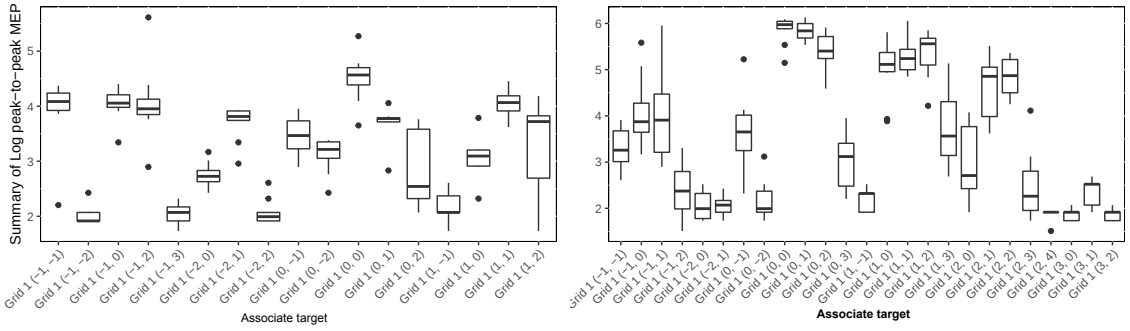


Figure 27 – Box plot of the historical data (left) and the current data (right) according to the "associate targets".

5.6.2 Results

The average computational run time for *no*, *full*, *p*, *sc*, and *c* are 4.20, 4.68, 6.64, 12.11, and 18.94 minutes accounting for 9%, 10%, 14%, 26%, and 41% of the total runtime respectively on an Intel(R) Core(TM) i7- @ 2.80GHz with 8GB RAM. The result showed about a 15% gain in

the model based on scalable commensurate prior compared with the one based on commensurate prior. Figure 28 shows the box plot of the CPO for each location and replication and mlCPO criteria for all the competing models, including the baseline models. The result showed that the median of the CPO and the mlCPO of the model based on the power, commensurate, and scalable commensurate priors are higher than the baseline models. Figure 29 shows the model residual quantiles for the models. The result shows that the proposed models had a better fit to the data compared with the baseline models. The result indicates that despite both data being obtained from a single subject at different TMS sessions, the distribution of the response to TMS stimulation significantly varies. This is evident as the full baseline model performed least among the models, indicating a lack of consistency in both data. However, the proposed model is able to harness the important information in the historical data to improve inference and prediction.

The posterior mean and the 95% credible interval of the power parameter ω are $0.182\,(0.02, 0.447)$. The estimate can be interpreted as the percentage (18.2%) amount of information borrowed from the historical data for inference, which indicates low similarity between the data distributions. For the model based on spatial commensurate prior, the posterior mean and the credible interval for the commensurate parameter for fixed and nonlinear effects combined, $\lambda_{\beta\psi}$, is $2.385\,(0.651, 5.381)$ and spatial effect, λ_θ, is $0.304\,(0.297, 0.307)$, indicating higher commensurate in the fixed and nonlinear effects. Similarly with the scalable spatial commensurate prior, the posterior mean and 95% credible interval for $\lambda_{\beta\psi}$ is $1.794\,(0.169, 7.735)$ and λ_θ is $7.615\,(2.149, 19.873)$ indicating higher commensurate in the spatial effect. The posterior mean and 95% credible interval for τ in the "full" baseline model is $2.207\,(1.905, 2.527)$, "no" baseline is $2.977\,(2.455, 3.552)$, and model based on the power prior model is $2.202\,(1.908, 2.519)$, the commensurate prior is $3.562\,(3.080, 4.083)$, and the scalable commensurate prior is $3.7865\,(3.182, 4.434)$.

Figure 30 shows the posterior densities of the effect of associate targets. The model based on the power prior distribution indicates the associate targets (0,0), (0,1), and (0,2) as promising targets in producing optimal MEP. The model based on commensurate prior indicates (1,2), (2,1), and (2,2), while the model based on scalable spatial commensurate prior indicates (0,0), (1,2), and (2,2). The "full" baseline model had a similar pattern as the power prior model and the "no" baseline model had a similar pattern as commensurate prior.

Figure 31 shows the non-linear effect of replication on the log peak-to-peak MEP for the proposed model and baseline models. The middle line is the posterior mean and the dashed lines are the 95% credible intervals. Overall, the result shows that log peak-to-peak MEP decreases as the TMS replication pulse increases. The minimum effect varies with the models. The effect pattern produced by the model based on power prior is similar to the one produced by the "full" baseline model, whereas, the effect produced by the model based on commensurate prior is similar to one produced by the "no" baseline model. The model based on scalable commensurate prior distribution has a distinct shape between replication 4 and 8, however, like

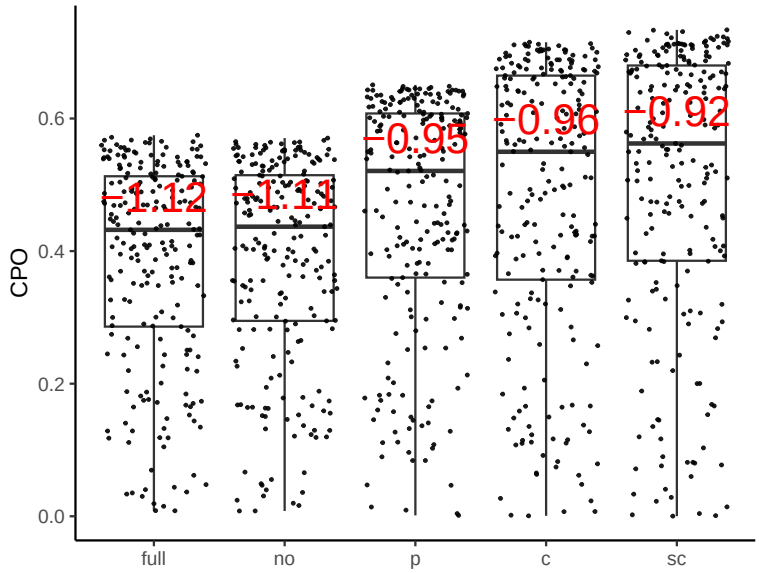


Figure 28 – The CPO of the proposed and baseline models. The dots indicate the CPO of a given model and the numerical figures are the mlCPO value.

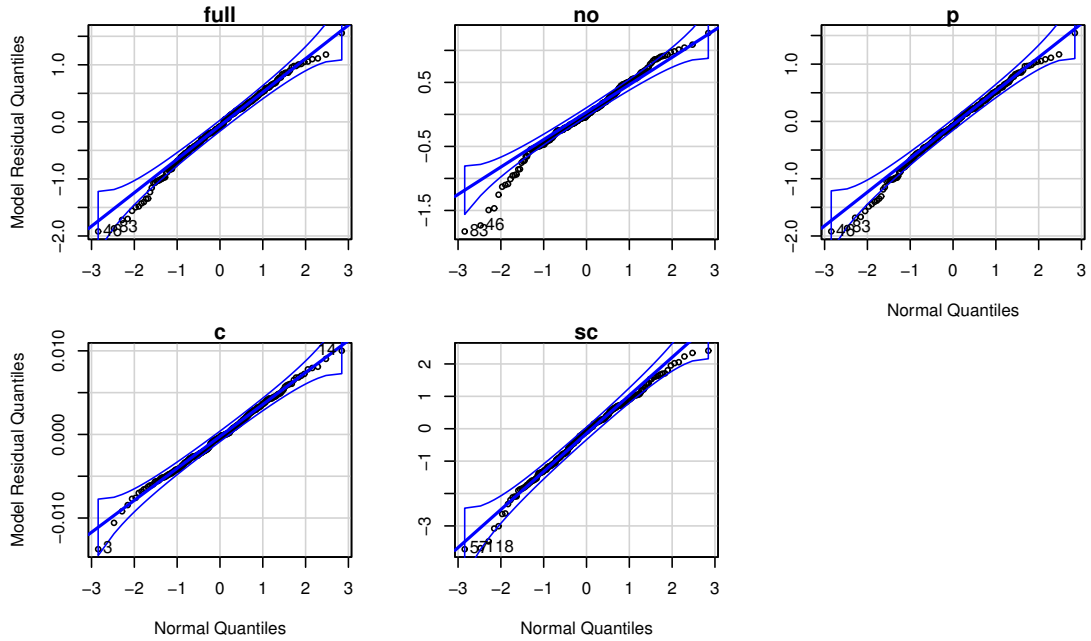


Figure 29 – The residual quantiles of the baseline and proposed models.

the commensurate model, it reached the minimum effect at replication 8. The result shows that the level of neuronal excitation initially decreases as the number of TMS pulses delivered increases. However, as more TMS pulses are administered to the motor cortex, the level of neuronal excitation gradually recovers.

Figure 32 shows the posterior prediction of the spatial effect over the whole spatial window after adjusting for other covariates using proposed models and baseline models. The posterior predicted spatial effect of the model based on power and commensurate prior distri-

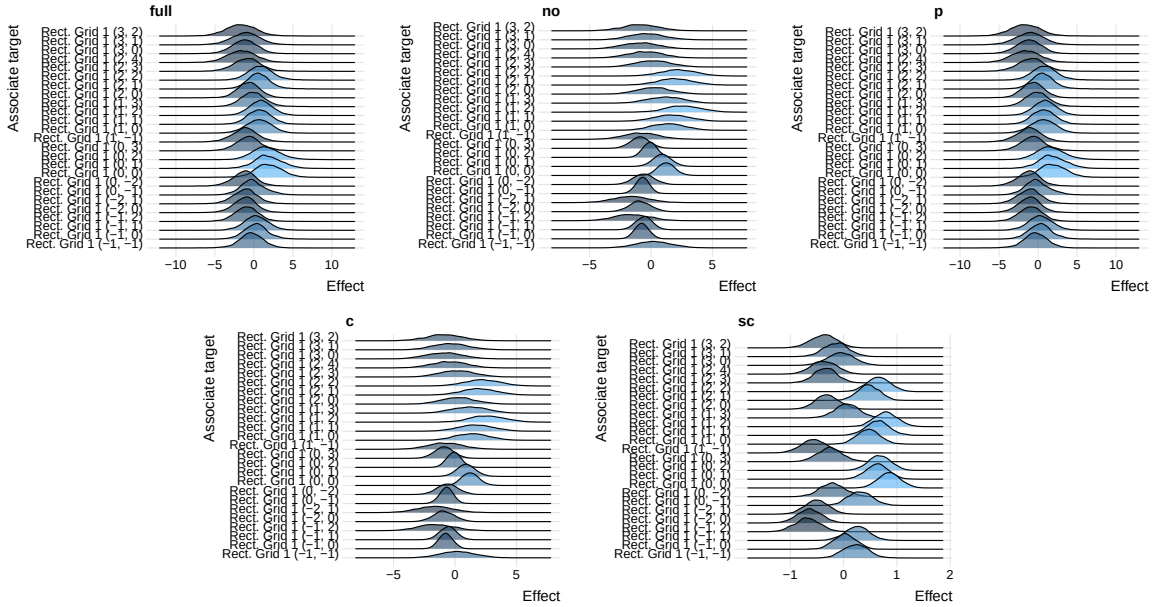


Figure 30 – Posterior density of the effects of associate targets on the log peak-to-peak MEP obtained based on the proposed models and the baseline models. The brighter the density, the higher the likelihood that the associate target will contain the locations that produce higher MEP.

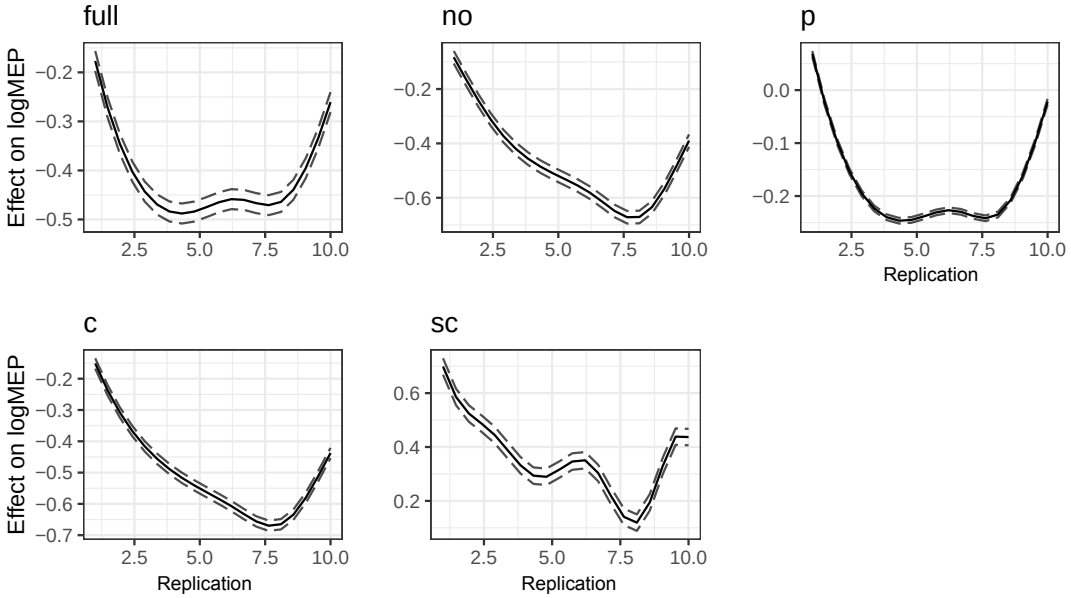


Figure 31 – The non-linear effect of replication on the MEP. The middle line indicates the posterior mean and the dashed lines indicate the 95% credible interval. On average, the result indicates that more repetitive stimulation lowers the peak-to-peak MEP until the 7th pulse.

butions have similar patterns with the baseline models. However, the scalable commensurate prior distribution showed a slightly different pattern. The locations with the high effect indicate candidate coordinates of the motor cortex that could produce high MEP. Among all the posterior predictive patterns, locations around coordinate (55,-10) consistently had a high spatial effect across all models, with a predicted MEP of 551.02μV for the model based on the power prior model, 672.68μV and 460.80μV for the model based on the commensurate and scalable com-

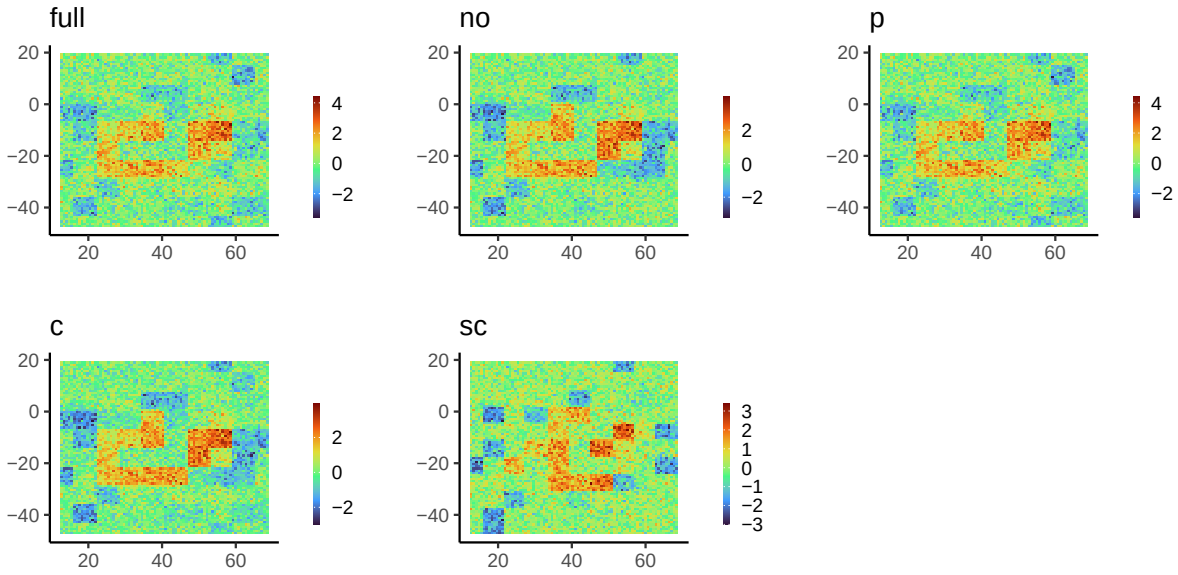


Figure 32 – Posterior prediction of the spatial effect based on proposed models and baseline models.

mensurate prior respectively. This location is referred to as the primary motor cortex, $M1$. Figure 33 shows the posterior exceedance predictive probabilities of a log MEP obtained at a given spatial location exceeds a threshold of $\log(400)$. This probability was computed over the spatial window using (5.26). The "no" baseline model and models based on commensurate and scalable commensurate priors exhibit bi-modal locations with relatively high exceedance probability. The exceedance probability was estimated to be at least 0.5 at these peaks, indicating the regions that respond mostly to TMS stimulation of the motor cortex. Particularly, location $M1$ has a 0.65 probability of producing above 400 μV MEP.

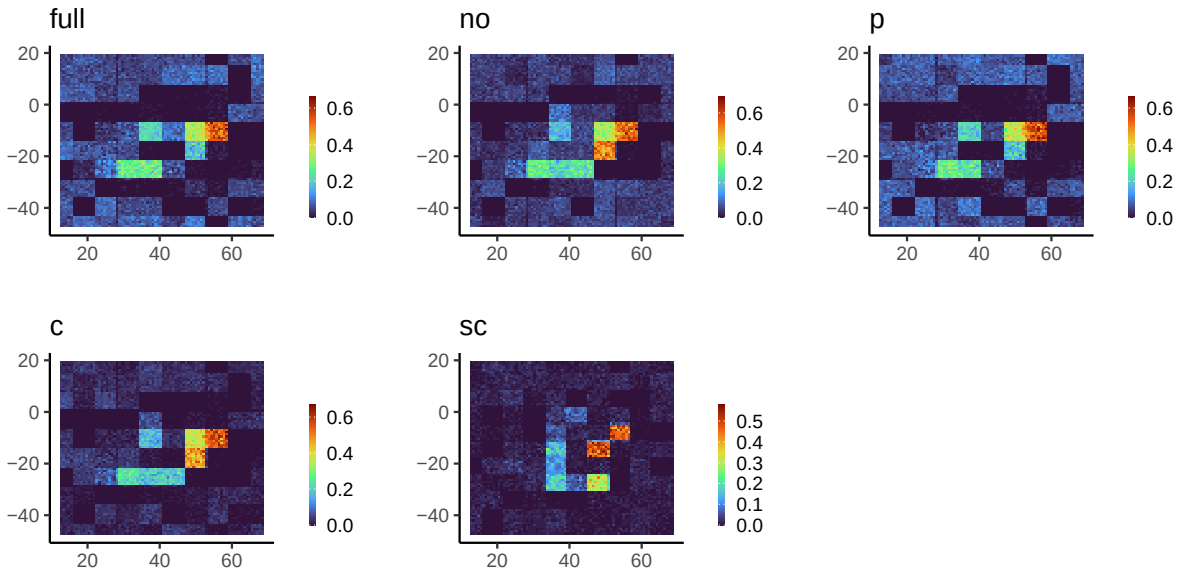


Figure 33 – The probability of an MEP obtained at a given location exceeding 400 μV.

5.6.3 Discussion of result

Although the TMS sessions that generate the historical and current data were conducted by the same practitioner on the same subject, there appears to be a weak relationship in the response pattern between the two TMS therapy sessions. This can be attributed to the dynamics of the brain which has the tendency to react to stimulation differently at different brain states as a result of intra-individual variability that may change in a time scale of minutes to hours (BERGMANN, 2018). The finding suggested that the location of the primary motor cortex is responsible for most movement of the FDI muscle, which is visually identified as it is predominantly located in the middle knee of the central sulcus (CS), formed by the posterior curvature of the precentral gyrus (PreCG) around the hand knob (SILVA *et al.*, 2020).

Moreover, findings revealed that the effect of stimulation replications on the peak-to-peak MEP at a given AT of the participant decreases non-linearly as the replication increases. That is, in the mesoscopic scale, the release of action potentials by the populations of neurons under stimulation lowers with an increase in stimulation replication at a given AT location, indicating that the number of depolarized neurons depreciates with the increase in stimulation, and thus, the aggregate neuronal activities over the mesoscopic level depreciate. It could be associated with the effect of repetitive suppression from the brain (KARIMINEZHAD *et al.*, 2020). However, the finding also indicates an ascension in the effect after replication 7. Findings also showed that there is heterogeneity across AT locations. It shows the differences in the capacity of neuronal activation at different ATs. The closer the ATs are to $M1$, the more probable they produce high MEP.

5.7 Conclusion

This work proposed three informative spatial prior distributions based on available historical geostatistical spatial data for performing Bayesian spatial analysis. Models based on these priors allow modeling linear, nonlinear, and spatial covariates in a single framework, and scales with the availability of many historical data. As demonstrated by the simulation studies, the proposed approach assures that the historical data do not obscure the true spatial pattern in the current data and provide evidence of improvement in the estimation compared with the naive pooling of these data. This technique is applicable to any geostatistical data and can be used conveniently to model large spatial data by partitioning the large data into batches, each considered as historical data. Each batch can be processed in parallel and then linked to one another using the scalable commensurate model. In the data analysis of the TMS data, though the same practitioner conducted both TMS sessions on the same subject, the response pattern was not consistent. This is because the brain is constantly changing and reacting differently to stimulation depending on its current state. However, the study did find that the primary motor cortex is responsible for most of the movement of the FDI muscle, and this location can be

identified visually and could facilitate TMS therapy and speed up the whole process in future stimulation sessions. The analysis provides a deeper understanding of the neural mechanisms underlying motor function and ultimately improves treatment options for individuals with motor impairments. Though the proposed approach only accommodates Gaussian processes, it can be extended to accommodate other spatial processes such as the student-t process (ORDOÑEZ *et al.*, 2023). The main drawback of the proposed methods is the computational complexity incurred by incorporating the historical data. Although this chapter suggests elegant ways to lower the complexity, future work may focus on faster estimation techniques such as variational inference.

Acquiring TMS data is expensive, and thus it is important to develop models that could capture all the possible variations in the data. While this chapter presents a parametric spatial model to better harness these variations, the spatial process was assumed to be stationary and Gaussian. Hence, the next chapter relaxes this assumption and proposed a nonparametric spatial model to further improve the existing model.

CHAPTER

6

MIXTURES OF DIRICHLET PROCESS PRIORS FOR SPATIAL MODELS

The spatial pattern of subjects' responses to TMS stimulation possesses a complex structure that has the tendency to be modulated by stimulation in different brain states as studies have shown intra-individual variability in a time scale of minutes. Thus, it is challenging to use the spatial pattern for one subject to gain insight into the improvement of a TMS treatment experience of a different subject. It is believed that the recorded responses to TMS stimulation may be due to mixtures of latent spatial patterns that are common to every subject and those particular to each subject. Hence, the goal of this chapter is to distinguish these latent processes in the TMS datasets so that the common pattern that is consistent across subjects could give the required insight into improving the treatment experiences of different subjects. This chapter developed a mixture of Dirichlet process models to arrive at a non-stationary and non-Gaussian spatial model that statistically combines multiple related geostatistical data within a single modeling framework. The model creates a probabilistic informed pathway for information sharing among multiple spatial sub-models to improve inferential and predictive performances. The proposed model was adopted to analyze subjects 1R1, 2, and 3 TMS datasets.

6.1 Spatial model with single Dirichlet process prior

This section first presents a nonparametric spatial model for analyzing a single dataset and laid the statistical foundation that will be adopted in the subsequent sections for modeling multiple datasets.

6.1.1 Model specification

Suppose for spatial locations $s_1, s_2, ..., s_I \in \mathscr{G}$, $\mathbf{Y}_t = (Y_{1t}, Y_{2t}, ..., Y_{It})^T$ is the corresponding random response vector for each replicate $t = 1, 2, 3, ..., n$ and the probability distribution of $\mathbf{Y}_t$

belong to a class of distributions $\mathscr{F}$, such that $\mathbf{Y}_t \sim f \in \mathscr{F}$. $Y_{it} \in \mathbb{R}$ is the response observed at location i in replication t. In relation to TMS data, s_i represents the grid cell on the scalp where TMS pulses were delivered, and Y_{it} is the corresponding recorded peak-to-peak motor-evoked potential (MEP) at that location in replication t. The goal is to model the underlying spatial process that generates data $\mathbf{Y}_t$. Y_{it} was assumed to have a conditional independent Gaussian distribution, so that, $\mathbf{Y}_t \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}_t \sim N(\boldsymbol{\mu}_t, \tau_t^{-1}\mathbf{I})$, where τ_t is the precision at replication t and $\boldsymbol{\mu}_t = (\mu_{1t}, \mu_{2t}, ..., \mu_{It})^T \in \mathbb{R}^I$ is defined by a structural additive predictor. That is, $\mu_{it} = \mathbf{x}_{it}^T\boldsymbol{\beta} + g(z_{it}, \boldsymbol{\phi}) + h(s_i, \boldsymbol{\theta}_t)$, where μ_{it} is the conditional expected value of the response at location s_i in replication t. $\boldsymbol{\beta}$ is a $p \times 1$ regression coefficient of covariate vector $\mathbf{x}_{it}$, $g(z_{it}, \boldsymbol{\phi})$ is an unknown nonlinear function with $R \times 1$ parameter vector $\boldsymbol{\phi}$ introduced to describe the smooth function of metrical covariate z_{it}, and $h(s_i, \boldsymbol{\theta}_t)$ is a spatial effect at location s_i in replication t, governed by a spatial process denoted by a $L \times 1$ vector $\boldsymbol{\theta}_t$. The covariate $\mathbf{x}_{it}$ could represent the angular readings of the TMS coil positioning and the intensity, and z_{it} could be the coil positioning error. Both effects were introduced in the model to avail the opportunity to adjust for TMS technical noise.

Let $\boldsymbol{\theta}_t = (\theta_{1t}, \theta_{2t}, ..., \theta_{Lt})^T$, where θ_{lt} is the spatial parameter corresponding to lth fixed spatial knot location $v_l \in \mathscr{G}$ in replication t. The approximate of the spatial effect function is given by $h(s_i, \boldsymbol{\theta}_t) = \mathbf{a}_i\boldsymbol{\theta}_t$, where $\mathbf{a}_i = (a_i^{(1)}, a_i^{(2)}, ..., a_i^{(L)})$ is a deterministic vector and $a_i^{(l)}$ is obtained by approximating h using a Gaussian kernel basis function with a bandwidth ψ evaluated at data location s_i and spatial field location v_l. That is,

$$a_i^{(l)} = \exp\left(-\frac{d(s_i, v_l)^2}{\psi^2}\right), \psi > 0, \tag{6.1}$$

where $d(s_i, v_l)$ is the euclidean distance between data location s_i and knot v_l. Given the kernel (6.1), Reich, Bandyopadhyay and Bondell (2013) proved that for a fixed number of knots L and bandwidth ψ, and any continuous marginal distribution for Y_{it}, there exists a density function for θ_{it} that leads to that marginal distribution of Y_{it}. Hence, this representation leads to a valid probability distribution. The approximation $h(s_i, \boldsymbol{\theta}_t) = \mathbf{a}_i\boldsymbol{\theta}_t$ is similar in spirit to the Karhunen–Loeve decomposition of stochastic processes (SIMPSON *et al.*, 2016). Equation (6.1) provides a structure for combining the analysis of multiple geostatistical spatial data with different spatial windows.

The function g is approximated semi-parametrically by a spline function to model the nonlinear relationship of metrical covariates (ELIERS; MARX, 1996). For a metrical covariate z_{it} in the model, $g(z_{it}, \boldsymbol{\phi}) = \sum_{r=1}^{R} B_r^d(z_{it})\phi_r$, where $B_r^d(z_{it})$ is the rth spline basis function of degree d, evaluated at z_{it} and ϕ_r is the corresponding coefficient. Particularly in this work, $R = d + K_d + 1$, where K_d is the number of equidistant knots and degree $d = 3$, giving rise to a cubic B-spline. The values of the basis functions are obtained using Equation (5.3). Let $\boldsymbol{\phi} = (\phi_1, \phi_2, ..., \phi_R)^T$ and $\mathbf{b}_{it} = (B_1^d(z_{it}), B_2^d(z_{it}), ..., B_R^d(z_{it}))$, then $g(z_{it}, \boldsymbol{\theta}) = \mathbf{b}_{it}\boldsymbol{\phi}$ in vector form.

Therefore, the structural additive form for all the spatial locations at replication t is given

as

$$\boldsymbol{\mu}_t = \mathbf{X}_t^T \boldsymbol{\beta} + \mathbf{B}_t \boldsymbol{\phi} + \mathbf{A}\boldsymbol{\theta}_t, \tag{6.2}$$

where each row of $\mathbf{X}_t^T$, $\mathbf{B}_t$ and $\mathbf{A}$ is the corresponding $\mathbf{x}_{it}^T$, $\mathbf{b}_{it}$ and $\mathbf{a}_i$ vectors respectively. The generating model is given as

$$\begin{aligned}
\mathbf{Y}_t \mid \boldsymbol{\beta}, \boldsymbol{\phi}, \boldsymbol{\theta}_t &\sim N(\boldsymbol{\mu}_t, \tau_t^{-1}\mathbf{I}), \\
\boldsymbol{\mu}_t &= \mathbf{X}_t^T \boldsymbol{\beta} + \mathbf{B}_t \boldsymbol{\phi} + \mathbf{A}\boldsymbol{\theta}_t, \\
\boldsymbol{\theta}_t &\sim H, \\
H &\sim DP(\alpha, G), \\
(\boldsymbol{\beta}, \boldsymbol{\phi}) &\sim \pi(\boldsymbol{\beta}, \boldsymbol{\phi}), \\
\boldsymbol{\Lambda} &\sim \pi(\boldsymbol{\Lambda}),
\end{aligned} \tag{6.3}$$

where DP is a Dirichlet process with concentration parameter α and base probability distribution G, $\pi(\boldsymbol{\beta}, \boldsymbol{\phi})$ is the joint prior distribution for the linear and nonlinear effects and $\pi(\boldsymbol{\Lambda})$ is the prior distribution of the vector of hyperparameters, $\boldsymbol{\Lambda}$. In Equation (6.3), $\boldsymbol{\theta}_t$ incorporates the spatial dependence of $\mathbf{Y}_t$ through the base distribution G. The prior models are discussed specifically in Subsection 6.1.2.

6.1.2 Prior model

The main model parameters are the $\boldsymbol{\beta}$, $\boldsymbol{\phi}$, and $\boldsymbol{\theta}_t$. A Gaussian prior distribution with zero mean and a precision matrix $\mathbf{Q}_\beta = \tau_\beta \mathbf{I}_p$ was assumed for $\boldsymbol{\beta}$, where $\mathbf{I}_p$ is an identity matrix of size p. A random walk model I (RW1) with a Gaussian error as prior distribution was assigned for $\boldsymbol{\phi}$. In general, higher orders of random walks and autoregressive models can be considered (LANG; BREZGER, 2004). Specifically, let $\boldsymbol{\phi}$ be a vector whose elements correspond to the parameters of the cubic spline basis function. Then the RW1 model is given as

$$\phi_r = \phi_{r-1} + e_r,\ \phi_1 = e_1,\ e_r \sim N(0, \tau_\phi^{-1}), r = 1, 2, ..., R. \tag{6.4}$$

Solving recursively, $\boldsymbol{\phi} = \mathbf{Re}$, $\mathbf{e} \sim N(\mathbf{0}, \tau_\phi^{-1}\mathbf{I}_R)$, $\mathbf{I}_R$ is an identity matrix of size R, and the matrix $\mathbf{R} = (I_{rr'})$, where $I_{rr'} = 1$ if $r \leq r'$ and 0 otherwise. Therefore, the prior distribution on $\boldsymbol{\phi} \sim N(\mathbf{0}, \mathbf{Q}_\phi^{-1})$, $\mathbf{Q}_\phi^{-1} = \tau_\phi^{-1}\mathbf{R}\mathbf{R}^\mathbf{T}$. The prior distribution is rank one deficient, and thus, a sum-to-zero constraint on the effect is imposed. The nonlinear effect with a random walk prior was considered to determine how the angular variation from a reference point in the coil positioning on the motor cortex affects the elicited MEP.

A random prior distribution H is assumed for $\boldsymbol{\theta}_t$, which is modeled non-parametrically. That is $\boldsymbol{\theta}_t \sim H$ and $H \sim DP(\alpha, G), t = 1, 2, ..., n$, where $DP(\alpha, G)$ is a Dirichlet process with concentration parameter α and a base distribution G. Since $\boldsymbol{\theta}_t$ is linked to Y_{it} through f, the model is referred to as a DP mixture with kernel f. Sethuraman (1994) provided a stick-breaking

representation of the DP. Recall that $DP(\alpha, G)$ could be represented as $\sum_{b=1}^{\infty} \omega_b \delta_{\boldsymbol{\theta}_t^{(b)}}$. δ_a denotes a point mass at a and $\boldsymbol{\theta}_t^{(b)} \sim G$. Let $\omega_1 = a_1$ and $a_b \sim Beta(1,\alpha)$, $b = 2,3,...$ independently, then, the stick-breaking representation follows as $\omega_b = a_b \prod_{k=1}^{b-1}(1-a_k)$. It is easy to arrive at the probability distribution for $\boldsymbol{\theta}_t$ following this representation. For example, suppose that the base distribution G is a spatial process, then draw samples $\boldsymbol{\theta}_t^{(b)}$ from G and independently draw a_bs from $Beta(1,\alpha)$ and evaluate ω_b. Repeating the procedure for several b; the collection $\{(\omega_1, \boldsymbol{\theta}_t^{(1)}), (\omega_2, \boldsymbol{\theta}_t^{(2)}), ...\}$ gives rise to the desired probability distribution, however discrete. The base distribution G helps to retain the spatial structure. The specified DP is thus referred to as Dependent Dirichlet Process (DDP).

Integrating out H from $\boldsymbol{\theta}_t \sim H$ and $H \sim DP(\alpha, G)$ using the Pólya Urn scheme (BLACKWELL; MACQUEEN, 1973) the conditional prior distribution becomes

$$\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{-t} \sim \frac{1}{n-1+\alpha} \sum_{k \neq t} \delta_{\boldsymbol{\theta}_k} + \frac{\alpha}{n-1+\alpha} G, \tag{6.5}$$

where $\boldsymbol{\theta}_{-t} = (\boldsymbol{\theta}_1, ...\boldsymbol{\theta}_{t-1}, \boldsymbol{\theta}_{t+1}, ..., \boldsymbol{\theta}_n)$. The conditional prior distribution given in (6.5) simply means that there is $1/(n-1+\alpha)$ probability that $\boldsymbol{\theta}_t$ assume value from $\boldsymbol{\theta}_{-t}$ and $\alpha/(n-1+\alpha)$ it assumes value from the base distribution G. It implies that as the concentration parameter α increases, $\boldsymbol{\theta}_t$ is more likely to follow G, hence, new clusters are created.

A stationary Gaussian process centered at zero with $L \times L$ variance-covariance matrix $\boldsymbol{\Sigma}$ was assumed for G. $\boldsymbol{\Sigma}$ is governed by a stationary Matérn covariance function given as:

$$\boldsymbol{\Sigma}_{(l,l')} = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \left(\frac{\sqrt{2\nu} d(v_l, v_{l'})}{\kappa} \right)^{\nu} K_{\nu} \left(\frac{\sqrt{2\nu} d(v_l, v_{l'})}{\kappa} \right), \tag{6.6}$$

where $\sigma^2, \kappa, \nu > 0$, and $K_{\nu}(.)$ is the modified Bassel function of the second kind with an order of differentiability ν. A particular case of the Matérn covariance function is the exponential covariance function for $\nu = 1/2$. More generally, the limiting function as ν approaches infinity is the squared exponential covariance function. The κ parameter controls the spatial range, while ν controls the smoothness, and σ^2 is the partial sill.

Besides the main model effect's parameters, there are associated hyperparameters $\boldsymbol{\Lambda}$ =(α, τ_t, τ_ϕ, τ_β, σ^2, κ, ν)^T in the model hierarchy. Throughout this work, $\alpha \sim Gamma(a_\alpha, b_\alpha)$ was assumed for the concentration parameter, which guarantees a faster technique of posterior update (ESCOBAR; WEST, 1995). Moreover, $\tau_t \sim Gamma(a_\tau, b_\tau)$, $\tau_\phi \sim Gamma(a_\phi, b_\phi)$, and $\tau_\beta \sim Gamma(a_\beta, b_\beta)$ for known parameters. A lognormal prior distribution was assigned to $\kappa \sim \log normal(\mu_\kappa, \sigma_\kappa^2)$ and $\nu \sim \log normal(\mu_\nu, \sigma_\nu^2)$. A possible prior distribution for σ^2 is the inverse gamma.

6.2 Spatial model with mixture of Dirichlet process priors

Combining sub-spatial models within a single modeling framework is important, especially in situations where the data generation process possesses shared and specific spatial characteristics. Consider for example the TMS mapping dataset, which is illustrated in Figure 34. The top panel shows the TMS data of three subjects corresponding to $\mathscr{G}_1, \mathscr{G}_2$, and $\mathscr{G}_3$ whose location of the primary motor cortex (M1) is fixed and known, and it is visually located over a subject's scalp. For each subject $j\,(j = 1,2,3)$, practitioners usually form a spatial window denoted by $\mathscr{G}_j$ around M1. The TMS responses $(\mathbf{y}_{11}, \mathbf{y}_{12}, \mathbf{y}_{13})$, $(\mathbf{y}_{21}, \mathbf{y}_{22})$, and $(\mathbf{y}_{31}, \mathbf{y}_{32}, \mathbf{y}_{33})$ are recorded respectively for subject $1,2$, and 3 within the windows for each subject. $\mathbf{y}_{ji} = (y_{ji1}, y_{ji2}, ..., y_{jiT_j})$, and y_{jit} is the recorded response at replication t, spatial location i for subject j. The spatial pattern of subjects' responses to TMS stimulation possesses a complex structure that has the tendency to be modulated by stimulation in different brain states as studies have shown intra-individual variability in a time scale of minutes (BERGMANN, 2018). Thus, it is challenging to use the spatial pattern for one subject to gain insight into the improvement of a TMS treatment experience of a different subject. It is believed that the recorded responses to TMS stimulation may be due to a spatial pattern that is common to every subject and those particular to each subject. Thus, the common pattern, which is similar across subjects could give the required insight into improving treatment experiences of different subjects.

The bottom panel of Figure 34 illustrates an imaginary overlapping of these spatial windows with reference to the primary motor cortex. In the figure, $\mathscr{G}$ denotes the window that encloses all the sub-windows. It is assumed that combining the shared spatial pattern in $\mathscr{G}$ with the subject-specific spatial patterns leads to the recorded responses. For example, it is possible that the recording leading to response y_{111} in $\mathbf{y}_{11}$ was elicited by a process shared across subjects, whereas y_{112} could be a response elicited by a subject-specific process as a result of a rapid change in brain state during a TMS therapy session. Hence, the goal of this section is to distinguish these latent processes.

Suppose there are J subjects that participated in the experiment. Let $\mathbf{Y}_{jt} = (Y_{j1t}, Y_{j2t}, ..., Y_{jI_jt})^T$ be a vector of the responses at the spatial locations $(s_{j1}, s_{j2}, ..., s_{jI_j})$ with corresponding spatial field $\boldsymbol{\theta}_{jt} = (\theta_{j1t}, \theta_{j2t}, ..., \theta_{jL_jt})^T$ in the tth replication for the jth subject. The proposed

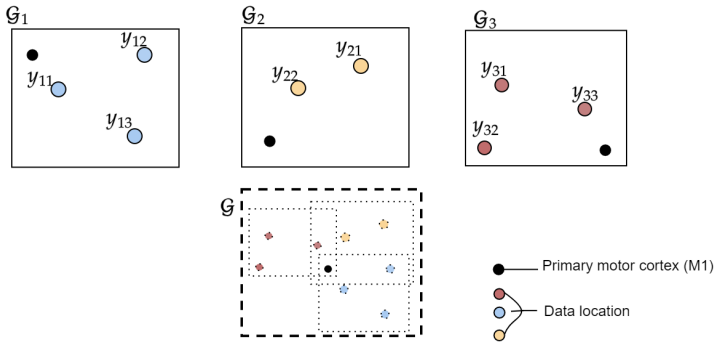


Figure 34 – The top panel illustrates three TMS brain mapping point referenced data for three different subjects. The data points $\{\mathbf{y}_{11},\mathbf{y}_{12},\mathbf{y}_{13}\}$, $\{\mathbf{y}_{21},\mathbf{y}_{22}\}$, and $\{\mathbf{y}_{31},\mathbf{y}_{32},\mathbf{y}_{33}\}$ correspond to the observed peak-to-peak of the motor evoked potential response signal in spatial window $\mathcal{G}_1,\mathcal{G}_2$, and $\mathcal{G}_3$ formed for each subject during TMS therapy session. The bottom panel illustrates an imaginary overlap of these windows with reference to the primary motor cortex for each subject.

hierarchical model follows as:

$$
\begin{aligned}
\mathbf{Y}_{jt} \mid \boldsymbol{\beta}_j, \boldsymbol{\theta}_{jt}, \boldsymbol{\phi}_j &\sim N(\boldsymbol{\mu}_{jt}, \tau_{jt}^{-1}\mathbf{I}),\\
\boldsymbol{\mu}_{jt} &= \mathbf{X}_{jt}^T\boldsymbol{\beta}_j + \mathbf{B}_{jt}\boldsymbol{\phi}_j + \mathbf{A}_j\boldsymbol{\theta}_{jt},\\
\boldsymbol{\theta}_{jt} &\sim \varepsilon_j H_0 \mathbb{I}_{(w_{jt}=1)} + (1-\varepsilon_j) H_j \mathbb{I}_{(w_{jt}=0)},\\
H_j &\sim DP(\alpha_j, G_j),\\
H_0 &\sim DP(\alpha_0, G_0),\\
(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j) &\sim N(\mathbf{0}, \tau_j^{-1}\mathbf{Q}_{\beta_j}^{-1}) N(\mathbf{0}, \tau_j^{-1}\mathbf{Q}_{\phi_j}^{-1}),\\
\boldsymbol{\Lambda} &\sim \pi(\boldsymbol{\Lambda}), t = 1,2,...,T_j, j = 1,2,...,J,
\end{aligned}
\tag{6.7}
$$

where G_j and G_0 are Gaussian spatial processes for the subject-specific and shared prior distributions. In (6.7), $\boldsymbol{\Lambda}$ is a vector of all hyperparameters with joint density $\pi(\boldsymbol{\Lambda})$, and w_{jt} is a latent variable. With probability ε_j, $w_{jt} = 1$ and with probability $1-\varepsilon_j$, $w_{jt} = 0$. It is assumed that $s_{ji} \in \mathcal{G}_j$, $v_l \in \mathcal{G}$ and $\cup\mathcal{G}_j \subset \mathcal{G}$, as shown in Figure 34. The mixture of DP priors in (6.7) allows spatial information sharing between multiple sub-models. Moreover, the model implies that the window where data are observed for different spatial data do not necessarily have to overlap but be in the same global spatial domain $\mathcal{G}$. $\varepsilon_j \in [0,1]$ is the weight parameter that can be interpreted as a measure of the marginal relationship between data j and others. The proposed model is flexible and it encompasses several sub-spatial models. For example, setting $\varepsilon_j = \varepsilon \forall j$ leads to a simpler model that allows information sharing at a constant rate across sub-models. Further, at the extreme value $\varepsilon = 0$, it implies that all the studies are independent and share no spatial relationship, whereas, at the other extreme, $\varepsilon = 1$, it implies that the spatial pattern of all the studies are perfectly related and are assumed to be generated by the same process or originated from the same population. The proposed model leads to a mixture of a non-Gaussian and non-stationary spatial process and allows information sharing between subjects. In the model, the linear and non-linear effects are not involved in the mixing of the observed response and

can be used to adjust for subject-specific effects and noise introduced by differences in TMS therapeutic procedures across subjects. This assumption is valid since data were obtained from subjects with different independent therapies. For example, TMS stimulation intensity was set on a given percentage of a subject-specific resting motor threshold, which varies for different patients. However, the proposed model can be easily extended to accommodate the linear and nonlinear effects in the mixing.

Conditioning on the latent variable w_{jt}, the prior model for the spatial effects in Equation 6.7 reduces to $\boldsymbol{\theta}_{jt} \mid w_{jt} = 1 \sim H_0$, and $\boldsymbol{\theta}_{jt} \mid w_{jt} = 0 \sim H_j$, with w_{jt} having a Bernoulli probability distribution with success probability ε_j. Let $\boldsymbol{\Delta}_{0j} = \{\boldsymbol{\theta}_{jt} : w_{jt} = 1\}$, $\boldsymbol{\Delta}_j = \{\boldsymbol{\theta}_{jt} : w_{jt} = 0\}$, and $\boldsymbol{\Delta}_{0j}^{-t}$ be $\boldsymbol{\Delta}_{0j}$ but excluding $\boldsymbol{\theta}_{jt}$, and $\boldsymbol{\Delta}_j^{-t}$ be $\boldsymbol{\Delta}_j$ but excluding $\boldsymbol{\theta}_{jt}$. Furthermore, let $\mathscr{N}_j = \{(jt) : w_{jt} = 1\}$, $n_{0j} = |\mathscr{N}_j|$, and $n_j = |\mathscr{N}_j^c|$ for data j. Integrating out H_j and H_0 from (6.7), then

$$\begin{aligned}
[\boldsymbol{\theta}_{jt} \mid w_{jt} = 0, \boldsymbol{\Delta}_j^{-t}] &\sim \frac{1}{n_j - 1 + \alpha_j} \sum_{k \neq t} \delta_{\boldsymbol{\theta}_{jk}} + \frac{\alpha_j}{n_j - 1 + \alpha_j} G_j, (jk) \in \mathscr{N}_j^c, \\
[\boldsymbol{\theta}_{jt} \mid w_{jt} = 1, \boldsymbol{\Delta}_{0j}^{-t}] &\sim \frac{1}{n_{0j} - 1 + \alpha_0} \sum_{k \neq t} \delta_{\boldsymbol{\theta}_{jk}} + \frac{\alpha_0}{n_{0j} - 1 + \alpha_0} G_0, (jk) \in \mathscr{N}_j,
\end{aligned} \tag{6.8}$$

for $j = 1, 2, ..., J$. Equation (6.8) implies that conditioning on $w_{jt} = 1$, $\boldsymbol{\theta}_{jt}$ is shared among all the data, whereas $w_{jt} = 0$, $\boldsymbol{\theta}_{jt}$ is specific to data j. Therefore, the joint distribution for $(\boldsymbol{\theta}_{jt}, w_{jt})$ follows as

$$\begin{aligned}
[\boldsymbol{\theta}_{jt}, w_{jt} = 0 \mid \boldsymbol{\Delta}_j^{-t}] &\sim \frac{(1 - \varepsilon_j)}{n_j - 1 + \alpha_j} \sum_{k \neq t} \delta_{\boldsymbol{\theta}_{jk}} + \frac{\alpha_j (1 - \varepsilon_j)}{n_j - 1 + \alpha_j} G_j, (jk) \in \mathscr{N}_j^c, \\
[\boldsymbol{\theta}_{jt}, w_{jt} = 1 \mid \boldsymbol{\Delta}_{0j}^{-t}] &\sim \frac{\varepsilon_j}{n_{0j} - 1 + \alpha_0} \sum_{k \neq t} \delta_{\boldsymbol{\theta}_{jk}} + \frac{\alpha_0 \varepsilon_j}{n_{0j} - 1 + \alpha_0} G_0, (jk) \in \mathscr{N}_j.
\end{aligned} \tag{6.9}$$

Restraining all $\varepsilon_j = \varepsilon$, the joint distribution (6.8) is straightforward to adjust to allow an equal sharing rate across all datasets. That is, the prior distribution becomes

$$\begin{aligned}
[\boldsymbol{\theta}_{jt}, w_{jt} = 0 \mid \boldsymbol{\Delta}_j^{-t}] &\sim \frac{(1 - \varepsilon)}{n_j - 1 + \alpha_j} \sum_{k \neq t} \delta_{\boldsymbol{\theta}_{jk}} + \frac{\alpha_j (1 - \varepsilon)}{n_j - 1 + \alpha_j} G_j, (jk) \in \mathscr{N}_j^c, \\
[\boldsymbol{\theta}_{jt}, w_{jt} = 1 \mid \boldsymbol{\Delta}_0^{-jt}] &\sim \frac{\varepsilon}{n_0 - 1 + \alpha_0} \sum_{(uk) \neq jt} \delta_{\boldsymbol{\theta}_{uk}} + \frac{\alpha_0 \varepsilon}{n_0 - 1 + \alpha_0} G_0, (uk) \in \bigcup_{j=1}^{J} \mathscr{N}_j,
\end{aligned} \tag{6.10}$$

where $n_0 = \sum_{j=1}^{J} n_{0j}$ and $\boldsymbol{\Delta}_0^{-jt}$ is $\bigcup_{j=1}^{J} \boldsymbol{\Delta}_{0j}$ but excluding $\boldsymbol{\theta}_{jt}$.

G_js and G_0 were assumed to be stationary Gaussian spatial processes centered at zero with variance-covariance matrix, $\lambda_j^{-1} \boldsymbol{\Sigma}(\nu_j, \kappa_j)$ and $\lambda_0^{-1} \boldsymbol{\Sigma}(\nu_0, \kappa_0)$ respectively. These matrices are constructed from a Matérn covariance function in (6.6) based on known common knots locations in $\mathscr{G}$. Given these specifications, $\mathbf{Y}_{jt}$ is a non-stationary and non-Gaussian spatial process, and thus, the model is capable of capturing the processes that naturally generate the TMS datasets.

Based on the above specification and from (6.6), $\sigma_j^2 = \frac{1}{\lambda_j \tau_\theta}$ for a fixed common $\tau_\theta > 0$ for all j. Similar prior distributions to those discussed in Section 6.1.2 for $(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \alpha_j, \tau_{\beta_j}, \tau_{\phi_j}, \kappa_j, \nu_j)$ were utilized for each j, with unique hyperparameters. A lognormal prior distribution was assigned to $\kappa_0 \sim \log normal(\mu_{\kappa_0}, \sigma_{\kappa_0}^2)$ and $\nu_0 \sim \log normal(\mu_{\nu_0}, \sigma_{\nu_0}^2)$. $\tau_{jt} \sim Gamma(a_{\tau_{jt}}, b_{\tau_{jt}})$ and $\lambda_j \sim Gamma(a_{\lambda_j}, b_{\lambda_j})$. A possible prior distribution for ε_j is $Beta(\varepsilon_j; a_\varepsilon, b_\varepsilon)$. Hence, $\boldsymbol{\Lambda} = (\{\tau_{jt}\}, \{\alpha_j\}, \{\tau_{\beta_j}\}, \{\tau_{\phi_j}\}, \{\kappa_j\}, \{\nu_j\}, \{\lambda_j\}, \{\varepsilon_j\}, \kappa_0, \nu_0)^T$.

The theoretical derivation of the predictive distribution of the proposed model for new replicates of the response on new spatial locations is shown in Appendix 7.5.1. A Metropolis Hasting algorithm within Gibbs sampling of the Markov Chain Monte Carlo method (MCMC) was developed to draw samples from the joint posterior distribution. In the estimation chain, there is a positive probability that $\boldsymbol{\theta}_{jt}$ would take a value in $\boldsymbol{\Delta}_j^{-t}$ or $\boldsymbol{\Delta}_{j0}^{-t}$ and thus, after several iterations, there would be only a fixed distinct value referred to as clusters values of data j, and the probability of sampling a new value becomes relatively smaller, and the cluster values are rarely updated. This occurs when the chain is balanced on a small set of clusters. To circumvent this problem, it is important to remix the cluster values from their posterior distributions. The posterior distribution is obtained by conditioning on the cluster indicators, and using the likelihood of only the response in the cluster and the base prior distribution. A data augmentation technique was adopted for a smooth update of α_j and α_0 from their posterior distributions. The details of the MCMC algorithm and the link to the `R` code are shown in Appendix 7.5.2.

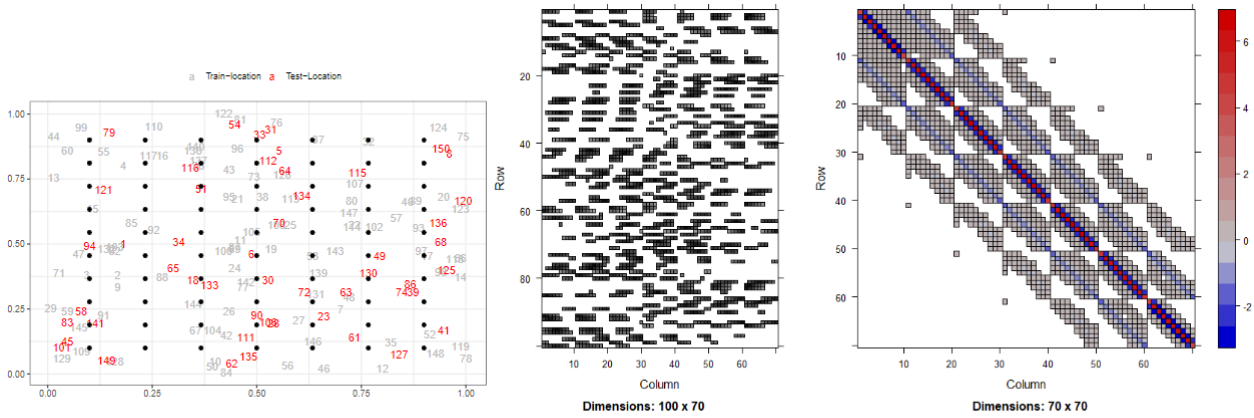
A possible computational concern is the increase in model complexity with the number of spatial knots. A way to ensure control over the complexity is to introduce sparseness by considering only data locations in the nearest neighborhood of the knot positions in constructing the projection matrix $\mathbf{A}_{jt}$ (GOTWAY; YOUNG, 2002). In addition, the Nearest Neighbor Gaussian Process (NNGP) (FINLEY *et al.*, 2019) is adopted for the base distribution using the sparse representation of the covariance matrices as leverage.

6.3 Simulation study

Two simulation experiments were conducted. The first was aimed to extensively assess the adequacy of the proposed model with the independent Dirichlet process prior as a reference, asses the predictive performance, and the impact of spatial knot positions. The second experiment aimed to compare the proposed model performance with independent and exchangeable Dirichlet process priors and assess the sensitivity of the hyperparameters.

6.3.1 Experiment 1

Two non-stationary related spatial data were simulated. Let $\mathscr{G}_1$ and $\mathscr{G}_2$ represent two spatial regions each within the subregion of $\mathscr{G}$. For example, Figure 35a shows the spatial window ([0,1]×[0,1]) adopted and the selected data locations used for training and testing. Let's denote



(a) (b) $\mathbf{A}_j$ (c) $\tilde{\mathbf{\Sigma}}_j^{-1}$

Figure 35 – (a) Simulated data locations and spatial knots adopted to construct the (b) sparse projection matrix $\mathbf{A}_j$ and (c) the NNGP precision matrix $\tilde{\mathbf{\Sigma}}_j^{-1}$ using 10 nearest neighbors.

simulated spatial effects at I different locations for two patients as $I \times 1$ vectors $\boldsymbol{\vartheta}_{1t}$ and $\boldsymbol{\vartheta}_{2t}$ and their shared spatial effect as $\boldsymbol{\vartheta}_{0t}$ in replication t. Here, the ith entry of $\boldsymbol{\vartheta}_{jt}$ can be likened to $h(s_i, \boldsymbol{\theta}_t)$. The spatial effects were drawn from a zero-mean non-stationary Gaussian process with a covariance function

$$\sigma^2(s_i)\sigma^2(s_{i'}) \exp\Big(-\rho_j d(s_i, s_{i'})\Big), \rho_j > 0, i, i' = 1, 2, ..., L, j = 0, 1, 2. \tag{6.11}$$

First, set $\sigma^{2\star}(s_i) = \lambda^{-1}((\text{xc}(s_i) - \text{midxc}(s_i))^2 + (\text{yc}(s_i) - \text{midyc}(s_i))^2)$, where $\text{xc}(s_i)$ is the $x-$coordinate of location s_i and $\text{yc}(s_i)$ is the $y-$coordinate. midxc = (max $\text{xc}(s_i)$+min $\text{xc}(s_i)$)/2 and midyc = (max $\text{yc}(s_i)$+min $\text{yc}(s_i)$)/2. This implies that the $\sigma^{2\star}(s_i)$ approaches zero towards the center of the spatial window and increases away from the center. To avoid degenerating spatial term ($\sigma^2(s_i) = 0$) in Equation 6.11, set $\sigma^2(s_i) = \max(\sigma^{2\star}(s_i), 1)$.

We assumed $\lambda^{-1} = 2.5^2$, $\tau_{1t} = 0.5, \tau_{2t} = 0.03 \, \forall t$, $\rho_1 = 2.5, \rho_2 = 3$, and $\rho_0 = 5$. In addition, linear effects were included to introduce TMS technical noise. That is $\boldsymbol{\beta}_1 = (2, -2, 1)$ and $\boldsymbol{\beta}_2 = (3.5, 1, -2.5, 0.1)$, and the associated covariates were drawn from a standard normal distribution. The responses were drawn as follows: draw $w_{jt} \sim \text{Bernoulli}(e_j)$, and set $\mathbf{Y}_{jt} = \mathbf{X}_{jt}^T \boldsymbol{\beta}_j + \boldsymbol{\vartheta}_{jt} + \boldsymbol{\varepsilon}_{jt}$ if $w_{jt} = 0$ and $\mathbf{Y}_{jt} = \mathbf{X}_{jt}^T \boldsymbol{\beta}_j + \boldsymbol{\vartheta}_{0t} + \boldsymbol{\varepsilon}_{jt}$ if $w_{jt} = 1$, where $\boldsymbol{\varepsilon}_{jt} \sim N(\mathbf{0}, \tau_{jt}\mathbf{I})$, $j = 1, 2$ and $t = 1, 2, ...20$. 150 spatial sites ($i = 1, 2, ...150$) were selected, out of which 100 were used for training and the rest was used for testing.

The hyperparameters assumed are as follow: $\tau_\theta = 1$, $\tau_\beta = 10^{-3}$, and for all j, t,

τ_{jt}	λ_j	κ_j	ν_j	τ_{ϕ_j}	ε	α_j
Gamma	Gamma	Log-Normal	log-Normal	Gamma	Beta	Gamma
$a_{\tau_{jt}} = 2$	$a_{\lambda_j} = 2$	$\mu_{\kappa_j} = 0$	$\mu_{\nu_j} = 0$	$a_\phi = 2$	$a_\varepsilon = 5$	$a_\alpha = 2$
$b_{\tau_{jt}} = 1$	$b_{\lambda_j} = 1$	$\sigma^2_{\kappa_j} = 2.5$	$\sigma^2_{\nu_j} = 2$	$b_\phi = 1$	$b_\varepsilon = 5$	$b_\alpha = 1$.

$L = 70$ equidistant spatial knots were selected, which are represented in Figure 35 as dots, and used to construct $\mathbf{\Sigma}$ whose inverse is a sparse matrix shown in Figure 35c for shared and

specific spatial effects. The element $\mathbf{\Sigma}_{(l,l')}$ was obtained using the Matérn covariance function with l, l' being the locations of the chosen spatial knots. $\psi_j = 4.0$ and the nearest 20 field locations around a given data point were used in the spatial kernel to obtain $\mathbf{A}_j$ shown in Figure 35b, $\forall j$. In the simulation $e_j = e \forall j$ and $e \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$. The proposed model was compared with the model based on independent Dirichlet process priors for each data j. The independent Dirichlet process served as a reference for comparison.

Based on the hyperparameter choices and knot positions, Figure 36a shows the posterior mean and the 95% credible interval of probability ε against the mixing probability used in the simulation. Figure 36b shows the root mean square error (RMSE) of the models for the spatial effect across all spatial locations data were simulated. Figure 36c is the RMSE of the out-of-sample prediction of the simulated response. Each scenario was repeated 100 times and the fainted lines in the figure are the RMSE 95% confidence intervals for the reference model derived from the repeated simulation. The result shows that the proposed model outperformed the reference model, and their performance decreases with an increase in the mixing, though more severe for the reference model. A similar conclusion was obtained for the posterior predictive of the response.

To have a deeper view of the proposed model performance, a simulation for fixed $e = 0.4$ was conducted. Figure 37 shows the posterior predictive density for the specific and shared spatial effects. The histograms are the true distribution of these effects across the simulated locations. For the shared spatial effect (right panel), the predictive density represented by black curves are the shared predictive densities estimated from data 1 and the blue curves are those from data 2. The posterior mean and 95% credible intervals of the spatial range and smoothness for spatial effects 1, 2 and shared are $\kappa_1 = 3.248\,(0.042,\, 11.731)$, $\nu_1 = 0.648\,(0.029,\, 2.081)$, $\kappa_2 = 1.188\,(0.079,\, 3.154)$, $\nu_2 = (0.106,\, 5.244)$, and $\kappa_0 = 9.528\,(4.355,\, 14.643)$, $\nu_0 = 0.689\,(0.031, 3.084)$, given rise to the auto-correlation curves in Figure 38a and covariance in Figure 38b. The correlation function for the shared spatial effect exhibits a short range, which agrees with ρ_0 since it was relatively larger than the specific effects. On the contrary, auto-correlation of spatial effect 2 had the longest range despite ρ_2 being relatively moderate. However, the estimated covariance closely represents the actual covariance pattern in the simulated data. Figure 39 shows the in-sample and out-of-sample posterior predictions of the responses. The histograms are the true simulated values and the density overlaid are the predictive densities. To aid visualization, five spatial sites from training and testing sites were randomly chosen. The grey histograms correspond to data 1, and the white correspond to data 2. The black posterior densities correspond to the in-sample predictive densities, and the red corresponds to the out-of-sample predictive densities. The result shows that the model performed well in describing the true data distribution. In the same simulation scenario, but repeated 100 times, the average posterior mean, standard deviation, mode, bias, and coverage probability (CP) of the corresponding linear effects and precision parameters are shown in Table 8.

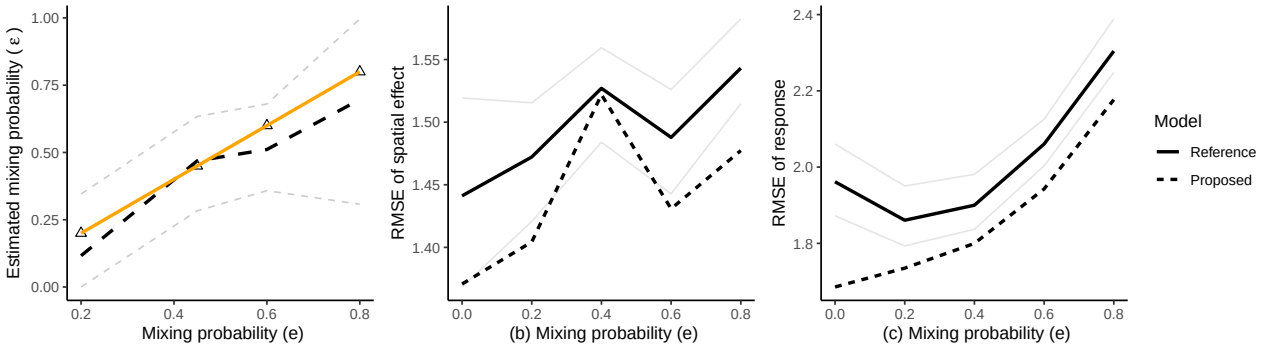


Figure 36 – (a) The posterior means and 95% credible interval of the probability ε from the proposed model against the mixing probability e used in the simulation. The triangles indicate the true value of ε. (b) The root mean square error of the predicted spatial effect across training locations using the proposed model and the independent Dirichlet processes (Reference) for each j, and (c) is the out-of-sample RMSE of the predicted response at testing locations for the proposed and reference model. The x-axis is the mixing probability (e) used for data simulation. The fainted lines are the 95% credible interval of the RSME obtained from simulating 100 scenarios. The triangular points indicate the expected mixing probability.

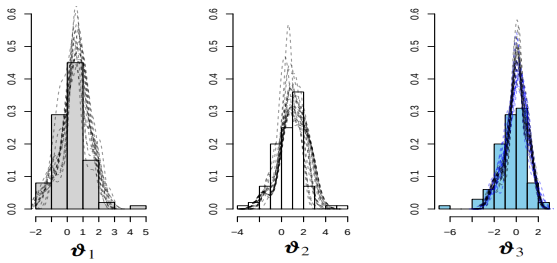
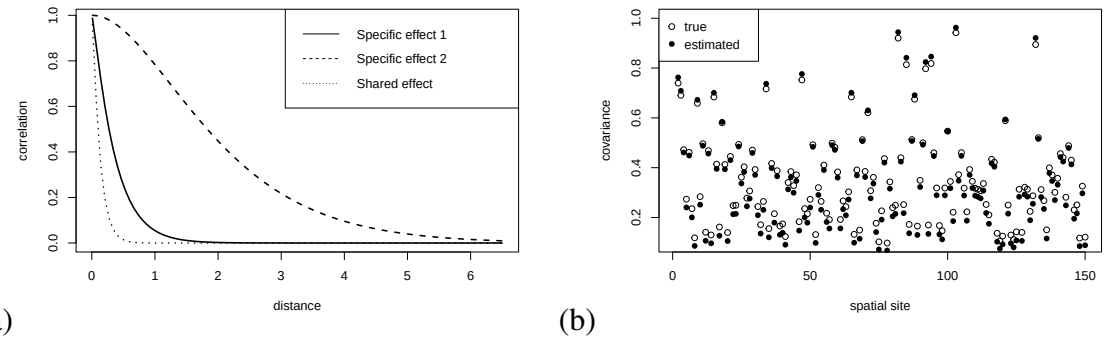


Figure 37 – The posterior predictive density of the spatial effects overlaid on the histogram of the simulated spatial effects. For the shared spatial effect, the black and blue densities are respectively of data 1 and 2.



(a) (b)

Figure 38 – (a) The estimated spatial correlation. (b) Estimate of the covariance between spatial site 1 and every other simulated spatial site for data 1.

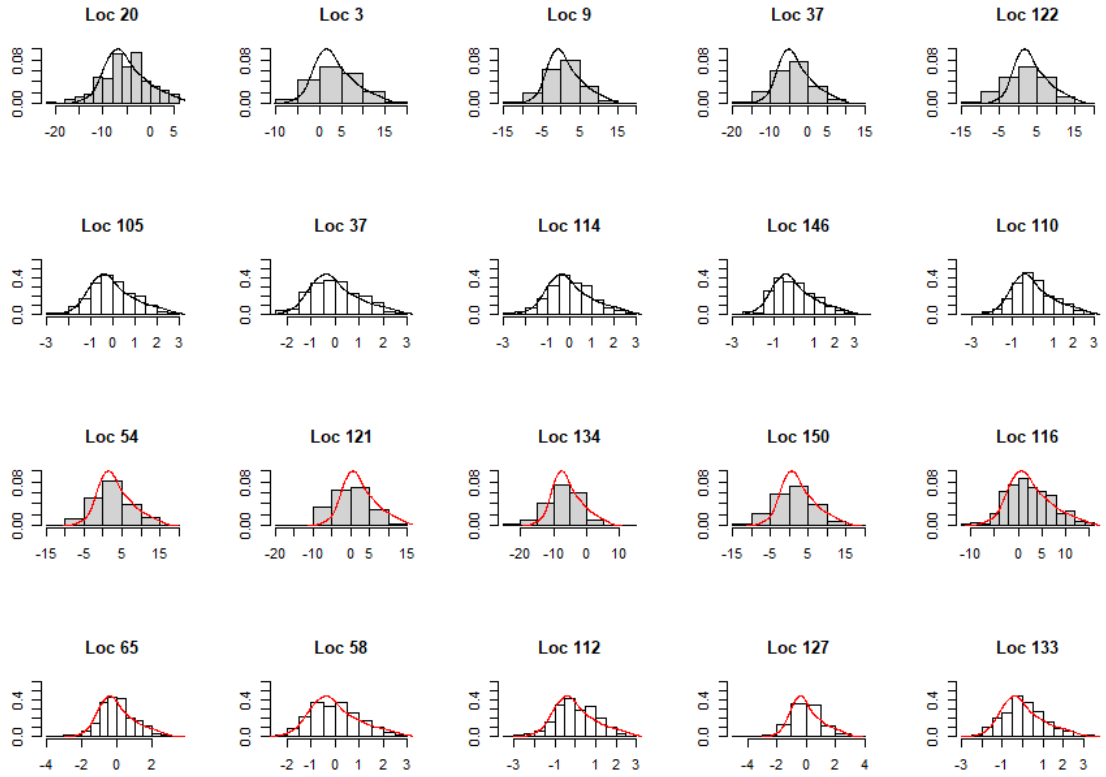


Figure 39 – Histogram of simulated response and posterior predictive densities: the histograms represent the true distribution of the response and the overlaid densities represent the corresponding posterior predictive density. The grey and white histograms correspond to data 1 and data 2 respectively, and the black and red densities correspond to in-sample and out-of-sample predictive densities respectively.

The impact of spatial knot position on the model's predictive performance was examined by considering multiple simulation scenarios for different choices of the number of spatial knots and positions. The types of knot positions covered were (1) regular spatial knots and (2) random spatial knot positions. In (1), the spatial knots are placed at equidistant points from one another; whereas, in (2) they are randomly placed over the spatial region. The overall impact on the posterior prediction performance in simulated data 1 and 2 was measured through the RMSE. As expected, the result shows that the RMSE decreases as the number of chosen spatial knot locations increases. This confirms that the model performance increase with increases in the number of spatial knots; however, the model complexity increases. Thus, there is a trade-off between prediction performance and complexity. In the simulation scenarios, the model based on randomly selected spatial knot locations consistently outperformed the model based on regularly selected spatial knots.

Table 8 – Average posterior mean, standard deviation, mode, bias, and coverage probability (CP) over hundred repeated simulations.

Parameters	True value	Mean (sd)	Mode	Bias	CP
β_{11}	2.0	2.008 (0.044)	2.010	0.008	0.97
β_{12}	-2.0	-2.006 (0.044)	-1.995	-0.006	0.96
β_{13}	1.0	0.998 (0.044)	1.007	-0.002	0.93
β_{21}	3.5	3.500 (0.086)	3.500	-0.000	0.92
β_{22}	1.0	1.006 (0.084)	1.014	0.006	0.91
β_{23}	-2.5	-2.500 (0.082)	-2.488	0.000	0.93
β_{24}	0.1	0.103 (0.084)	0.106	0.003	0.90
ε	0.4	0.457 (0.071)	0.456	0.157	0.90
τ_1	0.5	0.705 (0.198)	0.704	0.205	0.96
τ_2	0.03	0.042 (0.009)	0.042	-0.158	0.95

6.3.2 Experiment 2

Three related data were simulated in the following scheme. Let $\mathscr{G}_1, \mathscr{G}_2$, and $\mathscr{G}_3$ represent the three spatial regions each within the subregion of $\mathscr{G}$, and $\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}_1, \boldsymbol{\vartheta}_2, \boldsymbol{\vartheta}_3$ be the spatial effect obtained from a stationary Gaussian random field centered at zero with Matérn covariance function with partial sill $\sigma_0^2 = 0.1$ and range $\kappa_0 = 1.0$ for shared effect ($\boldsymbol{\vartheta}_0$), partial sill $\sigma_1^2 = 0.5$ and range $\kappa_1 = 5.0$ for specific effect 1 ($\boldsymbol{\vartheta}_1$), partial sill $\sigma_2^2 = 0.2$ and range $\kappa_2 = 0.25$ for spatial effect 2 ($\boldsymbol{\vartheta}_2$), and partial sill $\sigma_3^2 = 0.25$ and range $\kappa_3 = 0.1$ for spatial effect 3 ($\boldsymbol{\vartheta}_3$), with a constant smoothness $\nu = 0.5$. The arbitrarily chosen linear effects were $\boldsymbol{\beta}_1 = (2.0, 1.0, -1.4)$, $\boldsymbol{\beta}_2 = (0.2, -2.5, -3.0, 4.0)$, $\boldsymbol{\beta}_3 = (-2.0, 0.5)$. The corresponding design matrices for the linear effect were drawn from the standard normal distribution. A random walk model was adopted for the nonlinear effect. That is, $z_{jr} \sim N(z_{j(r-1)}, \sigma_\phi^2), z_{j1} \sim N(0, \sigma_\phi^2)$, $j = 1,2,3$, where $\sigma_\phi = 0.5$.

In this example, $w_{jt} \sim \text{Bernouli}(e_j)$, $e_1 = 0.2, e_2 = 0.4$, and $e_3 = 0.5$, $\tau_{jt} \sim \text{gamma}(0.5, 2), \forall\, j,t$. If $w_{jt} = 0$, $\mathbf{Y}_{jt} = \mathbf{X}_j^T \boldsymbol{\beta}_j + \mathbf{z}_j + \boldsymbol{\vartheta}_j + \mathbf{e}_{jt}$, otherwise $\mathbf{Y}_{jt} = \mathbf{X}_j^T \boldsymbol{\beta}_j + \mathbf{z}_j + \boldsymbol{\vartheta}_0 + \mathbf{e}_{jt}$. $\mathbf{e}_{jt} \sim N(\mathbf{0}, \tau_{jt}^{-1}\mathbf{I}), t = 1, 2, ..., 50$. The number of data locations was fixed at $I_j = 100$ for all j. Similar hyperparameters in experiment 1 were used, except that $a_{\varepsilon_j} = 0.3$ and $b_{\varepsilon_j} = 0.3$.

To relatively evaluate the proposed model's adequacy, the performances of the "no sharing" and "full sharing" reference models were obtained alongside the proposed model. The "no sharing" is the model which assigns an independent Dirichlet process prior to the spatial fields corresponding to each j data, and the "full sharing" assigns a single (exchangeable) Dirichlet process prior to the spatial effect corresponding to all j data.

Table 9 shows the posterior mean, 95% credible interval, and the bias of the linear effect estimates. As expected in Table 9, the linear effect estimates are only slightly different from one another. This is because these effects do not directly involve in the sharing component of the model. Thus, the sharing associated with the spatial effect has little impact on the linear effects estimates. The nonlinear effect estimate from the proposed model is shown in Figure 40. Similar to the linear effect, the estimate of the nonlinear effect from the "no sharing" and "full sharing"

Table 9 – Posterior estimates of the fixed effects.

Parameter		Proposed				No sharing (Independent)				Full sharing (Exchangeable)			
	True	mean	2.5%	97.5%	bias	mean	2.5%	97.5%	bias	mean	2.5%	97.5%	bias
β_{11}	2.00	1.976	1.953	2.001	-0.024	2.010	1.977	2.046	0.010	2.008	1.976	2.038	0.008
β_{12}	1.00	0.923	0.900	1.095	-0.077	0.902	0.879	0.926	-0.098	0.909	0.885	0.941	-0.091
β_{13}	-1.40	-1.406	-1.427	-1.384	-0.006	-1.396	-1.430	-1.365	0.004	-1.403	-1.438	-1.377	-0.003
β_{21}	0.20	0.209	0.1983	0.220	0.009	0.168	0.148	0.187	-0.032	0.160	0.139	0.174	-0.040
β_{22}	-2.50	-2.406	-2.521	-2.393	0.092	-2.514	-2.531	-2.500	-0.014	-2.522	-2.537	-2.506	-0.022
β_{23}	-3.00	-3.001	-3.011	-2.990	-0.001	-2.937	-2.950	-2.923	0.063	-2.950	-2.965	-2.937	0.049
β_{24}	4.00	3.998	3.986	4.008	-0.002	4.051	4.032	4.067	0.051	4.0234	4.008	4.036	0.023
β_{31}	-2.00	-1.929	-1.943	-2.091	0.071	-2.009	-2.028	-1.995	-0.009	-1.997	-2.012	-1.983	0.003
β_{32}	0.500	0.520	0.5062	0.535	0.020	0.476	0.459	0.491	-0.024	0.483	0.467	0.497	-0.017

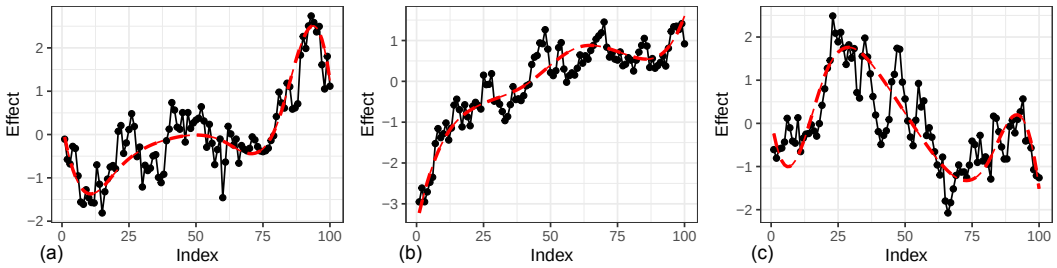


Figure 40 – Posterior estimates of the nonlinear effects from the proposed model for (a) $j = 1$, (b) $j = 2$, and (c) $j = 3$. The black lines are the true effects and the red lines are the estimated effects.

also had similar patterns.

Figure 41 shows the spatial effect estimates for the three competing models. Panel (a) is the true spatial effect. The estimate of the spatial effect through the proposed model is shown on panels (b), while the estimate from the "no sharing" and "full sharing" models are respectively shown on panels (c) and (d). The proposed model could capture the pattern in the true spatial effect. It shows more accurate estimates of the regions with very high or low effects. The posterior mean and 95% credible interval for $\varepsilon_j, j = 1,2,3$ are $0.167\,(0.000, 0.550)$, $0.320\,(0.024, 0.730)$, and $0.524\,(0.426, 0.827)$ respectively, and $\alpha_j s$ were $7.675\,(4.510, 11.554)$ for the shared component, $1.005\,(0.035, 3.642)$, $0.966\,(0.029, 3.790)$, and $4.910\,(2.449, 8.764)$ for $j = 1,2,3$ respectively. The density plots of the 50 replicates of the predicted responses were overlaid on the simulated responses and are shown in Figure 42. The estimated density showed a considerably good fit to the true density. The performances of the model for other mixing probabilities $e \in \{0.6, 0.9\}$ were again investigated using 30 replicates ($t = 1,2,...,30$) at each scenario. The result shows a slight decline in the recovery accuracy of the spatial effects as e increases, which conforms with experiment 1. That is, the spatial effects appear more similar to the estimates obtained with "full sharing" ($\varepsilon = 1$) as e increases. However, the fixed and the nonlinear effects were unaffected.

The sensitivity of the hyperparameters involved in the prior distributions was investigated. The variance assigned to the prior distribution for the linear effects varied from $\sigma^2_\beta = 10, 100,$ and 10000 for all the related spatial data, and the results showed no significant impact on the estimates. Similarly, for the nonlinear effects and the hyperparameter of the likelihood (τ_{jt} and

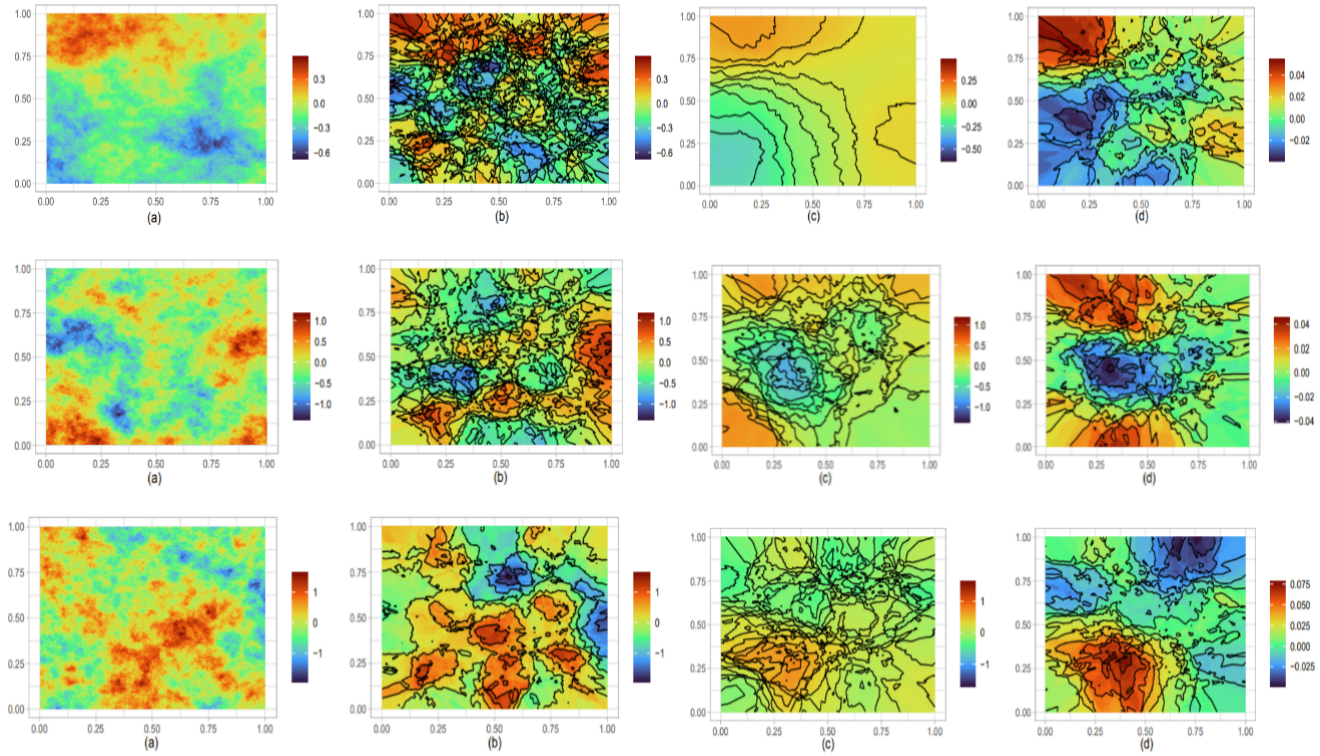


Figure 41 – (a) The true simulated spatial effect. The corresponding projected spatial effect using the (b) proposed, (c) no sharing, and (d) full sharing reference models. Rows 1, 2, and 3 correspond to data 1, 2, and 3.

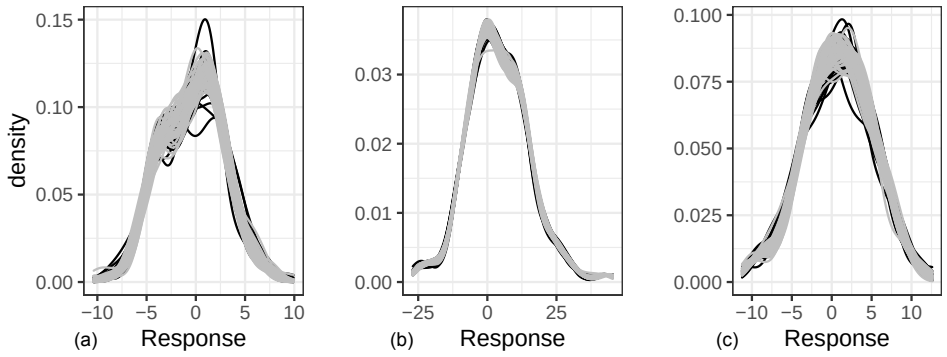


Figure 42 – True response density (black) and the proposed model predictive density (grey) for (a) $j = 1$, (b) $j = 2$, and (c) $j = 3$.

τ_ϕ), the shape and the rate parameter of the gamma prior distribution assigned were varied in the set $\{2,10,20\}$ and $\{1,10\}$ respectively. While no significant impact on the result was observed for τ_{jt}, the combination of the shape and scale that lead to a higher prior mean of τ_ϕ increases the smoothness of the nonlinear effect. One essential parameter of the spatial effect is the ε. The hyperparameters of the beta distribution were varied such that the prior variance is large. Specifically, $a_{\varepsilon_j} = g_a \times T_j$ and $b_{\varepsilon_j} = g_b \times T_j$, $g_a, g_b \in (0,1)$ were varied in the set $\{1/300, 1/30, 1/3\}$. Choosing a relatively high proportion of the total replicate (T_j) for the hyperparameters a_{ε_j} and b_{ε_j} pulls the prior distribution around 0.5 and whereas smaller proportions pull the prior mean away for 0.5 with higher variance. It was observed in both situations that the changes affect the posterior distribution of the parameter. It is recommended that a careful prior distribution choice be considered for g_a and g_b to avoid bias. A possible choice is the simplex and truncated normal distributions. Moreover, the concentration parameter α was also sensitive to the prior hyperparameters. However, the estimated spatial effects were not significantly affected. Based on these findings, the choice of hyperparameters should be carefully considered and model performance must be evaluated using metrical criteria in a real data application.

6.4 Data analysis

6.4.1 Data

The developed model was applied to analyze TMS data from three subjects (1R1, 2,3), observed independently by the same TMS practitioner. Hence, $j = 1,2,3$. Here, the subjects' datasets were tagged "Data 1", "Data 2", and "Data 3". The analysis aimed to unravel the spatial pattern of the subjects' response to TMS at the motor cortex. Such findings may help quantify the cortical and spinal excitability over the motor cortex, which is essential for improving patients' treatment experience. 10 TMS pulses were delivered at each spatial location, and therefore, $t = 1,2,...,10$. The spatial covariate considered in this analysis is the `Cartesian coordinate`. The `peak-to- peak MEP` (μV) at those locations serves as the target variable. The linear covariate, `replicate index`, indicates what stimulation replicate gave rise to the corresponding MEP response, and the nonlinear covariate, `angular error`, is the angular positioning error of the TMS coil, measured in degree. The response variable is defined as $\mathbf{y} = \log$ (`peak-to-peak MEP`). The box plot of the response according to some selected rectangular grid groupings (Associate Targets) is shown in Figure 43. It is evident that there exist variations in the patient's responses to TMS pulses across locations and participating subjects. The developed method was adopted to allow spatial information sharing between the subjects while estimating the subject-specific spatial pattern.

Three different structural functional forms were considered. In structural form 1, the functional form $\mu_{jit} = x_{jit}\beta_j + g(z_{jit}, \boldsymbol{\phi}_j) + h(s_{ji}, \boldsymbol{\theta}_{jt})$ was adopted, where x_{jit} is $1 \forall j,i,t$ and

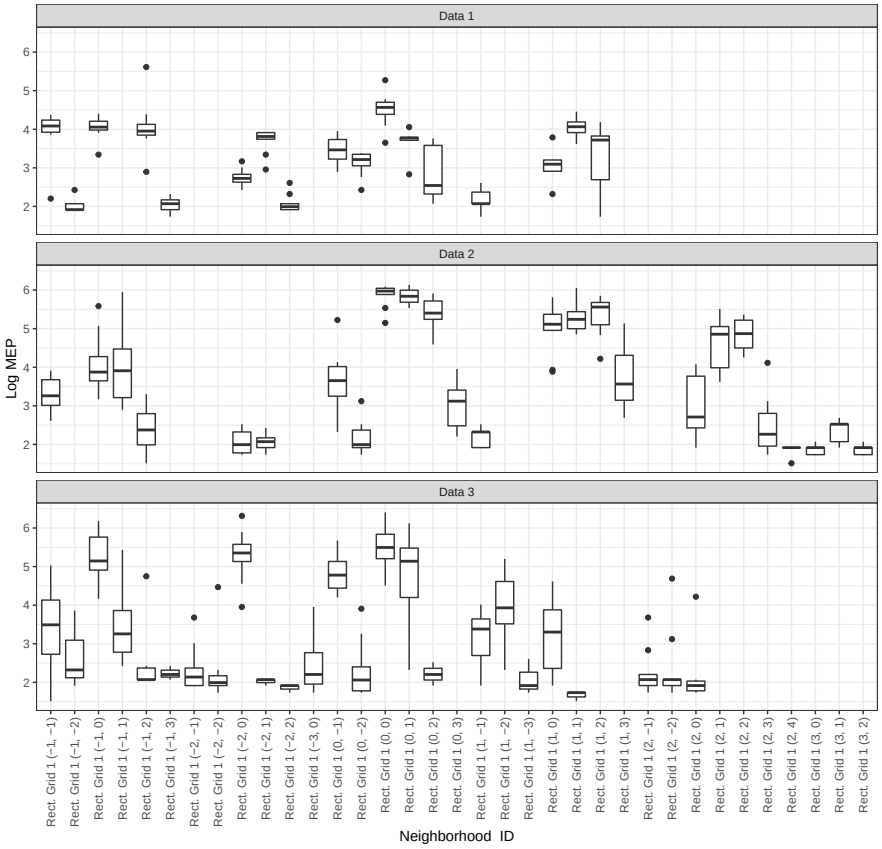


Figure 43 – Descriptive plot of the log peak-to-peak MEP according to rectangular grid cells for the purpose of better visualization. The neighborhood with an empty box plot indicates that no data was captured in that rectangular grid. The grid is labeled according to the coordinate of the cell centroid.

the coefficient β_j is a scalar. This model assumed that there is a common intercept over all the spatial locations in $\mathscr{G}_j$ and across replications, and it is relatively simple and parsimonious. Here z_{jit} represents the `angular error` modeled using cubic spline, and $\mathbf{s}_{ji}$ is the spatial location i (`Cartesian coordinate`) of the tth replication in the jth data. The structural form 2 follows as $\mu_{jit} = x_{jit}\beta_{jt} + g(z_{jit}, \boldsymbol{\phi}_j) + h(s_{ji}, \boldsymbol{\theta}_{jt})$. Here x_{jit} and β_{jt} have similar dimensions as in form 1, however, it is assumed that the intercept varies across replication, and each x_{jit} is constructed from the `replicate index` indicator. Lastly, the structural form 3, $\mu_{jit} = x_{jit}\beta_{jit} + g(z_{jit}, \boldsymbol{\phi}_j) + h(\mathbf{s}_{ji}, \boldsymbol{\theta}_{jt})$. Here x_{jit} and β_{jit} have a similar dimension as form 1, however, the model assumes that the intercept varies across location and replication. Again, x_{jit} is constructed from the `replication index` and the data coordinate (`Cartesian coordinate`). Other components were retained for all the models. Each β was assigned a Gaussian prior distribution. $\tau_\beta^{-1} = 10^3$, $a_{\phi_j} = 2, b_{\phi_j} = 1$, $a_{\lambda_j} = 50$, and $b_{\lambda_j} = 10$, $a_{\tau_{jt}} = 5$, $b_{\tau_{jt}} = 5, \forall j,t$, $a_\alpha = 2$, and $b_\alpha = 1$.

Each dataset was allowed to determine its sharing rate. That is, $a_{\varepsilon_j} = b_{\varepsilon_j} = (1/10) * T_j$ and $T_j = 10, \forall j$. In the data, $I_1 = 171$, $I_2 = 225$, $I_3 = 201$. The covariance matrices of the base prior distribution of the DP process were constructed using randomly chosen spatial knots.

The Conditional Predictive Ordinate (CPO) was used to investigate the performance of

the models. The CPO for observation ji was approximated as

$$CPO_{ji} \approx \left[\frac{1}{M} \sum_{k=1}^{M} \frac{1}{f(\mathbf{y}_{ji} \mid D_{-ji}, \mathscr{X}_{(k)})} \right]^{-1}, \tag{6.12}$$

where $\mathscr{X}_{(k)}, k = 1,2,...,M$ represents all the model parameters drawn from their respective marginal posterior distributions; M is the total number of draws; D_{-ji} is the data excluding jith observation and $f()$ is the data model. Higher values of CPO indicate a better fit. For a chosen MEP, the uncertainty of the joint posterior distribution about y was computed. That is the exceedance probability given a threshold y was computed as

$$P(\mathbf{Y}_j > y\mathbf{1} \mid D) = \mathbf{E}(\mathbb{I}_{(\mathbf{Y}_j > y\mathbf{1})} \mid D) = \int \mathbb{I}_{(\mathbf{Y}_j > y\mathbf{1})} P(\mathbf{Y}_j \mid D) d\mathbf{Y}_j, \tag{6.13}$$

where $\mathbf{1}$ is a column vector of ones and $\mathbb{I}_{(\mathbf{Y}_j > y\mathbf{1})}$ is an indicator random vector. Equation 6.13 was approximated using Monte Carlo samples from the posterior predictive distribution $P(\mathbf{Y}_j \mid D)$. An MEP of $400\mu V$ was chosen, which implies that $y = \log(400)$. The chosen threshold was informed based on the resting motor thresholds of the participants.

Two MCMC chains were used to check for the convergence of all the parameters through the behavior of the joint log-likelihood. The Gelman and Rubin (1992) diagnostic test was used to verify the convergence. The trace plots are shown in the Appendix (Figure 51).

6.4.2 Result

Figure 44 shows the *CPO* for each observation. The higher the score, the better the model. From the figure, the proposed model based on the structural form 1 and 3 performed nearly equally, however, they underperformed relatively to form 2. To numerically quantify the performance of the model and in comparison with exchangeable priors, Table 10 shows the Watanabe-Akaike information criterion (WAIC) and the *mlCPO*. The *mlCPO* was defined as the negative mean of log CPO, indicating that the lower the *mlCPO* and WAIC, the better the model. Based on these measures, the proposed model outperformed the exchangeable prior, indicating that there is a significant mixture of latent processes in the acquired data. Figure 45 shows the quantile residual analysis. From the Figure, the model fits the data adequately, though better for *Data* 1 and *Data* 2.

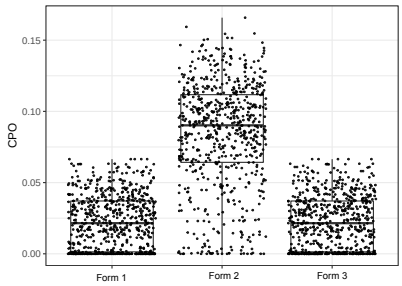


Figure 44 – Model comparison using *CPO*.

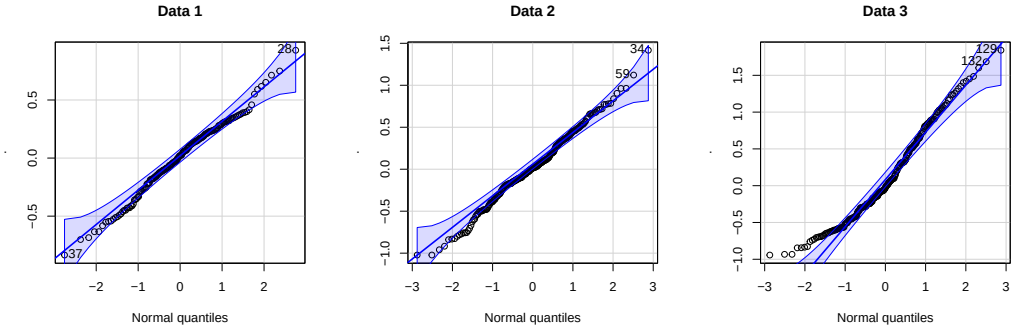


Figure 45 – Quantile residual from the data analysis.

Table 10 – Adequacy measure.

		WAIC	*mlCPO*
Proposed	Structural form 1	10194.127	53.919
	Structural form 2	-373.5657	1.362
	Structural form 3	3040.114	9.973
Exchangeable	Structural form 2	1960.287	3.324

The posterior mean and 95% credible intervals of the sharing probability for each Data are $\varepsilon_1 = 0.918\,[0.715, 0.980]$, $\varepsilon_2 = 0.819\,[0.723, 0.970]$ and $\varepsilon_3 = 0.750\,[0.722, 0.890]$, and the scale parameters of the DP prior are $\alpha_1 = 3.887\,[2.373, 5.793]$, $\alpha_2 = 3.860\,[2.302, 5.782]$, $\alpha_3 = 3.87\,[2.353, 5.751]$, $\alpha_4 = 3.583\,[2.268, 5.212]$. The posterior mean and 95% credible interval for τ_1, τ_2 and τ_3 are $0.636\,[0.083, 1.718]$, $0.662\,[0.098, 1.773]$, and $0.335\,[0.053, 1.003]$ respectively. Figure 46a & b show the posterior mean and the 95% credible interval for the linear and nonlinear effects. Figure 46a shows that the replication effects for *Data* 1 and *Data* 2 are significant since the credible intervals do not contain zero. However, this is not true for *Data* 3. In *Data* 1 and *Data* 2, the replications effect decreases from replication 1 to 2 but increases from replication 2 to 4. While the fixed effect for *Data* 2 reaches the minimum at replication 6 and then gradually increases afterward until replication 10, the replication effect for *Data* 1 has a sinusoidal pattern until it reaches the minimum at replication 7 and thereafter gradually increases until replication 10. The replication effects for *Data* 3 had a sinusoidal pattern and imitates the patterns in *Data* 1 and 2, and more closely from replicate 6 to 10. Figure 46b shows the nonlinear pattern of angular positioning error. The result shows that angular positioning does not significantly affects the MEP of DATA 2 and 3 since the pattern behaved sinusoidally around the null effect; however, it significantly alters the MEP of DATA 1, which leads to higher positive effects as angular error increases. Moreover, findings also showed that stimulation depth is not significant (result not shown).

Figure 47 shows the mean of the posterior predictive distribution of the response over the whole spatial domain for each data. The locations with high values signify the candidate locations where optimal cortical and spinal excitability can be obtained for the corresponding

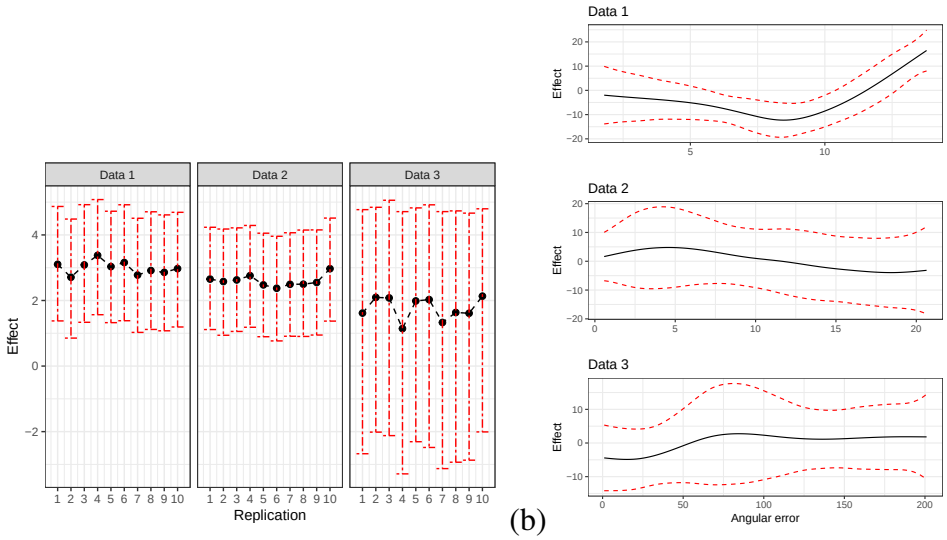


Figure 46 – Posterior mean and 95% credible interval for the (a) replicate effect, and (b) TMS angular positioning error.

subjects. The locations with low values indicate the non-candidate locations and should not be considered optimal in a TMS motor mapping for these subjects. The result shows that the primary motor cortex, M1, is a candidate location that could be considered for optimal excitability of the right first dorsal interosseous muscle. For *Data* 1, it is located at $(52.102, -14.105)$, and $(53.224, -8.217)$ and $(-47.479, -16.880)$ for *Data* 2 and 3 respectively. The locations with optimal excitability are concentrated in *Data* 2 and 3, whereas they are dispersed in *Data* 1.

Figure 48 shows the uncertainty about an MEP exceeding $400\mu V$. The higher the exceedance probability, the higher the possibility a subject's motor evoked potential exceeds $400\mu V$. The uncertainty map produced helps to compare the stimulation effectiveness and response rate to TMS pulses among the subjects. The result revealed that the subject associated with Data 2 has the highest exceedance probability compared with other subjects. Moreover, the subject of *Data* 1, which had the lowest probability, has its peak probability at locations away from M1, unlike other subjects. Figure 49 shows the shared spatial effect with reference to the primary motor cortex. The primary motor cortex is located at the center (0,0) of the plot, which coincides with all the subjects. The x and y axes are the distance away from this location. The result shows that at the primary motor cortex, there is a relatively lesser shared effect compared with its surroundings, especially the first and fourth quadrants. This indicates that, at this location, responses to TMS stimulation tend to be more particular to each subject and less likely to respond at a similar frequency across the subjects.

6.4.3 Discussion of result

The results obtained unveil the spatial patterns of cortical and spinal excitability among the three subjects. The finding supports the theory of the delivery of repetitive TMS (rTMS) pulses for brain treatment. This is evident in the replication effects, where the first replicate did

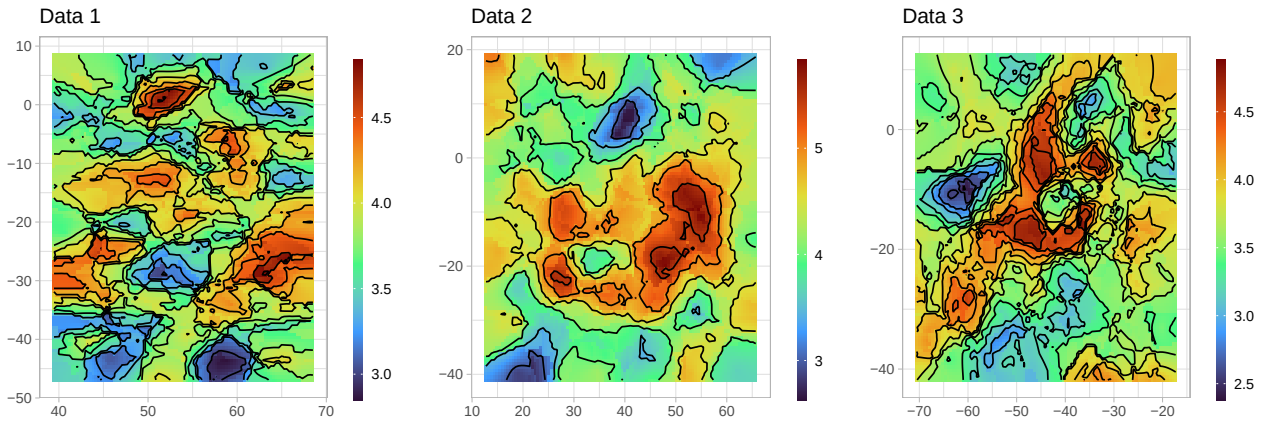


Figure 47 – The posterior mean of the predictive distribution of the response.

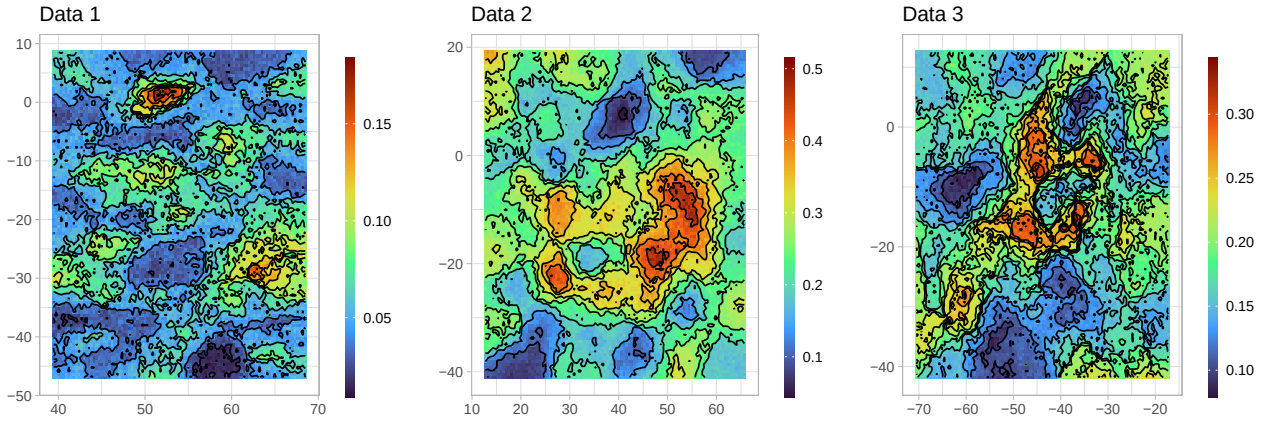


Figure 48 – Exceedance probability over MEP threshold of $400\mu V$.

not produce the highest average excitability response and revealed an ascension tendency at the end of replication 10. This may be associated with why rTMS causes a long-lasting after-effect and is thought to induce plasticity (KLOMJAI; KATZ; LACKMY-VALLÉE, 2015). The non-significance of the effect of stimulation depth is expected as the TMS pulses are electromagnetic and pass through several layers of the brain unimpeded, which has made TMS a popular non-invasive neurophysiological technique (NAJIB *et al.*, 2011). The finding shows that the location of the primary motor cortex of the hand, M1, was consistently obtained as a candidate location for achieving optimal functional excitability, however, vary significantly with relatively low shared patterns. This implies that M1 could be used as a location to attain significantly high functionality, especially in cases where cost and exploration time are to be minimized. The visibility of the M1 location to practitioners makes this efficiency gain even more realistic. The shared spatial effect pattern reveals the existence of consistent patterns around the M1. That is, locations around the primary motor cortex tend to better share information and behave more similarly across subjects. The spatial pattern produced in this work reveals new insight that could be used by TMS practitioners to informatively guide brain stimulation for a better patient

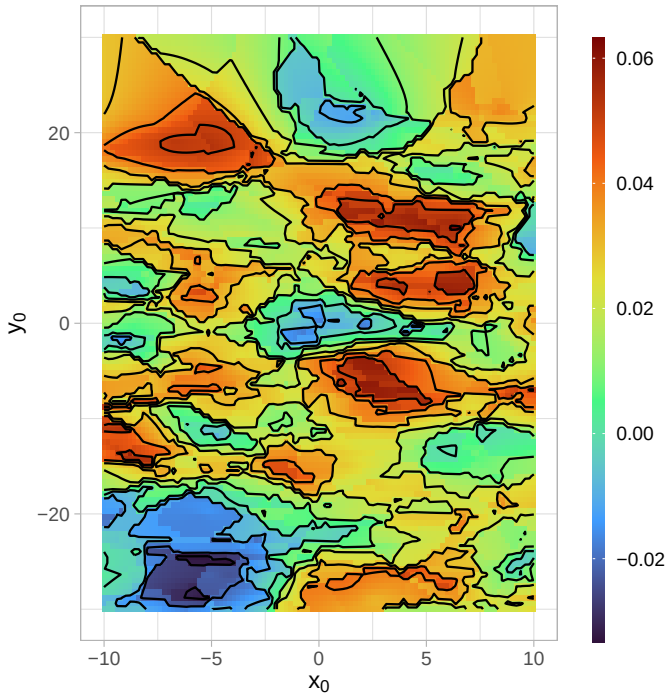


Figure 49 – The projected posterior mean of the shared spatial effect. In the figure, x_0 denotes how far away from M1 in the x-axis and y_0 denotes how far away from M1 in the y-axis. That is, location $(0,0)$ is the primary motor cortex common to all subjects.

treatment experience.

6.5 Conclusion

This study presents a novel technique for combining and modeling geostatistical spatial datasets using a Dirichlet process within a single Bayesian framework. This approach offers several advantages over traditional models by relaxing modeling assumptions and allowing for more flexibility. To account for model covariates' impact, a structural additive predictor with linear, nonlinear, and spatial terms was implemented. The developed model probabilistically switches data replicates between shared and specific components of the model, allowing information sharing among sub-spatial models. A Markov chain Monte Carlo algorithm was developed for model estimation. The model was applied to quantify the spatial pattern of patients' brain activation in a motor mapping of the right first dorsal interosseous muscle. The results revealed that the primary motor cortex consistently showed optimal neuron activation, making it a promising candidate area for motor cortex stimulation. These findings help practitioners tailor their approach accordingly to improve the efficacy of treatment, while also reducing the risk of adverse effects. The generated plots provided valuable insights into the response capacity of different areas of the motor cortex, which can be used to improve patients' medical experiences.

CHAPTER

7

CONCLUSION

7.1 Remarks

This project developed Bayesian spatial statistical models for analyzing Transcranial magnetic stimulation (TMS) geostatistical datasets, to unveil new insight into responses of the right first dorsal interosseous muscle to TMS stimulation pulses for the improvement of patients' treatment experiences. This is crucial given that a TMS therapy session heavily relies on accurately determining the true spatial response pattern of the patient undergoing treatment. This factor directly impacts the effectiveness of the therapy. Therefore, it is imperative to prioritize the precise identification of the patient's spatial response pattern to ensure optimal outcomes.

The first contribution of this project was a systematic review and meta-analysis of the existing Bayesian spatial models that could be considered for analyzing TMS datasets. Several state-of-the-art Bayesian spatial models and methodological gaps were identified. Based on some selected existing models in the literature, the second contribution of this project was the development of a user interface for performing Bayesian spatial modeling for analyzing TMS datasets with state-of-the-art models. The interface was documented as an R package and was used to analyze the TMS data of this project. Findings from the analysis identified, with uncertainties, the hotspot regions of patients' motor cortexes that are responsible for the movement of the right first dorsal interosseous muscle.

The third contribution proposed novel spatial statistical models for integrating geostatistical datasets in the form of prior elicitation in a Bayesian analysis. It creates a medium through which inference can be improved by elicitating coherent information from historical point-referenced data. The proposed models were validated using simulation studies. The study shows that naively integrating geostatistical datasets without ensuring the consistency of the data is detrimental to the desired inferences; however, the proposed model showed evidence of ensuring appropriate integration necessary for obtaining better results. The models were

used to analyze the TMS dataset for this project and the results showed significantly superior performance over comparable methods in the literature. The models unveiled new insights into the spatial patterns and replication significance effects of patient responses to TMS pulses.

The final contribution of this project is the proposition of a nonparametric spatial model with a nonstationary spatial process for the joint analysis of geostatistical datasets. The method used a Dependent Dirichlet process with some latent variables to share information across sub-spatial models. The proposed model is capable of harnessing and sharing coherent information across sub-spatial models, necessary to improve inference and simultaneously analyze multiple datasets. Two simulation studies were used to validate the model performance in comparison with a single analysis of individual data. The model showed superior performance over equivalent nonparametric models. The proposed model was used to simultaneously analyze multiple subjects' TMS data by sharing spatial information about the response of the motor cortex across sub-spatial models. The result unveils new insights that could be highly relevant for improving the stimulation skill of TMS practitioners and thereby improving patients' treatment experiences.

Though the statistical models developed in this work were motivated by the TMS detests, they can be adapted for any similar geostatistical or point-referenced spatial dataset to unveil spatial patterns and make predictions on unsampled locations. It is important to note that while the TMS dataset does not suffer from spatial confounding issues, Appendix 7.6 provides valuable steps to overcome spatial confounding problems in similar geostatistical datasets.

7.2 Limitations and future consideration

The first limitation associated with this project is the few number of participants' data available. This is a result of the cost related to TMS therapy. Hence, findings on the project may require further validation with additional numbers of participants. Recall that the peak-to-peak measure of signal variation was considered in this work, other stable and robust measures of variations such as those proposed in Rousseeuw and Croux (1993) can be considered. Moreover, the raw signal data can be equally considered, so long it is interpretable.

Another limitation of the work is the computational cost associated with analyzing the data using the proposed nonparametric model since it was required to loop through each observation during estimation. For large datasets, this limitation could be a significant drawback. Hence future work can develop fast algorithms to loop through the datasets in a systematic and block-type manner. In addition, future work can consider the Variational Bayes estimation algorithms for the proposed models for analyzing a high volume of data.

A future version of the developed `TMSBrainApp` could implement more response models and more robust signal variation estimators which could improve on the generality of the interface. In addition, the future version can consider the implementation of shape files for analyzing and plotting disease patterns in epidemiological datasets. Lastly, Non-Gaussian random field models

could be implemented.

Future work can consider a mixture of hierarchical Dirichlet processes, which is able to detect shared hierarchical clusters across multiple data sources with multiple sources of variation. For example, this could be adopted in Single-cell RNA sequencing data to reveal shared cell interactions.

BIBLIOGRAPHY

ADEGBOYE, O. A.; ADEKUNLE, A. I.; PAK, A.; GAYAWAN, E.; LEUNG, D. H.; ROJAS, D. P.; ELFAKI, F.; MCBRYDE, E. S.; EISEN, D. P. Change in outbreak epicentre and its impact on the importation risks of covid-19 progression: A modelling study. **Travel Medicine and Infectious Disease**, Elsevier, v. 40, p. 101988, 2021. Citation on page 46.

AKSEER, N.; BHATTI, Z.; MASHAL, T.; SOOFI, S.; MOINEDDIN, R.; BLACK, R. E.; BHUTTA, Z. A. Geospatial inequalities and determinants of nutritional status among women and children in afghanistan: an observational study. **The Lancet Global Health**, Elsevier, v. 6, n. 4, p. e447–e459, 2018. Citation on page 52.

ALEXANDER, N.; MOYEED, R.; STANDER, J. Spatial modelling of individual-level parasite counts using the negative binomial distribution. **Biostatistics**, Oxford University Press, v. 1, n. 4, p. 453–463, 2000. Citation on page 50.

ALLEN, C.; CHANG, Y.; NEELON, B.; CHANG, W.; KIM, H. J.; LI, Z.; MA, Q.; CHUNG, D. A Bayesian multivariate mixture model for high throughput spatial transcriptomics. **Biometrics**, Wiley Online Library, 2022. Citation on page 57.

ARELLANO-VALLE, R. B.; AZZALINI, A. On the unification of families of skew-normal distributions. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 33, n. 3, p. 561–574, 2006. Citation on page 49.

ARPÒN, J.; SAKAI, K.; GAUDIN, V.; ANDREY, P. Spatial modeling of biological patterns shows multiscale organization of arabidopsis thaliana heterochromatin. **Scientific reports**, Nature Publishing Group, v. 11, n. 1, p. 1–17, 2021. Citation on page 46.

ASWI, A.; CRAMB, S.; DUNCAN, E.; HU, W.; WHITE, G.; MENGERSEN, K. Bayesian spatial survival models for hospitalisation of dengue: A case study of wahidin hospital in makassar, indonesia. **International journal of environmental research and public health**, Multidisciplinary Digital Publishing Institute, v. 17, n. 3, p. 878, 2020. Citation on page 50.

ATTIAS, H. A variational Bayesian framework for graphical models. **Advances in neural information processing systems**, v. 12, 1999. Citation on page 32.

AUCHINCLOSS, A. H.; GEBREAB, S. Y.; MAIR, C.; ROUX, A. V. D. A review of spatial methods in epidemiology, 2000–2010. **Annual review of public health**, Annual Reviews, v. 33, p. 107–122, 2012. Citation on page 25.

AWISZUS, F. Tms and threshold hunting. In: **Supplements to Clinical neurophysiology**. [S.l.]: Elsevier, 2003. v. 56, p. 13–23. Citation on page 39.

AZEVEDO, D. R.; PRATES, M. O.; BANDYOPADHYAY, D. Alleviating spatial confounding in frailty models. **Biostatistics**, 2022. Citations on pages 26, 46, and 159.

BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. **Hierarchical modeling and analysis for spatial data**. [S.l.]: CRC press, 2014. Citations on pages 30, 31, 32, 46, and 48.

BENASSI, F.; NACCARATO, A. Households in potential economic distress. a geographically weighted regression model for italy, 2001–2011. **Spatial Statistics**, Elsevier, v. 21, p. 362–376, 2017. Citation on page 46.

BERGER, J.; GARCÍA, M. J. B. The interplay of Bayesian and frequentist analysis. **Statistical science**, Institute of Mathematical Statistics, n. 1, p. 58–80, 2004. Citation on page 25.

BERGER, J. O.; MORENO, E.; PERICCHI, L. R.; BAYARRI, M. J.; BERNARDO, J. M.; CANO, J. A.; HORRA, J. De la; MARTÍN, J.; RÍOS-INSÚA, D.; BETRÒ, B. *et al.* An overview of robust Bayesian analysis. **Test**, Springer, v. 3, n. 1, p. 5–124, 1994. Citation on page 32.

BERGMANN, T. O. Brain state-dependent brain stimulation. **Frontiers in psychology**, Frontiers Media SA, v. 9, p. 2108, 2018. Citations on pages 37, 97, and 103.

BESAG, J. Spatial interaction and the statistical analysis of lattice systems. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 36, n. 2, p. 192–225, 1974. Citations on pages 47, 49, and 64.

BESAG, J.; YORK, J.; MOLLIÉ, A. Bayesian image restoration, with two applications in spatial statistics. **Annals of the institute of statistical mathematics**, Springer, v. 43, p. 1–20, 1991. Citations on pages 49 and 64.

BLACKWELL, D.; MACQUEEN, J. B. Ferguson distributions via pólya urn schemes. **The annals of statistics**, Institute of Mathematical Statistics, v. 1, n. 2, p. 353–355, 1973. Citation on page 102.

BLANGIARDO, M.; CAMELETTI, M. **Spatial and spatio-temporal Bayesian models with R-INLA**. [S.l.]: John Wiley & Sons, 2015. Citation on page 47.

BOLKER, B. M.; BROOKS, M. E.; CLARK, C. J.; GEANGE, S. W.; POULSEN, J. R.; STEVENS, M. H. H.; WHITE, J.-S. S. Generalized linear mixed models: a practical guide for ecology and evolution. **Trends in ecology & evolution**, Elsevier, v. 24, n. 3, p. 127–135, 2009. Citation on page 30.

BONAT, W. H.; JR, P. J. R.; ZEVIANI, W. M. Likelihood analysis for a class of beta mixed models. **Journal of Applied Statistics**, Taylor & Francis, v. 42, n. 2, p. 252–266, 2015. Citation on page 30.

BRADLEY, J. R.; HOLAN, S. H.; WIKLE, C. K. Computationally efficient multivariate spatio-temporal models for high-dimensional count-valued data (with discussion). **Bayesian Analysis**, v. 13, n. 1, p. 253–310, 2018. Citation on page 49.

BRESLOW, N. E.; CLAYTON, D. G. Approximate inference in generalized linear mixed models. **Journal of the American statistical Association**, Taylor & Francis, v. 88, n. 421, p. 9–25, 1993. Citation on page 29.

BROOKS, S.; GELMAN, A.; JONES, G.; MENG, X.-L. **Handbook of Markov Chain Monte Carlo**. [S.l.]: CRC press, 2011. Citation on page 32.

BURR, I. W. Some approximate relations between terms of the hypergeometric, binomial and poisson distributions. **Communications in Statistics-Theory and Methods**, Taylor & Francis, v. 1, n. 4, p. 297–301, 1973. Citation on page 50.

CLEMENTS, A. C.; LWAMBO, N. J.; BLAIR, L.; NYANDINDI, U.; KAATANO, G.; KINUNG'HI, S.; WEBSTER, J. P.; FENWICK, A.; BROOKER, S. Bayesian spatial analysis and disease mapping: tools to enhance planning and implementation of a schistosomiasis control programme in tanzania. **Tropical medicine & international health**, Wiley Online Library, v. 11, n. 4, p. 490–503, 2006. Citation on page 46.

CRESSIE, N. **Statistics for spatial data**. [S.l.]: John Wiley & Sons, 2015. Citation on page 48.

CRESSIE, N. A. **Statistics for Spatial Data**. [S.l.]: John Wiley & Sons, 1993. Citations on pages 25 and 30.

DEAN, C.; UGARTE, M.; MILITINO, A. Detecting interaction between random region and fixed age effects in disease mapping. **Biometrics**, Wiley Online Library, v. 57, n. 1, p. 197–202, 2001. Citation on page 49.

DERADO, G.; BOWMAN, F. D.; ZHANG, L.; INITIATIVE, A. D. N. Predicting brain activity using a Bayesian spatial model. **Statistical methods in medical research**, Sage Publications Sage UK: London, England, v. 22, n. 4, p. 382–397, 2013. Citation on page 57.

DUAN, Y. **A modified Bayesian power prior approach with applications in water quality evaluation**. Phd Thesis (PhD Thesis) — Virginia Tech, Blacksburg, Virginia, 2005. Citations on pages 33 and 34.

EGBON, O. A.; SOMO-AINA, O.; GAYAWAN, E. Spatial weighted analysis of malnutrition among children in Nigeria: A Bayesian approach. **Statistics in Biosciences**, Springer, p. 1–29, 2021. Citations on pages 49, 50, and 51.

EL-BAZ, A.; FARAG, A. A. Image segmentation using gmrf models: parameters estimation and applications. In: IEEE. **Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)**. [S.l.], 2003. v. 2, p. II–177. Citation on page 47.

ELIERS, P.; MARX, B. Flexible smoothing using b-splines and penalized likelihood (with comments and rejoinder). **Statistical Science**, v. 11, n. 2, p. 89–121, 1996. Citation on page 100.

ESCOBAR, M. D.; WEST, M. Bayesian density estimation and inference using mixtures. **Journal of the American statistical association**, Taylor & Francis, v. 90, n. 430, p. 577–588, 1995. Citations on pages 37 and 102.

FABIO, L. C.; PAULA, G. A.; CASTRO, M. de. A poisson mixed model with nonnormal random effect distribution. **Computational Statistics & Data Analysis**, Elsevier, v. 56, n. 6, p. 1499–1510, 2012. Citation on page 30.

FAIRLEY, L.; FORMAN, D.; WEST, R.; MANDA, S. Spatial variation in prostate cancer survival in the northern and yorkshire region of england using Bayesian relative survival smoothing. **British journal of cancer**, Nature Publishing Group, v. 99, n. 11, p. 1786–1793, 2008. Citation on page 53.

FEISST, M.; KRISAM, J.; KIESER, M. Incorporating historical two-arm data in clinical trials with binary outcome: A practical approach. **Pharmaceutical Statistics**, Wiley Online Library, 2020. Citation on page 25.

FENG, C.; LI, L.; SADEGHPOUR, A. A comparison of residual diagnosis tools for diagnosing regression models for count data. **BMC Medical Research Methodology**, BioMed Central, v. 20, n. 1, p. 1–21, 2020. Citation on page 67.

FERGUSON, T. S. A Bayesian analysis of some nonparametric problems. **The annals of statistics**, JSTOR, p. 209–230, 1973. Citation on page 36.

FINLEY, A. O.; DATTA, A.; COOK, B. D.; MORTON, D. C.; ANDERSEN, H. E.; BANERJEE, S. Efficient algorithms for Bayesian nearest neighbor Gaussian processes. **Journal of Computational and Graphical Statistics**, Taylor & Francis, v. 28, n. 2, p. 401–414, 2019. Citations on pages 78 and 106.

FISHER, R. A. Design of experiments. **British Medical Journal**, BMJ Publishing Group, v. 1, n. 3923, p. 554, 1936. Citation on page 45.

FONTANELLA, L.; IPPOLITI, L.; SARRA, A.; VALENTINI, P.; PALERMI, S. Hierarchical generalized latent spatial quantile regression models with applications to indoor radon concentration. **Stochastic environmental research and risk assessment**, Springer, v. 29, n. 2, p. 357–367, 2015. Citations on pages 49, 50, and 51.

FRAGOSO, T. M.; BERTOLI, W.; LOUZADA, F. Bayesian model averaging: A systematic review and conceptual classification. **International Statistical Review**, Wiley Online Library, v. 86, n. 1, p. 1–28, 2018. Citations on pages 51 and 52.

GARCIA, J. S.; SOLANO, J. J. P.; SERRANO, M. C.; CAMBA, E. A. N.; CASTELL, S. F.; ASENSI, A. S.; SUAY, F. M. Spatial statistical analysis of urban noise data from a wasn gathered by an iot system: Application to a small city. **Applied Sciences**, MDPI, v. 6, n. 12, p. 380, 2016. Citation on page 25.

GELFAND, A. E.; KOTTAS, A.; MACEACHERN, S. N. Bayesian nonparametric spatial modeling with Dirichlet process mixing. **Journal of the American Statistical Association**, Taylor & Francis, v. 100, n. 471, p. 1021–1035, 2005. Citations on pages 25, 37, and 49.

GELMAN, A.; MENG, X.-L.; STERN, H. Posterior predictive assessment of model fitness via realized discrepancies. **Statistica sinica**, JSTOR, p. 733–760, 1996. Citation on page 52.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. **Statistical science**, JSTOR, p. 457–472, 1992. Citation on page 116.

GEWEKE, J. F. *et al.* **Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments**. USA, 1991. Citation on page 92.

GÓMEZ-RUBIO, V. **Bayesian inference with INLA**. [S.l.]: CRC Press, 2020. Citation on page 66.

GONÇALVES, F.; AGUILAR, G.; PRATES, M. Beyond Gaussian processes: Flexible Bayesian modeling and inference for geostatistical processes. **arXiv e-prints**, p. arXiv–2203, 2022. Citations on pages 25, 46, and 49.

GOTWAY, C. A.; YOUNG, L. J. Combining incompatible spatial data. **Journal of the American Statistical Association**, Taylor & Francis, v. 97, n. 458, p. 632–648, 2002. Citation on page 106.

GRACIA, E.; LÓPEZ-QUÍLEZ, A.; MARCO, M.; LLADOSA, S.; LILA, M. The spatial epidemiology of intimate partner violence: do neighborhoods matter? **American journal of epidemiology**, Oxford University Press, v. 182, n. 1, p. 58–66, 2015. Citation on page 46.

GREEN, P. J.; RICHARDSON, S. Hidden markov models and disease mapping. **Journal of the American statistical association**, Taylor & Francis, v. 97, n. 460, p. 1055–1070, 2002. Citation on page 49.

GREGORY, A.; LAU, F. D.-H.; GIROLAMI, M.; BUTLER, L. J.; ELSHAFIE, M. Z. The synthesis of data from instrumented structures and physics-based models via Gaussian processes. **Journal of Computational Physics**, Elsevier, v. 392, p. 248–265, 2019. Citation on page 47.

GUPTA, S. K.; PHUNG, D.; VENKATESH, S. A Bayesian nonparametric joint factor model for learning shared and individual subspaces from multiple data sources. In: SIAM. **Proceedings of the 2012 SIAM International Conference on Data Mining**. [S.l.], 2012. p. 200–211. Citation on page 37.

HALLETT, M. Transcranial magnetic stimulation and the human brain. **Nature**, Nature Publishing Group UK London, v. 406, n. 6792, p. 147–150, 2000. Citation on page 27.

HANKS, E. M.; SCHLIEP, E. M.; HOOTEN, M. B.; HOETING, J. A. Restricted spatial regression in practice: geostatistical models, confounding, and robustness under model misspecification. **Environmetrics**, Wiley Online Library, v. 26, n. 4, p. 243–254, 2015. Citation on page 159.

HOBBS, B. P.; CARLIN, B. P.; MANDREKAR, S. J.; SARGENT, D. J. Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. **Biometrics**, Wiley Online Library, v. 67, n. 3, p. 1047–1056, 2011. Citations on pages 26, 33, 34, and 35.

HOEGH, A.; CRANDELL, I.; KLOPFER, S.; FIES, M. Model selection with missing covariates for policy considerations in fox enclosures. **Journal of Applied Statistics**, Taylor & Francis, v. 44, n. 15, p. 2645–2658, 2017. Citation on page 25.

HSIEH, H.-F.; SHANNON, S. E. Three approaches to qualitative content analysis. **Qualitative health research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 15, n. 9, p. 1277–1288, 2005. Citation on page 45.

HUERTAS, I.; OLDEHINKEL, M.; OORT, E. S. van; GARCIA-SOLIS, D.; MIR, P.; BECKMANN, C. F.; MARQUAND, A. F. A Bayesian spatial model for neuroimaging data based on biologically informed basis functions. **NeuroImage**, Elsevier, v. 161, p. 134–148, 2017. Citation on page 46.

HUTTON, B.; SALANTI, G.; CALDWELL, D. M.; CHAIMANI, A.; SCHMID, C. H.; CAMERON, C.; IOANNIDIS, J. P.; STRAUS, S.; THORLUND, K.; JANSEN, J. P. *et al.* The prisma extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations. **Annals of internal medicine**, American College of Physicians, v. 162, n. 11, p. 777–784, 2015. Citation on page 43.

IBRAHIM, J. G.; CHEN, M.-H. Power prior distributions for regression models. **Statistical Science**, Institute of Mathematical Statistics, v. 15, n. 1, p. 46–60, 2000. Citations on pages 33 and 34.

IBRAHIM, J. G.; CHEN, M.-H.; GWON, Y.; CHEN, F. The power prior: theory and applications. **Statistics in medicine**, Wiley Online Library, v. 34, n. 28, p. 3724–3749, 2015. Citation on page 34.

IBRAHIM, J. G.; CHEN, M.-H.; SINHA, D. **Bayesian survival analysis**. New York: Springer Science & Business Media, 2001. Citations on pages 33, 34, and 82.

ISARD, W. The general theory of location and space-economy. **The Quarterly Journal of Economics**, MIT Press, v. 63, n. 4, p. 476–506, 1949. Citation on page 45.

KALTSA, V.; BRIASSOULI, A.; KOMPATSIARIS, I.; STRINTZIS, M. G. Multiple hierarchical Dirichlet processes for anomaly detection in traffic. **Computer Vision and Image Understanding**, Elsevier, v. 169, p. 28–39, 2018. Citation on page 35.

KANG, J.; JOHNSON, T. D.; NICHOLS, T. E.; WAGER, T. D. Meta analysis of functional neuroimaging data via Bayesian spatial point processes. **Journal of the American Statistical Association**, Taylor & Francis, v. 106, n. 493, p. 124–134, 2011. Citation on page 51.

KARIMI, O.; MOHAMMADZADEH, M. *et al.* Bayesian spatial regression models with closed skew normal correlated errors and missing observations. **Statistical Papers**, Springer Science+ Business Media, Van Godewijckstraat 30 Dordrecht 3311 GX . . . , v. 53, n. 1, p. 205–218, 2012. Citation on page 45.

KARIMINEZHAD, S.; KARHU, J.; SÄISÄNEN, L.; REIJONEN, J.; KÖNÖNEN, M.; JULKUNEN, P. Brain response induced with paired associative stimulation is related to repetition suppression of motor evoked potential. **Brain Sciences**, MDPI, v. 10, n. 10, p. 674, 2020. Citation on page 97.

KIBRIA, B. G.; SUN, L.; ZIDEK, J. V.; LE, N. D. Bayesian spatial prediction of random space-time fields with application to mapping pm2. 5 exposure. **Journal of the American Statistical Association**, Taylor & Francis, v. 97, n. 457, p. 112–124, 2002. Citation on page 51.

KIM, S.; SMYTH, P.; STERN, H. A Bayesian mixture approach to modeling spatial activation patterns in multisite fmri data. **IEEE transactions on medical imaging**, IEEE, v. 29, n. 6, p. 1260–1274, 2010. Citation on page 37.

KLOMJAI, W.; KATZ, R.; LACKMY-VALLÉE, A. Basic principles of transcranial magnetic stimulation (tms) and repetitive tms (rtms). **Annals of physical and rehabilitation medicine**, Elsevier, v. 58, n. 4, p. 208–213, 2015. Citation on page 119.

KNORR-HELD, L.; RASSER, G. Bayesian detection of clusters and discontinuities in disease maps. **Biometrics**, Wiley Online Library, v. 56, n. 1, p. 13–21, 2000. Citation on page 49.

KOTTAS, A. Dirichlet process mixtures of beta distributions, with applications to density and intensity estimation. In: **Workshop on Learning with Nonparametric Bayesian Methods, 23rd International Conference on Machine Learning (ICML)**. [S.l.: s.n.], 2006. v. 47. Citation on page 35.

KOUTSOURELAKIS, P.-S. A multi-resolution, non-parametric, Bayesian framework for identification of spatially-varying model parameters. **Journal of computational physics**, Elsevier, v. 228, n. 17, p. 6184–6211, 2009. Citation on page 47.

_____. A novel Bayesian strategy for the identification of spatially varying material properties and model validation: an application to static elastography. **International Journal for Numerical Methods in Engineering**, Wiley Online Library, v. 91, n. 3, p. 249–268, 2012. Citation on page 47.

KOZUBOWSKI, T. J.; PODGÓRSKI, K. A multivariate and asymmetric generalization of laplace distribution. **Computational Statistics**, Springer, v. 15, p. 531–540, 2000. Citation on page 49.

KRAINSKI, E. T.; GÓMEZ-RUBIO, V.; BAKKA, H.; LENZI, A.; CASTRO-CAMILO, D.; SIMPSON, D.; LINDGREN, F.; RUE, H. **Advanced spatial modeling with stochastic partial differential equations using R and INLA**. [S.l.]: CRC press, 2018. Citation on page 66.

KRIGE, D. G.; KRIGE, D. **Lognormal-de Wijsian geostatistics for ore evaluation**. [S.l.]: South African Institute of mining and metallurgy Johannesburg, 1981. Citation on page 46.

KRISZTIN, T.; PIRIBAUER, P.; WÖGERER, M. A spatial multinomial logit model for analysing urban expansion. **Spatial Economic Analysis**, Taylor & Francis, p. 1–22, 2021. Citation on page 50.

KWAK, S. G.; KIM, J. H. Central limit theorem: the cornerstone of modern statistics. **Korean journal of anesthesiology**, Korean Society of Anesthesiologists, v. 70, n. 2, p. 144, 2017. Citation on page 50.

LANG, S.; BREZGER, A. Bayesian p-splines. **Journal of computational and graphical statistics**, Taylor & Francis, v. 13, n. 1, p. 183–212, 2004. Citations on pages 78 and 101.

LAW, J.; QUICK, M.; JADAVJI, A. A Bayesian spatial shared component model for identifying crime-general and crime-specific hotspots. **Annals of GIS**, Taylor & Francis, v. 26, n. 1, p. 65–79, 2020. Citation on page 47.

LEE, D. A comparison of conditional autoregressive models used in Bayesian disease mapping. **Spatial and spatio-temporal epidemiology**, Elsevier, v. 2, n. 2, p. 79–89, 2011. Citation on page 64.

LEE, D.; MITCHELL, R. Boundary detection in disease mapping studies. **Biostatistics**, Oxford University Press, v. 13, n. 3, p. 415–426, 2012. Citation on page 49.

LEE, D.-J.; DURBÁN, M. Smooth-car mixed models for spatial count data. **Computational Statistics & Data Analysis**, Elsevier, v. 53, n. 8, p. 2968–2979, 2009. Citation on page 49.

LEE, J.; BAHRI, Y.; NOVAK, R.; SCHOENHOLZ, S. S.; PENNINGTON, J.; SOHL-DICKSTEIN, J. Deep neural networks as Gaussian processes. **arXiv preprint arXiv:1711.00165**, 2017. Citation on page 47.

LEFAUCHEUR, J.-P.; ANDRÉ-OBADIA, N.; ANTAL, A.; AYACHE, S. S.; BAEKEN, C.; BENNINGER, D. H.; CANTELLO, R. M.; CINCOTTA, M.; CARVALHO, M. de; RIDDER, D. D. *et al.* Evidence-based guidelines on the therapeutic use of repetitive transcranial magnetic stimulation (rtms). **Clinical Neurophysiology**, Elsevier, v. 125, n. 11, p. 2150–2206, 2014. Citation on page 27.

LEININGER, T. J.; GELFAND, A. E. Bayesian inference and model assessment for spatial point patterns using posterior predictive samples. **Bayesian Analysis**, International Society for Bayesian Analysis, v. 12, n. 1, p. 1–30, 2017. Citation on page 48.

LEROUX, B. G.; LEI, X.; BRESLOW, N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: SPRINGER. **Statistical models in epidemiology, the environment, and clinical trials**. [S.l.], 2000. p. 179–191. Citations on pages 49 and 65.

LI, Q.; ZHANG, M.; XIE, Y.; XIAO, G. Bayesian modeling of spatial molecular profiling data via Gaussian process. **Bioinformatics**, Oxford Academic, v. 37, n. 22, p. 4129–4136, 2021. Citations on pages 46 and 50.

LINDGREN, F.; RUE, H.; LINDSTRÖM, J. An explicit link between Gaussian fields and Gaussian markov random fields: the stochastic partial differential equation approach. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 73, n. 4, p. 423–498, 2011. Citations on pages 49 and 65.

LIU, J.; WADE, S.; BOCHKINA, N. Shared differential clustering across single-cell rna sequencing datasets with the hierarchical Dirichlet process. **arXiv preprint arXiv:2212.02505**, 2022. Citation on page 37.

LOPES, H. F.; MÜLLER, P.; ROSNER, G. L. Bayesian meta-analysis for longitudinal data models using multivariate mixture priors. **Biometrics**, Wiley Online Library, v. 59, n. 1, p. 66–75, 2003. Citation on page 26.

LÓPEZ, F. O. A Bayesian approach to parameter estimation in simplex regression model: a comparison with beta regression. **Revista Colombiana de Estadística**, Universidad Nacional de Colombia., v. 36, n. 1, p. 1–21, 2013. Citation on page 89.

LOUZADA, F.; NASCIMENTO, D. C. d.; EGBON, O. A. Spatial statistical models: an overview under the Bayesian approach. **Axioms**, MDPI, v. 10, n. 4, p. 307, 2021. Citation on page 30.

LUAN, H.; LAW, J.; LYSY, M. Diving into the consumer nutrition environment: A Bayesian spatial factor analysis of neighborhood restaurant environment. **Spatial and spatio-temporal epidemiology**, Elsevier, v. 24, p. 39–51, 2018. Citation on page 46.

MOALA, F. A.; O'HAGAN, A. Elicitation of multivariate prior distributions: A nonparametric Bayesian approach. **Journal of Statistical Planning and Inference**, Elsevier, v. 140, n. 7, p. 1635–1655, 2010. Citation on page 49.

MOHER, D.; LIBERATI, A.; TETZLAFF, J.; ALTMAN, D. G.; ALTMAN, D.; ANTES, G.; ATKINS, D.; BARBOUR, V.; BARROWMAN, N.; BERLIN, J. A. *et al.* Preferred reporting items for systematic reviews and meta-analyses: The prisma statement (chinese edition). **Journal of Chinese Integrative Medicine**, Science Press (China), v. 7, n. 9, p. 889–896, 2009. Citation on page 43.

MOON, T. K. The expectation-maximization algorithm. **IEEE Signal processing magazine**, IEEE, v. 13, n. 6, p. 47–60, 1996. Citation on page 33.

MORRIS, R. S. Diseases, dilemmas, decisions—converting epidemiological dilemmas into successful disease control decisions. **Preventive veterinary medicine**, Elsevier, v. 122, n. 1-2, p. 242–252, 2015. Citation on page 46.

MORRIS, T. P.; WHITE, I. R.; CROWTHER, M. J. Using simulation studies to evaluate statistical methods. **Statistics in medicine**, Wiley Online Library, v. 38, n. 11, p. 2074–2102, 2019. Citation on page 51.

MÜLLER, I.; BETUELA, I.; HIDE, R. Regional patterns of birthweights in papua new guinea in relation to diet, environment and socio-economic factors. **Annals of human biology**, Taylor & Francis, v. 29, n. 1, p. 74–88, 2002. Citation on page 46.

MÜLLER, P.; QUINTANA, F.; ROSNER, G. A method for combining inference across related nonparametric Bayesian models. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 66, n. 3, p. 735–749, 2004. Citation on page 26.

MUNOZ, F.; PENNINO, M. G.; CONESA, D.; LÓPEZ-QUÍLEZ, A.; BELLIDO, J. M. Estimation and prediction of the spatial occurrence of fish species using Bayesian latent Gaussian models. **Stochastic Environmental Research and Risk Assessment**, Springer, v. 27, n. 5, p. 1171–1180, 2013. Citations on pages 48 and 49.

NAJIB, U.; BASHIR, S.; EDWARDS, D.; ROTENBERG, A.; PASCUAL-LEONE, A. Transcranial brain stimulation: clinical applications and future directions. **Neurosurgery Clinics**, Elsevier, v. 22, n. 2, p. 233–251, 2011. Citation on page 119.

NEILL, D.; MOORE, A.; COOPER, G. A Bayesian spatial scan statistic. **Advances in neural information processing systems**, Citeseer, v. 18, p. 1003–1010, 2005. Citation on page 51.

NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. **Journal of the Royal Statistical Society: Series A (General)**, Wiley Online Library, v. 135, n. 3, p. 370–384, 1972. Citations on pages 29 and 30.

NETER, J.; KUTNER, M. H.; NACHTSHEIM, C. J.; WASSERMAN, W. *et al.* Applied linear statistical models. Irwin Chicago, 1996. Citation on page 29.

OBAROMI, D. Spatial modelling of some conditional autoregressive priors in a disease mapping model: the Bayesian approach. **Biomedical Journal of Scientific & Technical Research**, v. 14, n. 3, 2019. Citation on page 49.

OLIVEIRA, V. D.; ECKER, M. D. A non–stationary non–Gaussian hedonic spatial model for house selling prices. **Communications in Statistics-Simulation and Computation**, Taylor & Francis, v. 51, n. 6, p. 2888–2905, 2022. Citation on page 49.

ORDOÑEZ, J. A.; PRATES, M. O.; MATOS, L. A.; LACHOS, V. H. Objective Bayesian analysis for geostatistical student-t processes. **Journal of Spatial Science**, Taylor & Francis, p. 1–19, 2023. Citations on pages 25, 49, 56, and 98.

OSAMA, A.; SAYED, T. A novel approach for identifying, diagnosing, and treating active transportation safety issues. **Transportation research record**, Sage Publications Sage CA: Los Angeles, CA, v. 2673, n. 11, p. 813–823, 2019. Citation on page 46.

PACIOREK, C. J. The importance of scale for spatial-confounding bias and precision of spatial regression estimators. **Statistical science: a review journal of the Institute of Mathematical Statistics**, NIH Public Access, v. 25, n. 1, p. 107, 2010. Citation on page 159.

PALACIOS, M. B.; STEEL, M. F. J. Non-Gaussian Bayesian geostatistical modeling. **Journal of the American Statistical Association**, Taylor & Francis, v. 101, n. 474, p. 604–618, 2006. Citation on page 50.

PALMÍ-PERALES, F.; GÓMEZ-RUBIO, V.; MARTINEZ-BENEITO, M. A. Bayesian multivariate spatial models for lattice data with inla. **Journal of Statistical Software**, v. 98, p. 1–29, 2021. Citation on page 26.

PEREIRA, C. A. de B.; STERN, J. M. Evidence and credibility: full Bayesian significance test for precise hypotheses. **Entropy**, Molecular Diversity Preservation International, v. 1, n. 4, p. 99–110, 1999. Citation on page 67.

PERPEROGLOU, A.; SAUERBREI, W.; ABRAHAMOWICZ, M.; SCHMID, M. A review of spline function procedures in r. **BMC medical research methodology**, BioMed Central, v. 19, n. 1, p. 1–16, 2019. Citation on page 76.

PILZ, J.; KAZIANKA, H.; SPÖCK, G. Some advances in Bayesian spatial prediction and sampling design. **Spatial Statistics**, Elsevier, v. 1, p. 65–81, 2012. Citation on page 49.

POTTER, J. E.; SCHMERTMANN, C. P.; ASSUNÇÃO, R. M.; CAVENAGHI, S. M. Mapping the timing, pace, and scale of the fertility transition in brazil. **Population and development review**, Wiley Online Library, v. 36, n. 2, p. 283–307, 2010. Citation on page 47.

PRATES, M. O.; DEY, D. K.; WILLIG, M. R.; YAN, J. Transformed Gaussian markov random fields and spatial modeling of species abundance. **Spatial Statistics**, Elsevier, v. 14, p. 382–399, 2015. Citation on page 49.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. Available: <https://www.R-project.org/>. Citation on page 66.

RABAOUI, A.; VIANDIER, N.; DUFLOS, E.; MARAIS, J.; VANHEEGHE, P. Dirichlet process mixtures for density estimation in dynamic nonlinear modeling: Application to gps positioning in urban canyons. **IEEE Transactions on Signal Processing**, IEEE, v. 60, n. 4, p. 1638–1655, 2011. Citation on page 35.

RAO, R. P. **Brain-computer interfacing: an introduction**. [S.l.]: Cambridge University Press, 2013. Citation on page 69.

RAPPEL, H.; WU, L.; NOELS, L.; BEEX, L. A. A Bayesian framework to identify random parameter fields based on the copula theorem and Gaussian fields: Application to polycrystalline materials. **Journal of Applied Mechanics**, American Society of Mechanical Engineers, v. 86, n. 12, p. 121009, 2019. Citation on page 47.

REICH, B. J.; BANDYOPADHYAY, D.; BONDELL, H. D. A nonparametric spatial model for periodontal data with nonrandom missingness. **Journal of the American Statistical Association**, Taylor & Francis, v. 108, n. 503, p. 820–831, 2013. Citations on pages 77 and 100.

RILEY, S.; EAMES, K.; ISHAM, V.; MOLLISON, D.; TRAPMAN, P. Five challenges for spatial epidemic models. **Epidemics**, Elsevier, v. 10, p. 68–71, 2015. Citation on page 45.

ROBINSON, N.; RAMPANT, P.; CALLINAN, A.; RAB, M.; FISHER, P. Advances in precision agriculture in south-eastern australia. ii. spatio-temporal prediction of crop yield using terrain derivatives and proximally sensed data. **Crop and Pasture Science**, CSIRO Publishing, v. 60, n. 9, p. 859–869, 2009. Citation on page 47.

ROTENBERG, A.; HORVATH, J. C.; PASCUAL-LEONE, A. The transcranial magnetic stimulation (tms) device and foundational techniques. In: **Transcranial magnetic stimulation**. [S.l.]: Springer, 2014. p. 3–13. Citations on pages 27 and 37.

ROUSSEEUW, P. J.; CROUX, C. Alternatives to the median absolute deviation. **Journal of the American Statistical Association**, Taylor & Francis, v. 88, n. 424, p. 1273–1283, 1993. Citation on page 122.

RUE, H.; HELD, L. **Gaussian Markov Random Fields: Theory and Applications**. [S.l.]: CRC Press, 2005. Citation on page 25.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate Bayesian inference for latent Gaussian models by using integrated nested laplace approximations. **Journal of the royal statistical society: Series b (statistical methodology)**, Wiley Online Library, v. 71, n. 2, p. 319–392, 2009. Citations on pages 26, 32, 49, and 60.

RUE, H.; TJELMELAND, H. Fitting Gaussian markov random fields to Gaussian fields. **Scandinavian Journal of Statistics**, Wiley Online Library, v. 29, n. 1, p. 31–49, 2002. Citation on page 48.

SAAVEDRA, P.; SANTANA, A.; BELLO, L.; PACHECO, J.-M.; SANJUÁN, E. A Bayesian spatio-temporal analysis of mortality rates in spain: application to the covid-19 2020 outbreak. **Population Health Metrics**, BioMed Central, v. 19, n. 1, p. 1–10, 2021. Citation on page 46.

SCHLUTH, C. G.; STANDLEY, C. J.; BANSAL, S.; CARLSON, C. J. Spatial parasitology and the unmapped human helminthiases. **Parasitology**, Cambridge University Press, p. 1–30, 2023. Citation on page 25.

SERNEELS, S.; LAMBIN, E. F. Proximate causes of land-use change in narok district, kenya: a spatial statistical model. **Agriculture, Ecosystems & Environment**, Elsevier, v. 85, n. 1-3, p. 65–81, 2001. Citation on page 25.

SETHURAMAN, J. A constructive definition of Dirichlet priors. **Statistica sinica**, JSTOR, p. 639–650, 1994. Citations on pages 36 and 101.

SHARKEY, P.; WINTER, H. C. A Bayesian spatial hierarchical model for extreme precipitation in great britain. **Environmetrics**, Wiley Online Library, v. 30, n. 1, p. e2529, 2019. Citations on pages 47, 48, and 53.

SHORT, M.; CARLIN, B. P.; BUSHHOUSE, S. Using hierarchical spatial models for cancer control planning in minnesota (united states). **Cancer Causes & Control**, Springer, v. 13, p. 903–916, 2002. Citation on page 46.

SIDÉN, P.; LINDSTEN, F. Deep Gaussian markov random fields. In: PMLR. **International Conference on Machine Learning**. [S.l.], 2020. p. 8916–8926. Citation on page 47.

SILVA, A. R. F. D. A Dirichlet process mixture model for brain mri tissue classification. **Medical image analysis**, Elsevier, v. 11, n. 2, p. 169–182, 2007. Citation on page 37.

SILVA-BUTTKUS, P. D.; MARCELLI, G.; FRANKS, S.; STARK, J.; HARDY, K. Inferring biological mechanisms from spatial analysis: prediction of a local inhibitor in the ovary. **Proceedings of the National Academy of Sciences**, National Acad Sciences, v. 106, n. 2, p. 456–461, 2009. Citation on page 46.

SILVA, L. M.; SILVA, K. M. S.; LIRA-BANDEIRA, W. G.; COSTA-RIBEIRO, A. C.; ARAÚJO-NETO, S. A. Localizing the primary motor cortex of the hand by the 10-5 and 10-20 systems for neurostimulation: An mri study. **Clinical EEG and Neuroscience**, SAGE Publications Sage CA: Los Angeles, CA, p. 1550059420934590, 2020. Citations on pages 39 and 97.

SIMÕES, P.; CARVALHO, M.; ALEIXO, S.; GOMES, S.; NATÁRIO, I. A spatial econometric analysis of the calls to the portuguese national health line. **Econometrics**, Multidisciplinary Digital Publishing Institute, v. 5, n. 2, p. 24, 2017. Citations on pages 46 and 50.

SIMPSON, D.; ILLIAN, J. B.; LINDGREN, F.; SØRBYE, S. H.; RUE, H. Going off grid: Computationally efficient inference for log-Gaussian cox processes. **Biometrika**, Oxford University Press, v. 103, n. 1, p. 49–70, 2016. Citations on pages 56, 76, and 100.

SIMPSON, D.; RUE, H.; RIEBLER, A.; MARTINS, T. G.; SØRBYE, S. H. Penalising model component complexity: A principled, practical approach to constructing priors. **Statistical Science**, v. 32, n. 1, p. 1–28, 2017. Citations on pages 49 and 66.

SONG, Y.; NATHOO, F.; BABUL, A. A potts-mixture spatiotemporal joint model for combined magnetoencephalography and electroencephalography data. **Canadian Journal of Statistics**, Wiley Online Library, v. 47, n. 4, p. 688–711, 2019. Citation on page 57.

SPARKS, P. J.; SPARKS, C. S.; CAMPBELL, J. J. An application of Bayesian spatial statistical methods to the study of racial and poverty segregation and infant mortality rates in the us. **GeoJournal**, Springer, v. 78, p. 389–405, 2013. Citation on page 45.

STAUBACH, C.; SCHMID, V.; KNORR-HELD, L.; ZILLER, M. A Bayesian model for spatial wildlife disease prevalence data. **Preventive veterinary medicine**, Elsevier, v. 56, n. 1, p. 75–87, 2002. Citation on page 50.

STINEAR, C. M.; BYBLOW, W. D.; ACKERLEY, S. J.; SMITH, M.-C.; BORGES, V. M.; BARBER, P. A. Prep2: A biomarker-based algorithm for predicting upper limb function after stroke. **Annals of clinical and translational neurology**, Wiley Online Library, v. 4, n. 11, p. 811–820, 2017. Citation on page 27.

STROUP, W. W. **Generalized linear mixed models: modern concepts, methods and applications**. [S.l.]: CRC press, 2012. Citation on page 30.

SUN, W.; LE, N. D.; ZIDEK, J. V.; BURNETT, R. Assessment of a Bayesian multivariate interpolation approach for health impact studies. **Environmetrics: The official journal of the International Environmetrics Society**, Wiley Online Library, v. 9, n. 5, p. 565–586, 1998. Citations on pages 51 and 52.

TAYLOR, B. M.; DIGGLE, P. J. Inla or mcmc? a tutorial and comparative evaluation for spatial prediction in log-Gaussian cox processes. **Journal of Statistical Computation and Simulation**, Taylor & Francis, v. 84, n. 10, p. 2266–2284, 2014. Citation on page 54.

TEERAPABOLARN, K.; JAIOUN, K. Approximation of binomial distribution by an improved poisson distribution. **Int. J. Pure Appl. Math**, v. 97, p. 491–495, 2014. Citation on page 50.

TEH, Y. W. *et al.* Dirichlet process. **Encyclopedia of machine learning**, v. 1063, p. 280–287, 2010. Citations on pages 35 and 36.

VEMULAPALLI, R.; TUZEL, O.; LIU, M.-Y. Deep Gaussian conditional random field network: A model-based deep network for discriminative denoising. In: **Proceedings of the IEEE conference on computer vision and pattern recognition**. [S.l.: s.n.], 2016. p. 4801–4809. Citation on page 47.

WALKER, M.; CURTIS, A. Eliciting spatial statistics from geological experts using genetic algorithms. **Geophysical Journal International**, Oxford University Press, v. 198, n. 1, p. 342–356, 2014. Citation on page 49.

WARD, M. Spatial epidemiology: Where have we come in 150 years? **Geospatial technologies and homeland security: research frontiers and future challenges**, Springer, p. 257–282, 2008. Citation on page 46.

WOODS, J. Two-dimensional discrete markovian fields. **IEEE Transactions on Information Theory**, IEEE, v. 18, n. 2, p. 232–240, 1972. Citation on page 47.

XIE, M.; CAI, J.; LIU, Y.; WEI, W.; ZHAO, Z.; DAI, M.; WU, Y.; HUANG, Y.; TANG, Y.; XIAO, L. *et al.* Association between childhood trauma and white matter deficits in first-episode schizophrenia. **Psychiatry Research**, Elsevier, p. 115111, 2023. Citation on page 25.

YAU, K. K.; LEE, A. H.; NG, A. S. Theory & methods: A zero-augmented gamma mixed model for longitudinal data with many zeros. **Australian & New Zealand Journal of Statistics**, Wiley Online Library, v. 44, n. 2, p. 177–183, 2002. Citation on page 30.

YOUSEFI, K.; SWARTZ, T. B. Advanced putting metrics in golf. **Journal of Quantitative Analysis in Sports**, De Gruyter, v. 9, n. 3, p. 239–248, 2013. Citation on page 47.

7.3 Appendix of Chapter 3

Table 11 – List of Questions for Conceptual Classification Scheme

QUESTION 1	Is it only an application?
1.1	Yes
1.2	Both
1.3	No (only method)
QUESTION 2	What is the field of application?
2.1	Medical science
2.2	Economics and Humanity
2.3	Physical Science and Engineering
2.4	Agricultural and Environmental Science
2.5	Sport
QUESTION 3	What Spatial domain was employed?
3.1	Area or lattice
3.2	Geostatistical data
3.3	Spatial point patterns
3.4	Area and Geostatistical data
QUESTION 4	What type of spatial priors are used?
4.1	Conditional Autoregressive (CAR)
4.2	Besag York Mollié (BYM)
4.3	Leroux CAR
4.4	Gaussian Markov Random Field (Other specifications)
4.5	Covariance Function (Not GMRF)
4.6	Other (new methodology/proposed)
QUESTION 5	What type of response variable is used?
5.1	Discrete (Countable)
5.2	Continuous
5.3	Combined (Mixed)
5.4	Ordinal
QUESTION 6	What is the statistical model used?
6.1	Generalize Linear (mixed) model (or Hierarchical models)
6.2	Survival and Longitudinal models
6.3	Non-parametric models (Machine Learning models)
6.4	Spatial Econometrics
6.5	Proposed
6.6	Not stated
6.7	Other
QUESTION 7	How are model Prior specified?
7.1	Vague prior (Non-informative)
7.2	Used verbatim from the literature
7.3	Elicitated from experts or from the problem
7.4	No explicit use or reference/not applicable
QUESTION 8	What is the estimation method applied?
8.1	Markov Chain Monte Carlo (MCMC)
8.2	Integrated Nested Laplace Approximation (INLA)
8.3	Expectation-Maximization (EM)
8.4	Maximum (Penalized quasi) Likelihood Method
8.5	Not stated
8.6	Other
QUESTION 9	Is the model validated through simulation?
9.1	Yes
9.2	No
QUESTION 10	Is the application validated through data-driven procedures?
10.1	Cross-validation and data splitting (K-fold / Holdout)
10.2	Leave-One-Out Cross-Validation (LOOCV)
10.3	Posterior predictive check
10.4	Other
10.5	None or not applicable

7.4 Appendix of Chapter 5

7.4.1 Fill-reducing permutation matrix

Let $\mathbf{P}$ be a fill-reducing permutation matrix such that the Cholesky decomposition $\mathbf{L}$ of a sparse matrix $\boldsymbol{\Omega}$ has the least number of non-zero elements and $\mathbf{L}\mathbf{L}^T = \mathbf{P}\boldsymbol{\Omega}\mathbf{P}^T$ and the corresponding inverse is $\mathbf{L}^{-\mathbf{T}}\mathbf{L}^{-\mathbf{1}} = \mathbf{P}\boldsymbol{\Omega}^{-1}\mathbf{P}^T$. Since $\mathbf{L}$ has fewer non-zero elements, the computation cost of $\boldsymbol{\Omega}^{-1}$ is cheaper. In general, the modification is compensated for in a quadratic form. Suppose $\mathbf{r}$ is a conforming vector to $\boldsymbol{\Omega}$, the quadratic form $\mathbf{r}^T\boldsymbol{\Omega}\mathbf{r} = (\mathbf{P}\mathbf{r})^T\mathbf{L}\mathbf{L}^T\mathbf{P}\mathbf{r}$ is computationally cheaper to evaluate and the $det(\boldsymbol{\Omega}) = (det(\mathbf{L}))^2$. This technique was adopted in the estimation algorithm, presented in the next section.

7.4.2 Metropolis-Hasting within Gibbs Algorithm

7.4.2.1 Under joint spatial power prior

The posterior distribution derived from the model based on the spatial power prior distribution allows inference to be made on the current experiment while borrowing historical knowledge through the parameter ω_0. The hyper-parameters involved in the model are $a_\tau, b_\tau, \sigma^2_\beta, \sigma^2_\psi, \sigma^2_\theta, \tau_x, \mu_\phi, \sigma^2_\phi, \mu_\nu$, and σ^2_ν. To make an inference, the hyper-parameters are either fixed or elicited based on previous studies. If fixed, the hyper-parameters are chosen such that the prior distribution entails the true parameter. At the start of the chain, the initial values of every random parameter in the model are drawn from their respective prior distributions.

The main effect $\mathscr{X}$ is first updated using the posterior conditional distribution. That is $\mathscr{X}$ is then updated from a multivariate normal distribution given as $\mathscr{X} \mid . \sim N(\boldsymbol{\mu}_{\mathscr{X}}, \boldsymbol{\Sigma}_{\mathscr{X}})$, where $\boldsymbol{\Sigma}_{\mathscr{X}} = \tau^{-1}(\sum_{j=1}^n \mathbf{V}_j^T\mathbf{V}_j + \mathbf{D}^{-1})^{-1}$ and $\boldsymbol{\mu}_{\mathscr{X}} = \tau\boldsymbol{\Sigma}_{\mathscr{X}}(\sum_{j=1}^n \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{B})$. The matrix $\mathbf{D}^{-1}$ and $\sum_{j=1}^n \mathbf{V}_j^T\mathbf{V}_j$ are sparse matrices; however, they do not guarantee the sparseness of $\boldsymbol{\Sigma}_{\mathscr{X}}$. Since $\boldsymbol{\Sigma}_{\mathscr{X}}$ is a function of τ and ω_0, the matrix is updated at every step in the MCMC chain. To improve on the computational burden, we find a "one-time" optimal fill-reducing permutation matrix $\mathbf{P}_{\mathscr{X}}$ of $\boldsymbol{\Sigma}_{\mathscr{X}}$, and determine the Cholesky decomposition $\mathbf{L}_{\mathscr{X}}$, such that $\mathbf{L}_{\mathscr{X}}\mathbf{L}_{\mathscr{X}}^T = \mathbf{P}_{\mathscr{X}}\boldsymbol{\Sigma}_{\mathscr{X}}\mathbf{P}_{\mathscr{X}}^T$. Draw $q^\star = p+r+q$ independent uniform samples in the interval $(0,1)$ to form a vector $\mathbf{u}$. Compute $\mathbf{u}_{\mathscr{X}} = \mathbf{P}_{\mathscr{X}}^T\mathbf{L}_{\mathscr{X}}^{-T}\mathbf{u}$ and $\boldsymbol{\mu}_{\mathscr{X}} = \tau\mathbf{P}^T\mathbf{L}_{\mathscr{X}}\mathbf{L}_{\mathscr{X}}^T\mathbf{P}_{\mathscr{X}}(\sum_{j=1}^n \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{B})$ and set $\mathscr{X} = \boldsymbol{\mu}_{\mathscr{X}} + \mathbf{u}_{\mathscr{X}}$. In this setup, the fill-reducing permutation matrix is computed once throughout the MCMC chain.

Similarly, $\boldsymbol{\vartheta}$ is updated from the posterior conditional distribution. Assuming a multivariate normal distribution prior for $\boldsymbol{\vartheta} \sim \pi(\mathbf{0}, \tau^{-1}\mathbf{R})$, the conditional posterior distribution is given as $\boldsymbol{\vartheta} \mid . \sim N(\boldsymbol{\mu}_{\vartheta}, \boldsymbol{\Sigma}_{\vartheta})$, where $\boldsymbol{\Sigma}_{\vartheta} = \tau^{-1}(\mathbf{R}^{-1} + \sum_{j=1}^n \mathbf{U}_j\mathbf{U}_j^T)^{-1}$ and $\boldsymbol{\mu}_{\vartheta} = \tau\boldsymbol{\Sigma}_{\vartheta}\sum_{j=1}^n \mathbf{U}_j(\mathbf{Y}_j - \mathbf{V}_j\mathscr{X})$. Next, we update the precision parameter of the current data from its posterior conditional distribution. The conditional posterior distribution is given as $\tau \mid . \sim Gamma(e_\tau, f_\tau)$, where $e_\tau = \frac{(mn+m_0n_0\omega_0+q^\star+p_0+2a_\tau)}{2}$ and $f_\tau = \frac{1}{2}\Big(\sum_{j=1}^n(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X})^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X}) + (\mathscr{X} - \mathbf{D}\mathbf{B})^T\mathbf{D}^{-1}(\mathscr{X} - \mathbf{D}\mathbf{B}) + k + \boldsymbol{\vartheta}^T\mathbf{R}^{-1}\boldsymbol{\vartheta}\Big)$. A similar technique, using the permutation matrix, can be adopted to evaluate $\mathbf{D}\mathbf{B}$. Since $\mathbf{D}^{-1}$ is a sparse matrix, it is possible to obtain a fill-reducing permutation matrix $\mathbf{P}_D$. Thus, $\mathbf{D}\mathbf{B}$ can be evaluated as $\mathbf{P}_D^T\mathbf{L}_D\mathbf{L}_D^T\mathbf{P}_D\mathbf{B}$, where $\mathbf{L}_D$ is a Cholesky decomposition of $\mathbf{P}_D\mathbf{D}\mathbf{P}_D^T$.

The power parameter ω_0 is then updated from its posterior conditional distribution. Unlike the parameters previously updated, there is no close form for the posterior conditional distribution. Thus, we update ω_0 using the Metropolis-Hasting algorithm. The posterior conditional distribution follows as

$$\pi(\omega_0 \mid .) \propto \frac{(2k)^{\frac{\nu}{2}}\tau^{(m_0n_0\omega_0)/2}}{\Gamma(\frac{\nu}{2}) \mid \mathbf{D} \mid^{1/2}} \exp\Big\{ -\frac{1}{2}\Big(\tau(\mathscr{X} - \mathbf{D}\mathbf{B})^T\mathbf{D}^{-1}(\mathscr{X} - \mathbf{D}\mathbf{B}) + \tau k\Big)\Big\}\pi(\omega_0),$$

where $\pi(\omega_0)$ is the prior distribution. Notice that $\pi(\omega_0 \mid .)$ is indirectly linked to the current data through τ and $\mathscr{X}$, by explaining how much the historical data explains these parameters. The computation of the determinant of $\mathbf{D}$ is less expensive using $\mathbf{L}_D$ since the permutation matrix is only computed once. That is, $det(\mathbf{D})^{1/2} = det(\mathbf{L}_D)$.

An alternative method to compute the determinant is to compute the determinant of the sparse matrix $\mathbf{D}^{-1}$, and set $det(\mathbf{D})^{1/2} = det(\mathbf{D}^{-1})^{-1/2}$. A truncated normal distribution in the interval of $(0,1)$ was used for the proposal distribution. The Metropolis-Hasting step follows. Suppose that the chain is at t, sample from the proposal distribution $\omega_0^{(t*)} \sim q_{\omega_0}(.|\omega_0^{(t-1)})$ and compute the acceptance probability

$$\alpha_{\omega_0} = min\left(1, \frac{\pi(\omega_0^{(t*)}|.)q_{\omega_0}(\omega_0^{(t-1)}|\omega_0^{(t*)})}{\pi(\omega_0^{(t-1)}|.)q_{\omega_0}(\omega_0^{(t*)}|\omega_0^{(t-1)})}\right).$$

Draw $U \sim Uniform(0,1)$ and set $\omega_0^{(t)} = \omega_0^{(t*)}$ if $U < \alpha_{\omega_0}$ and set $\omega_0^{(t)} = \omega_0^{(t-1)}$ otherwise. A Beta distribution given in Equation 5.13 was considered for $\pi(\omega_0)$. Thus, the hyper-parameters were considered random and assigned prior distributions. Update the hyperparameters as follows. Draw a proposal $\sigma_{0b}^{(t*)} \sim q_{\sigma_{0b}}(.|\mu_{0b}) = Uniform(0, \mu_{0b}^{(t-1)}(1-\mu_{0b}^{t-1}))$ and compute the acceptance probability

$$\alpha_{\sigma_{0b}} = min\left(1, \frac{\pi(\sigma_0^{(t*)}|.)q_{\sigma_{0b}}(\sigma_{0b}^{(t-1)}|\mu_{0b})}{\pi(\sigma_0^{(t-1)}|.)q_{\sigma_{0b}}(\sigma_{0b}^{(t*)}|\mu_{0b})}\right).$$

Draw $U \sim Uniform(0,1)$ and set $\sigma_{0b}^{(t)} = \sigma_{0b}^{(t*)}$ if $U < \alpha_{\sigma_{0b}}$, otherwise set $\sigma_{0b}^{(t)} = \sigma_{0b}^{(t-1)}$, where

$$\pi(\sigma_{0b}|.) \propto \omega_0^{\mu_{0b}\left(\frac{\mu_{0b}(1-\mu_{0b})}{\sigma_{0b}}-1\right)-1}(1-\omega_0)^{\mu_{0b}\left(\frac{1-\mu_{0b}}{\mu_{0b}}\right)\left(\frac{\mu_{0b}(1-\mu_{0b})}{\sigma_{0b}}-1\right)-1}\mathbf{I}(\sigma_{0b} < \mu_{0b}(1-\mu_{0b})),$$

and $\mathbf{I}(.)$ denotes an indicator function.

To update μ_{0b}, draw a proposal $\mu_{0b}^{(t*)} \sim q_{\mu_{0b}}(.|a_0,b_0) = Uniform(a_0,b_0)$, $0 < a_0 < b_0 < 1$ until $\mu_{0b}^{(t*)}(1-\mu_{0b}^{(t*)}) > \sigma_{ob}^{(t)}$ and compute the acceptance probability

$$\alpha_{\mu_{0b}} = min\left(1, \frac{\pi(\mu_0^{(t*)}|.)q_{\mu_{0b}}(\mu_{0b}^{(t-1)}|a_0,b_0)}{\pi(\mu_0^{(t-1)}|.)q_{\mu_{0b}}(\mu_{0b}^{(t*)}|a_0,b_0)}\right).$$

Draw $U \sim Uniform(0,1)$ and set $\mu_{0b}^{(t)} = \mu_{0b}^{(t*)}$ if $U < \alpha_{\mu_{0b}}$, otherwise set $\mu_{0b}^{(t)} = \mu_{0b}^{(t-1)}$, where

$$\pi(\mu_{0b}|.) \propto \omega_0^{\mu_{0b}\left(\frac{\mu_{0b}(1-\mu_{0b})}{\sigma_{0b}}-1\right)-1}(1-\omega_0)^{\mu_{0b}\left(\frac{1-\mu_{0b}}{\mu_{0b}}\right)\left(\frac{\mu_{0b}(1-\mu_{0b})}{\sigma_{0b}}-1\right)-1}\mathbf{I}(0<\mu_{0b}<1)\mathbf{I}(\mu_{0b}(1-\mu_{0b}) > \sigma_{0b}).$$

Similarly, the parameters (ϕ, ν) are jointly updated. An independent log-normal distribution to the pair (ϕ, ν) was assigned. The conditional posterior distribution follows as

$$\pi(\phi, \nu \mid .) \propto \frac{1}{\mid \mathbf{D} \mid^{1/2}} \exp\left\{-\frac{1}{2}\left(\mathscr{X}^T \tau \mathbf{\Sigma}^{-1}(\boldsymbol{\kappa})\mathscr{X}\right)\right\}\pi(\phi, \nu \mid \mu_\phi, \sigma_\phi^2, \mu_\nu, \sigma_\nu^2).$$

Samples are drawn in a similar way using the Metropolis-Hasting algorithm with gamma proposal distributions. The R code used for the computation is accessible through the link
https://github.com/eosafu/SpatialPriorElicitation/blob/main/simulationPowerBorrow.R.

7.4.2.2 *Under joint spatial commensurate prior*

The spatial commensurate prior distribution provides a different scheme to borrow knowledge from historical point-referenced data. However, its inferential procedure follows a similar technique outlined in the previous section. To make an inference, samples are drawn from the joint posterior distribution using conditional distributions. The fixed hyper-parameters included in this model are $a_\tau, b_\tau, \sigma_\beta^2, \mu_\phi, \sigma_\phi^2, \mu_\nu, \sigma_\nu^2, \mathscr{M}, \upsilon$. At the start of the chain, the initial values of the parameters are drawn from their corresponding prior distributions.

First, we update the field $\mathscr{X}$ through its posterior conditional distribution given as $\mathscr{X} \mid . \sim N(\boldsymbol{\mu}_{\mathscr{X}}, \mathbf{\Sigma}_{\mathscr{X}})$, where $\mathbf{\Sigma}_{\mathscr{X}} = \tau^{-1}(\sum_{j=1}^n \mathbf{V}_j^T\mathbf{V}_j + \mathbf{E}^{-1})^{-1}$ and $\boldsymbol{\mu}_{\mathscr{X}} = (\sum_{j=1}^n \mathbf{V}_j^T\mathbf{V}_j + \mathbf{E}^{-1})^{-1}(\sum_{j=1}^n \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{F})$. In a similar

manner, we adopt the fill-reducing permutation matrix $\mathbf{P}_{\mathscr{X}}$ to ease the burden of factorizing $\mathbf{\Sigma}_{\mathscr{X}}$ several times. To sample from the posterior conditional, find the Cholesky decomposition $\mathbf{L}_{\mathscr{X}}$, such that $\mathbf{L}_{\mathscr{X}}\mathbf{L}_{\mathscr{X}}^T = \mathbf{P}_{\mathscr{X}}\mathbf{\Sigma}_{\mathscr{X}}\mathbf{P}_{\mathscr{X}}^T$. Draw $q^\star$ independent uniform samples in the interval $(0,1)$ to form a vector $\mathbf{u}$. Compute $\mathbf{u}_{\mathscr{X}} = \mathbf{P}_{\mathscr{X}}^T\mathbf{L}_{\mathscr{X}}^{-T}\mathbf{u}$ and $\boldsymbol{\mu}_{\mathscr{X}} = \tau\mathbf{P}^T\mathbf{L}_{\mathscr{X}}\mathbf{L}_{\mathscr{X}}^T\mathbf{P}_{\mathscr{X}}(\sum_{j=1}^n \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{F})$ and set $\mathscr{X} = \boldsymbol{\mu}_{\mathscr{X}} + \mathbf{u}_{\mathscr{X}}$.

Similarly, we update $\boldsymbol{\vartheta}$ from the posterior conditional distribution given as $\boldsymbol{\vartheta} \mid . \sim N(\boldsymbol{\mu}_{\vartheta}, \mathbf{\Sigma}_{\vartheta})$, where $\mathbf{\Sigma}_{\vartheta} = \tau^{-1}(\mathbf{R}^{-1} + \sum_{j=1}^n \mathbf{U}_j\mathbf{U}_j^T)^{-1}$ and $\boldsymbol{\mu}_{\vartheta} = \tau\mathbf{\Sigma}_{\vartheta}\sum_{j=1}^n \mathbf{U}_j(\mathbf{Y}_j - \mathbf{V}_j\mathscr{X})$. Moreover, we update the precision parameter τ from the posterior distribution given as $\tau \mid . \sim Gamma(e_\tau, f_\tau)$, where $e_\tau = \frac{(mn+m_0n_0+q^\star+p_0+2a_\tau)}{2}$ and

$$f_\tau = \frac{1}{2}\Big(\sum_{j=1}^n (\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X})^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X}) + \sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\mathbf{Y}_{0j} + \mathscr{X}^T\mathbf{E}^{-1}\mathscr{X} - 2\mathscr{X}^T\mathbf{F} - (\sum_{j=1}^{n_0}\mathbf{V}_{oj}^T\mathbf{Y}_{0j})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{V}_{oj}^T\mathbf{Y}_{0j}) + \boldsymbol{\vartheta}^T\mathbf{R}^{-1}\boldsymbol{\vartheta} + 2b_\tau. \tag{7.1}$$

Adopting the Metropolis-Hasting algorithm, next, we update the commensurate parameter $\boldsymbol{\lambda}$. The posterior conditional distribution follows as

$$\pi(\boldsymbol{\lambda} \mid .) \propto \frac{\mid \mathbf{D} \mid^{1/2}}{\mid \boldsymbol{\lambda}^{-1} \mid^{1/2}} \exp\Big\{-\frac{\tau}{2}\Big(\mathscr{X}^T\mathbf{E}^{-1}\mathscr{X} - 2\mathscr{X}^T\mathbf{F} - (\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{Y}_{0j})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{Y}_{0j})\Big)\Big\}\pi(\boldsymbol{\lambda}),$$

where $\pi(\boldsymbol{\lambda})$ is a Wishart distribution scale matrix $\mathscr{M}$ and degree of freedom υ, or a joint independent gamma distribution for λ_β, λ_ψ, and λ_θ. A possible proposal distribution is the Wishart distribution with scale matrix $\mathbf{I}$ and $q^\star$ degree of freedom or gamma proposals, as the case may be. The MH steps follow as previously discussed.

Finally, update the pair of parameters (ϕ, ν) from the joint conditional posterior distribution using the Metropolis-Hasting algorithm. Independent log-gamma prior distributions were assigned to (ϕ, ν). The joint conditional posterior distribution follows as

$$\pi(\phi, \nu \mid .) \propto \frac{1}{\mid \mathbf{\Sigma}(\boldsymbol{\kappa}) \mid^{1/2}} \exp\Big\{-\frac{\tau}{2}\mathscr{X}^T\mathbf{\Sigma}(\boldsymbol{\kappa})\mathscr{X}\Big\}\pi(\phi, \nu \mid \mu_\phi, \sigma_\phi^2, \mu_\nu, \sigma_\nu^2).$$

In a similar way, we adopted the fill-reducing permutation matrix for the computation of $det(\mathbf{\Sigma}(\boldsymbol{\kappa}))^{1/2}$ and used gamma distributions as proposals. The outlined steps are iterated over until convergence is reached. The `R` code used in the estimation is accessible through the link
https://github.com/eosafu/SpatialPriorElicitation/blob/main/simulationCommenBorrow.R.

7.4.2.3 Under joint scalable spatial commensurate prior

The scalable spatial commensurate prior distribution allows historical knowledge to be borrowed with lesser computation cost compared with the commensurate spatial prior distribution. The scalable prior allows the historical precision parameter τ_1 to be random. The hyper-parameters included in the model are $a_\tau, b_\tau, a_{\tau_1}, b_{\tau_1}, \mathscr{M}, \upsilon, \mu_\phi, \sigma_\beta^2, \sigma_\phi^2, \mu_\nu, \sigma_\nu^2$. The initial values of the random parameters are drawn from their respective prior distributions. The following steps outline the technique used for inference under the scalable prior distribution.

First, we update the historical field $\mathscr{X}^{'}$ using the posterior conditional distribution given as $\mathscr{X}^{'} \mid \tau_1, \mathscr{D}_0 \sim N\left(\boldsymbol{\mu}^{'}, \mathbf{\Sigma}^{'}\right)$, where $\mathbf{\Sigma}^{'} = \tau_1^{-1}(\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{V}_{0j} + \mathbf{\Sigma}_1^{-1}(\boldsymbol{\kappa}_1))^{-1}$ and $\boldsymbol{\mu}^{'} = \tau_1\mathbf{\Sigma}^{'}(\sum_{j=1}^{n_0}\mathbf{V}_{0j}^T\mathbf{Y}_{0j})$. Depending on the computational demand of factorizing $\mathbf{\Sigma}^{'}$, the fill-reducing permutation matrix could be adopted at this stage to lower the computational burden. Moreover, we update the historical precision τ_1 from the posterior conditional distribution given as $\tau_1 \mid . \sim Gamma(e_{\tau_1}, f_{\tau_1})$, where $e_{\tau_1} = \frac{m_0n_0+q_1^\star+2a_{\tau_1}}{2}$ and $f_{\tau_1} = \frac{1}{2}\Big(\sum_{j=1}^{n_0}(\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'})^T(\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'}) + \mathscr{X}^{'T}\mathbf{\Sigma}_1^{-1}\mathscr{X}^{'} + 2b_{\tau_1}\Big)$.

We find the set $N(s_0)$ that contains the index of the members of the historical field $\mathscr{X}^{'}$ that forms the nearest neighbors to the field $\mathscr{X}$, which is associated with the current data. Denote the subset of the historical field forming the nearest neighbor by $\mathscr{X}^{'}[N(s_0)]$ and the remaining as $\mathscr{X}_1^{'}$. Given $\mathscr{X}_1^{'}$, evaluate matrix $\mathscr{A}$ and $\mathscr{B}$ given in Section 5.3.3. Next, we update the current field using the conditional posterior distribution given as $\mathscr{X} \mid . \sim N(\boldsymbol{\mu}_{\mathscr{X}}, \boldsymbol{\Sigma}_{\mathscr{X}})$, where $\boldsymbol{\Sigma}_{\mathscr{X}} = (\mathbf{M}^{-1} + \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} + \sum_{j=1}^{n} \mathbf{V}_j^T \tau \mathbf{I} \mathbf{V}_j)^{-1}$, and $\boldsymbol{\mu}_{\mathscr{X}} = \boldsymbol{\Sigma}_{\mathscr{X}}(\tau \sum_{j=1}^{n} \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{W})$. A similar technique of using the fill-reducing permutation matrix is adopted to draw samples from the posterior conditional distribution.

Furthermore, the posterior samples of $\boldsymbol{\vartheta}$ are drawn from the posterior conditional distribution given as $\boldsymbol{\vartheta} \mid . \sim N(\boldsymbol{\mu}_{\vartheta}, \boldsymbol{\Sigma}_{\vartheta})$, where $\boldsymbol{\Sigma}_{\vartheta} = \tau^{-1}(\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \mathbf{U}_j^T)^{-1}$ and $\boldsymbol{\mu}_{\vartheta} = \tau\boldsymbol{\Sigma}_{\vartheta}(\sum_{j=1}^{n} \mathbf{U}_j(\mathbf{Y}_j - \mathbf{V}_j\mathscr{X}))$. We update the precision parameter τ using the posterior conditional distribution which follows as $\tau \mid . \sim \text{Gamma}(e_\tau, f_\tau)$, where $e_\tau = \frac{mn+q^\star+p_0+2a_\tau}{2}$ and $f_\tau = \frac{1}{2}\Big(\sum_{j=1}^{n}(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X})^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X}) + \mathscr{X}^T\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1}\mathscr{X} + \boldsymbol{\vartheta}^T\mathbf{R}^{-1}\boldsymbol{\vartheta} + 2b_\tau\Big)$.

Similarly, we update the commensurate parameter $\boldsymbol{\lambda}$ using a Metropolis-Hasting algorithm. Depending on the type of commensurate parameters adopted, the Wishart proposal distribution or independent gamma proposal distributions are assigned to $\boldsymbol{\lambda}$ or λ_β, λ_ψ, and λ_θ respectively. The posterior conditional distribution follows as

$$\pi(\lambda \mid .) \propto \frac{\mid \mathbf{E} \mid^{1/2} \mid \mathbf{K} \mid^{1/2}}{\mid \boldsymbol{\lambda}^{-1} \mid^{1/2}} \exp\left\{ -\frac{1}{2}\Big(\mathscr{X}^T\mathbf{M}^{-1}\mathscr{X} - \mathbf{W}^T\mathscr{X} - \mathscr{X}^T\mathbf{W} - \mathscr{A}^T\mathscr{B}^{-1}(\mathbf{K}\mathscr{B}^{-1} - \mathbf{I})\mathscr{A}\Big)\right\}\pi(\boldsymbol{\lambda}),$$

where $\pi(\boldsymbol{\lambda})$ is the Wishart prior probability distribution with scale matrix $\mathscr{M}$ and $\upsilon > q^\star - 1$ degree of freedom, or a joint independent gamma distribution for λ_β, λ_ψ and λ_θ for the diagonal representation form. The fill-reducing permutation matrix can be adopted to lower the computational burden of evaluating $\mid \mathbf{E} \mid$. The Metropolis-Hasting steps to update $\boldsymbol{\lambda}$ is similar to earlier discussed steps.

Lastly, we update the pair of parameters (ϕ, ν) and (ϕ_1, ν_1) from the joint conditional posterior distribution using the Metropolis-Hasting algorithm. Independent log-gamma prior distributions were assigned to (ϕ, ν) and (ϕ_1, ν_1). The joint conditional posterior distribution follows as

$$\pi(\phi, \nu \mid .) \propto \frac{1}{\mid \boldsymbol{\Sigma}(\boldsymbol{\kappa}) \mid^{1/2}} \exp\Big\{ -\frac{\tau}{2}\mathscr{X}^T\boldsymbol{\Sigma}(\boldsymbol{\kappa})\mathscr{X}\Big\}\pi(\phi, \nu \mid \mu_\phi, \sigma_\phi^2, \mu_\nu, \sigma_\nu^2),$$
$$\pi(\phi_1, \nu_1 \mid .) \propto \frac{1}{\mid \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1) \mid^{1/2}} \exp\Big\{ -\frac{\tau}{2}\mathscr{X}^{'T}\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)\mathscr{X}^{'}\Big\}\pi(\phi_1, \nu_1 \mid \mu_\phi, \sigma_\phi^2, \mu_\nu, \sigma_\nu^2).$$

Note that the dimensions of $\boldsymbol{\Sigma}(\boldsymbol{\kappa})$ and $\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)$ in the scalable commensurate prior are less or equal to the same matrix under the commensurate prior, thus, the factorization is cheaper, and a fill-reducing permutation matrix can be adopted for the computation of the determinants. These MCMC steps, except the ones involved in finding the permutation matrix and nearest neighbors, are repeated several times until convergence is obtained and samples at the burn-in stage are discarded. The `R` code used in the estimation is accessible through the link https://github.com/eosafu/SpatialPriorElicitation/blob/main/simulationScommenBorrow.R.

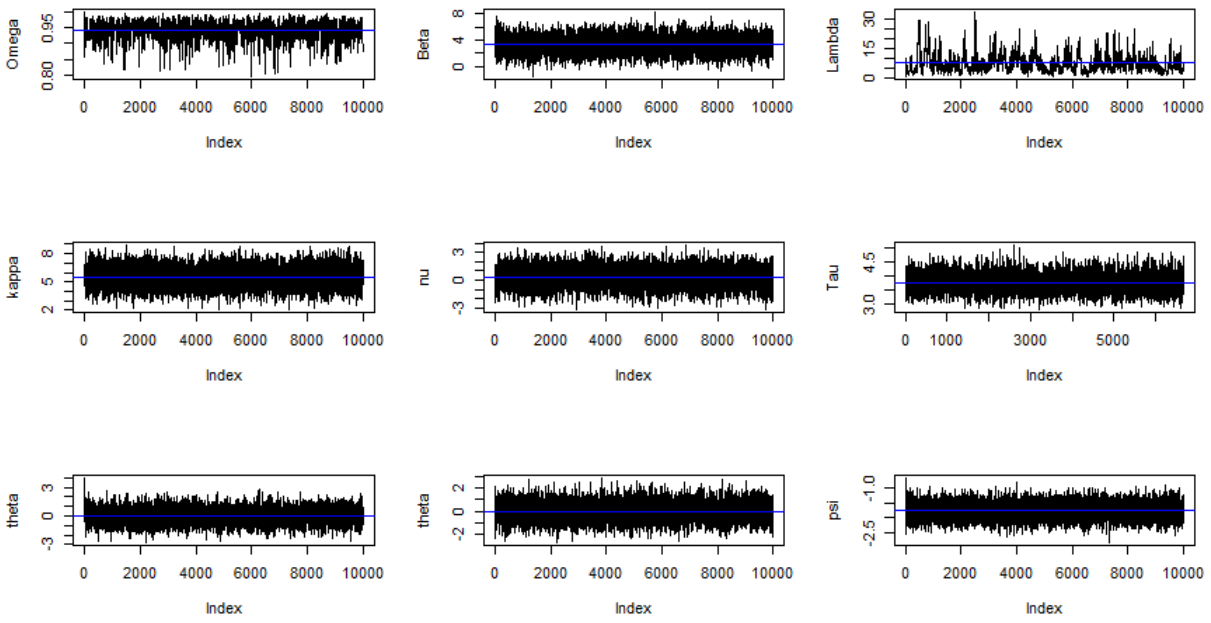


Figure 50 – Trace plot of some selected parameters of Chapter 5.

7.4.3 Derivation of the prior distributions

7.4.3.1 Derivation of the joint spatial power prior by marginalizing over observed confounding variables

Given that

$$
\begin{aligned}
\mathbf{Y}_{0j} \mid \boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\psi} &\sim N(\boldsymbol{\mu}_{0j}, \tau^{-1}\mathbf{I}), \boldsymbol{\mu}_{0j} = \mathbf{X}_{0j}^T\beta + \mathbf{Z}_{0j}^T\psi + \mathbf{A}_{0j}\boldsymbol{\theta}, \\
&\sim N(\mathbf{P}_{0j}^T\mathscr{X}_0 + \mathbf{A}_{0j}\boldsymbol{\theta}, \tau^{-1}\mathbf{I}), \mathbf{P}_{0j}^T = \text{horizontal stack}(\mathbf{X}_{0j}^T, \mathbf{Z}_{0j}^T), \mathscr{X}_0 = (\boldsymbol{\beta}^T, \boldsymbol{\psi}^T)^T, \\
\mathscr{X}_0 &\sim N(\mathbf{0}, \tau^{-1}\mathbf{C}), \\
\boldsymbol{\theta} &\sim N(\mathbf{0}, \tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa})), \\
\tau &\sim Gamma(a,b), \\
\omega_0 &\sim \pi(\omega_0).
\end{aligned}
$$

$$
\pi(\boldsymbol{\theta}, \mathscr{X}_0, \tau, \omega_0 \mid \boldsymbol{\Lambda}, \mathscr{D}_0) \propto
\frac{\prod_{j=1}^{n_0}\left[N(\mathbf{Y}_{0j}; \mu_{0j}, \tau^{-1}\mathbf{I})\right]^{\omega_0} N(\boldsymbol{\theta}; \mathbf{0}, \tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))N(\mathscr{X}_0; \mathbf{0}, \tau^{-1}\mathbf{C})Gam(\tau; a, b)\pi(\omega_0)}{\int \prod_{j=1}^{n_0}\left[N(\mathbf{Y}_{0j}; \mu_{0j}, \tau^{-1}\mathbf{I})\right]^{\omega_0} N(\boldsymbol{\theta}; \mathbf{0}, \tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))N(\mathscr{X}_0; \mathbf{0}, \tau^{-1}\mathbf{C})Gam(\tau; a, b)d\boldsymbol{\theta}\, d\mathscr{X}_0\, d\tau},
$$

Marginalizing over $\mathscr{X}_0$

$$
\pi(\boldsymbol{\theta}, \tau, \omega_0 \mid \boldsymbol{\Lambda}, \mathscr{D}_0) \propto \int \pi(\boldsymbol{\theta}, \mathscr{X}_0, \tau, \omega_0 \mid \boldsymbol{\Lambda}, \mathscr{D}_0) d\mathscr{X}_0.
$$

Considering the denominator and first integrating over $\mathscr{X}_0$,

$$\pi(\omega_0)N(\boldsymbol{\theta};\mathbf{0},\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))Gam(\tau;a,b)\int\prod_{j=1}^{n_0}\Big[N(\mathbf{Y}_{0j};\mu_{0j},\tau^{-1}\mathbf{I})\Big]^{\omega_0}N(\mathscr{X}_0;\mathbf{0},\tau^{-1}\mathbf{C})d\mathscr{X}_0$$

$$=\frac{\pi(\omega_0)N(\boldsymbol{\theta};\mathbf{0},\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))Gam(\tau;a,b)}{det(2\pi\tau^{-1}\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\tau^{-1}\mathbf{C})^{1/2}}\times$$

$$\int\exp\Big\{-\frac{1}{2}\Big(\sum_{j=1}^{n_0}\Big(\mathbf{Y}_{0j}-\mathbf{P}_{0j}^T\mathscr{X}_0-\mathbf{A}_{0j}\boldsymbol{\theta}\Big)^T\tau\mathbf{I}\Big(\mathbf{Y}_{0j}-\mathbf{P}_{0j}^T\mathscr{X}_0-\mathbf{A}_{0j}\boldsymbol{\theta}\Big)+\mathscr{X}_0^T\tau\mathbf{C}^{-1}\mathbf{X}_0\Big)\Big\}d\mathscr{X}_0$$

$$=\frac{N(\boldsymbol{\theta};\mathbf{0},\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))Gam(\tau;a,b)}{det(2\pi\tau^{-1}\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\tau^{-1}\mathbf{C})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\sum_{j=1}^{n_0}\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}+$$

$$n_0\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}\Big)\Big\}\times\int\exp\Big\{-\frac{1}{2}\Big((\mathscr{X}_0-\mathbf{D}\mathbf{B})^T\mathbf{D}^{-1}(\mathscr{X}_0-\mathbf{D}\mathbf{B})-\mathscr{B}^T\mathbf{D}\mathbf{B}\Big)\Big\}d\mathscr{X}_0$$

$$=\frac{N(\boldsymbol{\theta};\mathbf{0},\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))Gam(\tau;a,b)}{det(2\pi\tau^{-1}\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\tau^{-1}\mathbf{C})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\sum_{j=1}^{n_0}\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}+$$

$$n_0\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}\Big)\Big\}\times\exp\Big\{-\frac{1}{2}\Big(-\mathbf{B}^T\mathbf{D}\mathbf{B}\Big)\Big\}det(2\pi\mathbf{D})^{1/2},$$

where $B=\Big(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}\Big)$ and $\mathbf{D}^{-1}=\Big(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{P}_{0j}^T+\tau\mathbf{C}^{-1}\Big)$. It implies that

$$\frac{det(2\pi\mathbf{D})^{1/2}N(\boldsymbol{\theta};\mathbf{0},\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))Gam(\tau;a,b)}{det(2\pi\tau^{-1}\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\tau^{-1}\mathbf{C})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\sum_{j=1}^{n_0}\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}+$$

$$n_0\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}\Big)\Big\}\times\exp\Big\{-\frac{1}{2}\Big(-\mathbf{B}^T\mathbf{D}\mathbf{B}\Big)\Big\}\pi(\omega_0)$$

$$=\frac{det(2\pi\mathbf{D})^{1/2}N(\boldsymbol{\theta};\mathbf{0},\tau^{-1}\mathbf{Q}_\theta(\boldsymbol{\kappa}))Gam(\tau;a,b)}{det(2\pi\tau^{-1}\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\tau^{-1}\mathbf{C})^{1/2}}\times$$

$$\exp\Big\{-\frac{1}{2}\Big(\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\sum_{j=1}^{n_0}\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}+\quad n_0\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\mathbf{B}^T\mathbf{D}\mathbf{B}\Big)\Big\}\pi(\omega_0)$$

$$=\frac{det(2\pi\mathbf{D}_1)^{1/2}\tau^{\frac{n_0m_0\omega_0+(p+r)+q}{2}+a-1-(p+q)/2}b^a}{det(2\pi\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\mathbf{C})^{1/2}det(2\pi\mathbf{Q}_\theta(\boldsymbol{\kappa}))^{1/2}\Gamma(a)}\times$$

$$\exp\Big\{-\frac{1}{2}\Big(\boldsymbol{\theta}^T\tau\mathbf{Q}_\theta(\boldsymbol{\kappa})^{-1}\boldsymbol{\theta}+2\tau b+\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\sum_{j=1}^{n_0}\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{Y}_{0j}+$$

$$n_0\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\Big((\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{Y}_{0j})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{Y}_{0j})-(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{Y}_{0j})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta})-$$

$$(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}A_{0j}\boldsymbol{\theta})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}\mathbf{Y}_{0j})+(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}A_{0j}\boldsymbol{\theta})^T\mathbf{D}(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\tau\mathbf{I}A_{0j}\boldsymbol{\theta})\Big)\Big\},$$

Where $\mathbf{D}_1=\tau^{-1}\mathbf{D}$. Let $K=\frac{det(2\pi\mathbf{D}_1)^{1/2}b^a}{det(2\pi\mathbf{I})^{\frac{n_0\omega_0}{2}}det(2\pi\mathbf{C})^{1/2}det(2\pi\mathbf{Q}_\theta(\boldsymbol{\kappa}))^{1/2}\Gamma(a)}$, and

$M=\frac{1}{2}\Big(\boldsymbol{\theta}^T\mathbf{Q}_\theta(\boldsymbol{\kappa})^{-1}\boldsymbol{\theta}+2b+\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\mathbf{I}\mathbf{Y}_{0j}-\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\omega_0\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-\sum_{j=1}^{n_0}\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\mathbf{I}\mathbf{Y}_{0j}+n_0\boldsymbol{\theta}^T\mathbf{A}_{0j}^T\omega_0\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta}-$
$\Big((\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{Y}_{0j})^T\mathbf{D}_1(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{Y}_{0j})-(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{Y}_{0j})^T\mathbf{D}_1(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta})-$
$(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta})^T\mathbf{D}_1(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{Y}_{0j})+(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta})^T\mathbf{D}_1(\sum_{j=1}^{n_0}\mathbf{P}_{0j}\omega_0\mathbf{I}\mathbf{A}_{0j}\boldsymbol{\theta})\Big)$.
Factorizing M, we have that

$M = \frac{1}{2}\Big((\boldsymbol{\theta} - \mathbf{EF})^T \mathbf{E}^{-1}(\boldsymbol{\theta} - \mathbf{EF}) + k\Big)$, where

$k = -\mathbf{F}^T\mathbf{EF} + 2b + \omega_0 \sum_{j=1}^{n_0} \mathbf{Y}_{0j}^T \mathbf{Y}_{0j} - (\omega_0 \sum_{j=1}^{n_0} \mathbf{P}_{0j} Y_{0j})^T \mathbf{D}_1 (\omega_0 \sum_{j=1}^{n_0} \mathbf{P}_{0j} Y_{0j})$,

$\mathbf{E}^{-1} = (\mathbf{Q}_\theta(\boldsymbol{\kappa})^{-1} + \omega_0 n_0 \mathbf{A}_{0j}^T \mathbf{A}_{0j} - (\omega_0 \sum_{j=1}^{n_0} \mathbf{A}_{0j}^T \mathbf{P}_{0j}^T)\mathbf{D}_1(\omega_0 \sum_{j=1}^{n_0} \mathbf{P}_{0j}\mathbf{A}_{0j})$, and

$\mathbf{F} = \Big(\omega_0 \sum_{j=1}^{n_0} \mathbf{A}_{0j}^T \mathbf{Y}_{0j} - (\omega_0 \sum_{j=1}^{n_0} \mathbf{A}_{0j}^T \mathbf{P}_{0j}^T)\mathbf{D}(\omega_0 \sum_{j=1}^{n_0} \mathbf{P}_{0j} Y_{0j})\Big)$. Thus, the numerator is given as

$$K\tau^{\frac{n_0 m_0 \omega_0 + q + 2a}{2} - 1} \exp\Big\{ - \tau M \Big\}.$$

To obtain the denominator, we integrate out $\boldsymbol{\theta}$ and τ.

Let $a^* = \frac{v+q}{2}$ and $v = n_0 m_0 \omega_0 + 2a$, then

$$\begin{aligned}
\int\int K\tau^{\frac{n_0 m_0 \omega_0 + q + 2a}{2} - 1} \exp\Big\{ - \tau M \Big\} d\tau d\boldsymbol{\theta} &= K \int \frac{\Gamma(a^*)}{M^{a^*}} d\boldsymbol{\theta} \\
&= K\Gamma(a^*) \int \frac{1}{\Big[\frac{1}{2}\Big((\boldsymbol{\theta} - \mathbf{EF})^T \mathbf{E}^{-1}(\boldsymbol{\theta} - \mathbf{EF}) + k\Big)\Big]^{a^*}} d\boldsymbol{\theta} \\
&= K\Gamma(a^*) \int \frac{1}{\Big[\frac{k}{2}\Big(\frac{v(\boldsymbol{\theta} - \mathbf{EF})^T \mathbf{E}^{-1}(\boldsymbol{\theta} - \mathbf{EF})}{vk} + 1\Big)\Big]^{a^*}} d\boldsymbol{\theta} \\
&= K\Gamma(\frac{v+q}{2}) \int \frac{1}{\Big[\frac{k}{2}\Big(\frac{(\boldsymbol{\theta} - \mathbf{EF})^T \frac{v}{k}\mathbf{E}^{-1}(\boldsymbol{\theta} - \mathbf{EF})}{v} + 1\Big)\Big]^{\frac{v+q}{2}}} d\boldsymbol{\theta} \\
&= \frac{K\Gamma(\frac{v+q}{2})}{(k/2)^{\frac{v+q}{2}}} \int \frac{1}{\Big[\Big(\frac{(\boldsymbol{\theta} - \mathbf{EF})^T \frac{v}{k}\mathbf{E}^{-1}(\boldsymbol{\theta} - \mathbf{EF})}{v} + 1\Big)\Big]^{\frac{v+q}{2}}} d\boldsymbol{\theta} \\
&= \frac{K\Gamma(\frac{v+q}{2})\Gamma(v/2) v^{q/2} \pi^{q/2} \mid \frac{k}{v}\mathbf{E} \mid^{1/2}}{(k/2)^{\frac{v+q}{2}} \Gamma(\frac{v+q}{2})}.
\end{aligned}$$

Therefore, the prior distribution is then given as

$$\pi^p(\theta, \tau, \omega_0, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}_0) \propto \pi(\omega_0)\pi(\boldsymbol{\kappa}) \frac{2^{\frac{v+q}{2}} k^{\frac{v}{2}} \tau^{\frac{n_0 m_0 \omega_0 + q + 2a}{2} - 1}}{\Gamma(v/2) \mid \mathbf{E} \mid^{1/2}} \exp\left\{ - \frac{\tau}{2}\Big((\boldsymbol{\theta} - \mathbf{EF})^T \mathbf{E}^{-1}(\boldsymbol{\theta} - \mathbf{EF}) + k\Big) \right\}.$$

7.4.3.2 Derivation of joint spatial commensurate prior

Given that $\mathbf{Y}_{0j} \sim N(V_0\mathscr{X}_0, \mathbf{H})$, $\mathscr{X} \sim N(\mathscr{X}_0, \boldsymbol{\lambda})$, $\mathscr{X} \sim N(\mathbf{0}, \tau^{-1}\Sigma(\boldsymbol{\kappa}))$, $\pi(\tau)$, $\pi(\boldsymbol{\lambda})$, and $\pi(\boldsymbol{\kappa})$,

$$\pi(\mathscr{X}, \mathscr{X}_0, \tau, \boldsymbol{\lambda}, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}_0) = L(\mathscr{X}_0 \mid \mathscr{D}_0)N(\mathscr{X}; \mathscr{X}_0, \boldsymbol{\lambda})N(\mathscr{X}; \mathbf{0}, \tau^{-1}\Sigma(\boldsymbol{\kappa}))\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\tau)$$

$$\begin{aligned}
\pi(\mathscr{X}, \tau, \boldsymbol{\lambda}, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}_0) &= \int \frac{1}{\mid 2\pi \mid \mathbf{H} \mid^{n_0/2}} \frac{1}{\mid 2\pi\boldsymbol{\lambda}^{-1}\mid^{1/2}} \frac{1}{\mid 2\pi\tau^{-1}\Sigma(\boldsymbol{\kappa}) \mid^{1/2}} \exp\Big(-\frac{1}{2}(\mathscr{X}^T\tau\Sigma(\boldsymbol{\kappa})^{-1}\mathscr{X})\Big)\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\tau)\times \\
&\exp\Big(-\frac{1}{2}(\sum_j^{n_0}(\mathbf{Y}_{0j} - V_{0j}\mathscr{X}_0)^T\mathbf{H}^{-1}(\mathbf{Y}_{0j} - V_{0j}\mathscr{X}_0) + (\mathscr{X} - \mathscr{X}_0)^T\boldsymbol{\lambda}(\mathscr{X} - \mathscr{X}_0)\Big)d\mathscr{X}_0 \\
&= \int \frac{1}{\mid 2\pi\mathbf{H} \mid^{n_0/2}} \frac{1}{\mid 2\pi\boldsymbol{\lambda}^{-1}\mid^{1/2}} \frac{1}{\mid 2\pi\tau^{-1}\Sigma(\boldsymbol{\kappa}) \mid^{1/2}} \exp\Big(-\frac{1}{2}(\mathscr{X}^T\tau\Sigma(\boldsymbol{\kappa})^{-1}\mathscr{X})\Big)\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\tau)\times \\
&\exp\Big(-\frac{1}{2}\Big(\sum_j \mathbf{Y}_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0j} + \mathscr{X}^T\boldsymbol{\lambda}\mathscr{X} - \mathbf{B}^\mathbf{T}\mathbf{D}\mathbf{B}\Big)\exp\Big(-\frac{1}{2}(\mathscr{X}_0 - \mathbf{D}\mathbf{B})^T\mathbf{D}^{-1}(\mathscr{X}_0 - \mathbf{D}\mathbf{B})\Big)d\mathscr{X}_0 \\
&= \frac{1}{\mid 2\pi\mathbf{H} \mid^{n_0/2}} \frac{1}{\mid 2\pi\boldsymbol{\lambda}^{-1}\mid^{1/2}} \frac{1}{\mid 2\pi\tau^{-1}\Sigma(\boldsymbol{\kappa}) \mid^{1/2}} \exp\Big(-\frac{1}{2}(\mathscr{X}^T\tau\Sigma(\boldsymbol{\kappa})^{-1}\mathscr{X})\Big)\pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\tau)\times \\
&\exp\Big(-\frac{1}{2}\Big(\sum_j \mathbf{Y}_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0j} + \mathscr{X}^T\boldsymbol{\lambda}\mathscr{X} - \mathbf{B}^\mathbf{T}\mathbf{D}\mathbf{B}\Big) \mid 2\pi\mathbf{D} \mid^{1/2} \\
&= \mid 2\pi\mathbf{D} \mid^{1/2} \frac{1}{\mid 2\pi\mathbf{H} \mid^{n_0/2}} \frac{1}{\mid 2\pi\boldsymbol{\lambda}^{-1}\mid^{1/2}} \frac{1}{\mid 2\pi\tau^{-1}\Sigma(\boldsymbol{\kappa}) \mid^{1/2}} \pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\tau)\times \\
&\exp\Big(-\frac{1}{2}\Big(\sum_j \mathbf{Y}_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0j} + \mathscr{X}^T\boldsymbol{\lambda}\mathscr{X} - \mathbf{B}^\mathbf{T}\mathbf{D}\mathbf{B} + (\mathscr{X}^T\tau\Sigma(\boldsymbol{\kappa})^{-1}\mathscr{X})\Big)\Big),
\end{aligned}$$

where $\mathbf{B} = (\sum_j^{n_0} V_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0j} + \boldsymbol{\lambda}\mathscr{X})$ and $\mathbf{D}^{-1} = (\sum_{j=1}^{n_0} V_{0j}^T\mathbf{H}^{-1}V_0 + \boldsymbol{\lambda})$,

$$\mathbf{B}^\mathbf{T}\mathbf{D}\mathbf{B} = (\sum_{j=1}^{n_0} V_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0i})^T\mathbf{D}(\sum_{j=1}^{n_0} V_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0i}) + (\sum_{j=1}^{n_0} V_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0i})^T\mathbf{D}\boldsymbol{\lambda}\mathscr{X} + (\boldsymbol{\lambda}\mathscr{X})^T\mathbf{D}(\sum_{j=1}^{n_0} V_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0i}) + (\boldsymbol{\lambda}\mathscr{X})^T\mathbf{D}(\boldsymbol{\lambda}\mathscr{X}).$$

$$\begin{aligned}
\pi(\mathscr{X}, \tau, \boldsymbol{\lambda}, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}_0) = \frac{\mid 2\pi\mathbf{D} \mid^{1/2} \pi(\boldsymbol{\lambda})\pi(\boldsymbol{\kappa})\pi(\tau)}{\mid 2\pi\mathbf{H} \mid^{n_0/2} \mid 2\pi\boldsymbol{\lambda}^{-1}\mid^{1/2} \mid 2\pi\tau^{-1}\Sigma(\boldsymbol{\kappa}) \mid^{1/2}} \exp\Big(-\frac{1}{2}\sum_{j=1}^{n_0}\mathbf{Y}_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0j}\Big)\times \\
\exp\Big(-\frac{1}{2}\Big((\mathscr{X} - \mathbf{E}\mathbf{F})^T\mathbf{E}^{-1}(\mathscr{X} - \mathbf{E}\mathbf{F}) - \mathbf{F}^\mathbf{T}\mathbf{E}\mathbf{F} - (\sum_{j=1}^{n_0} V_{0i}^T\mathbf{H}^{-1}\mathbf{Y}_{0j})^T\mathbf{D}(\sum_{j=1}^{n_0} V_{0i}^T\mathbf{H}^{-1}\mathbf{Y}_{0j})\Big)\Big),
\end{aligned}$$

where $\mathbf{E}^{-1} = (\tau\Sigma(\boldsymbol{\kappa})^{-1} + \boldsymbol{\lambda} - \boldsymbol{\lambda}^T\mathbf{D}\boldsymbol{\lambda})$, $\mathbf{F} = \boldsymbol{\lambda}^T\mathbf{D}(\sum_j^{n_0} V_{0j}^T\mathbf{H}^{-1}\mathbf{Y}_{0j})$, and $\boldsymbol{\Lambda}$ is the vector of all hyperparameters.

For K historical data,

$$\begin{aligned}
\pi(\mathscr{X}, \tau, \boldsymbol{\kappa}, \boldsymbol{\lambda}_{01}, .., \boldsymbol{\lambda}_{0K} \mid \boldsymbol{\Lambda}, \{\mathscr{D}_{0i}\}) = \frac{\pi(\boldsymbol{\kappa})\pi(\tau)}{\mid 2\pi\tau^{-1}\Sigma(\boldsymbol{\kappa}) \mid^{1/2}} \prod_{i=1}^{K}\Bigg[\frac{\mid 2\pi\mathbf{D}_i \mid^{1/2} \pi(\boldsymbol{\lambda}_{0i})}{\mid 2\pi H_{0i} \mid^{n_{0i}/2} \mid 2\pi\boldsymbol{\lambda}_{0i}^{-1}\mid^{1/2}}\Bigg] \exp\Big(-\frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{n_0}\mathbf{Y}_{0ij}^T\mathbf{H}_{0i}^{-1}\mathbf{Y}_{0ij}\Big)\times \\
\exp\Big(-\frac{1}{2}\Big((\mathscr{X} - \mathbf{E}\mathbf{F})^T\mathbf{E}^{-1}(\mathscr{X} - \mathbf{E}\mathbf{F}) - \mathbf{F}^\mathbf{T}\mathbf{E}\mathbf{F} - \sum_{i=1}^{K}(\sum_{j=1}^{n_{0i}} V_{0ij}^T\mathbf{H}_{0i}^{-1}\mathbf{Y}_{0ij})^T\mathbf{D}_i(\sum_{j=1}^{n_{0i}} V_{0ij}^T\mathbf{H}_{0i}^{-1}\mathbf{Y}_{0ij})\Big)\Big)
\end{aligned}$$

where $\mathbf{E}^{-1} = (\tau\Sigma(\boldsymbol{\kappa})^{-1} + \sum_{i=1}^{K}(\boldsymbol{\lambda}_{0i} - \boldsymbol{\lambda}_{0i}^T\mathbf{D}_i\boldsymbol{\lambda}_{0i}))$, and $\mathbf{F} = \sum_{i=1}^{K}\boldsymbol{\lambda}_{0i}^T\mathbf{D}_i(\sum_j^{n_{0i}} V_{0ij}^T\mathbf{H}_{0i}^{-1}\mathbf{Y}_{0ij})$, and $\mathbf{D}_i^{-1} = (\sum_{j=1}^{n_{0i}} V_{0ij}^T\mathbf{H}_{0i}^{-1}V_{0ij} + \boldsymbol{\lambda}_{0i})$.

7.4.3.3 Derivation of the joint scalable spatial commensurate prior

Let $\mathbf{Y}_{01} \mid \mathscr{X}^{'} \sim N(\mathbf{V}_0\mathscr{X}^{'}, \tau_1^{-1}\mathbf{I})$ and $\mathscr{X}^{'} \sim N(\mathbf{0}, \tau_1^{-1}\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1))$, $\tau_1 \sim Gamma(a_1, b_1)$. Thus, the joint posterior distribution follows as

$$\begin{aligned}\pi(\mathscr{X}^{'}, \tau_1 \mid \mathscr{D}_0, .) \propto & L(\mathscr{X}^{'}, \tau_1 \mid \mathscr{D}_0) N(\mathscr{X}^{'}; \mathbf{0}, \tau_1^{-1}\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)) Gamma(\tau_1; a_1, b_1) \\ = & \frac{1}{det(2\pi\tau_1^{-1}\mathbf{I})^{1/2}} \frac{\tau_1^{a_1-1}}{det(2\pi\tau^{-1}\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1))^{1/2}} \exp\Big\{ -\frac{1}{2}\Big(\sum_{j=1}^{n_0} (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'})^T \tau_1 \mathbf{I} (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'}) + \\ & \mathscr{X}^{'T} \tau_1 \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1} \mathscr{X}^{'} + 2\tau_1 b_1 \Big) \Big\} \\ \propto & \tau_1^{\frac{m_0 n_0 + q_1^{\star}}{2} + a_1 - 1} \exp\Big\{ -\frac{1}{2}\Big(\sum_{j=1}^{n_0} (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'})^T \tau_1 \mathbf{I} (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'}) + \mathscr{X}^{'T} \tau_1 \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1} \mathscr{X}^{'} + 2\tau_1 b_1 \Big) \Big\}.\end{aligned}$$

Thus, the conditional posterior distribution of the updated $\mathscr{X}^{'}$ follows as

$$\pi(\mathscr{X}^{'} \mid \tau_1, \mathscr{D}_0, .) \propto \exp\Big\{ -\frac{1}{2}\mathscr{X}^{'T} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \tau_1 \mathbf{I} \mathbf{V}_{0j} + \tau_1 \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1}) \mathscr{X}^{'} - (\sum_{j=1}^{n_0} \mathbf{Y}_{0j}^T \tau_1 \mathbf{I} \mathbf{V}_{0j}) \mathscr{X}^{'} - \mathscr{X}^{'T} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \tau_1 \mathbf{I} \mathbf{Y}_{0j}) \Big\}$$

Therefore, $\mathscr{X}^{'} = (\mathscr{X}^{'}[N(s_0)], \mathscr{X}_1^{'})^T \mid \tau_1, \mathscr{D}_0 \sim N\Big(\boldsymbol{\mu}^{'}, \boldsymbol{\Sigma}^{'}\Big)$ where

$$\begin{aligned}\boldsymbol{\mu}^{'} &= \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix} = (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{V}_{0j} + \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1})^{-1} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{Y}_{0j}) \text{ and} \\ \boldsymbol{\Sigma}^{'} &= \begin{pmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12} \\ \mathbf{Q}_{21} & \mathbf{Q}_{22} \end{pmatrix} = \tau_1^{-1} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{V}_{0j} + \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1})^{-1}.\end{aligned}$$

Thus,

$$\mathscr{X}^{'}[N(s_0)] \mid \mathscr{X}_1^{'}, \tau_1 \sim N(\mathscr{A}, \mathscr{B}), \text{ where } \mathscr{A} = \boldsymbol{\mu}_1 + \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}(\mathscr{X}_1^{'} - \boldsymbol{\mu}_2) \text{ and } \mathscr{B} = \mathbf{Q}_{11} - \mathbf{Q}_{12}\mathbf{Q}_{22}^{-1}\mathbf{Q}_{21}.$$

$$\pi(\tau_1 \mid \mathscr{X}^{'}) \propto \tau_1^{\frac{m_0 n_0 + q_1^{\star}}{2} + a_1 - 1} \exp\Big\{ -\frac{\tau_1}{2}\Big(\sum_{j=1}^{n_0} (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'})^T (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'}) + \mathscr{X}^{'T} \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1} \mathscr{X}^{'} + 2b_1 \Big) \Big\}.$$

Therefore, $\tau_1 \mid . \sim Gamma(a_{\tau_1}, b_{\tau_1})$, where $a_{\tau_1} = \frac{m_0 n_0 + q_1^{\star} + 2a_1}{2}$ and $b_{\tau_1} = \frac{1}{2}\Big(\sum_{j=1}^{n_0} (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'})^T (\mathbf{Y}_{0j} - \mathbf{V}_{0j}\mathscr{X}^{'}) + \mathscr{X}^{'T} \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{-1} \mathscr{X}^{'} + 2b_1 \Big)$.
Thus, to obtain the posterior predictive distribution at locations in $\mathscr{G}$,

$$\begin{pmatrix} \mathscr{X}_0 \\ \mathscr{X}^{'} \end{pmatrix} = \begin{pmatrix} \mathscr{X}_0 \\ \mathscr{X}^{'}[N(s_0)] \\ \mathscr{X}_1^{'} \end{pmatrix} \sim N(\mathbf{0}, \tau_1^{-1}\boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{\star}), \text{ where } \boldsymbol{\Sigma}_1(\boldsymbol{\kappa}_1)^{\star} = \begin{pmatrix} \Sigma_{00} & \Sigma_{01} \\ \Sigma_{10} & \boldsymbol{\Sigma}^{'}. \end{pmatrix}$$

Let $\mathscr{X}^{'}[N(s_0)]$ be the spatial field in $\mathscr{G}_0$ having connection to the spatial field $\mathscr{X}_0$ in $\mathscr{G}$. So that, $\mathscr{X}_0 \mid \mathscr{X}^{'} = \mathscr{X}_0 \mid \mathscr{X}^{'}[N(s_0)] \sim N\Big(\mathbf{c}^T \mathscr{X}^{'}[N(s_0)], \mathbf{D}\Big)$, where $\mathbf{c}^T = \Sigma_{01}\mathbf{Q}_{11}^{'-1}$ and $\mathbf{D} = \Sigma_{00} - \Sigma_{01}\mathbf{Q}_{11}^{'-1}\Sigma_{10}$.

Given this background, the proposed commensurate prior distribution follows as

$$\pi(\mathscr{X}, \mathscr{X}_0, \mathscr{X}^{'}[N(s_0)], \boldsymbol{\lambda} \mid \mathscr{X}_1^{'}) = \pi(\mathscr{X} \mid \mathscr{X}_0, \boldsymbol{\lambda}) \pi(\mathscr{X}_0 \mid X^{'}[N(s_0)]) \pi(\mathscr{X}^{'} N(s_0) \mid \mathscr{X}_1^{'}, \mathscr{D}_0) \pi(\mathscr{X}) \pi(\boldsymbol{\lambda}).$$

Therefore,

$$\begin{aligned}
\pi(\mathscr{X},\boldsymbol{\lambda}\mid \mathscr{X}_1^{'}) &= \int\int \Big[\pi(\mathscr{X}\mid \mathscr{X}_0,\boldsymbol{\lambda})\pi(\mathscr{X}_0\mid \mathscr{X}^{'}[N(s_0)])\pi(\mathscr{X}^{'}[N(s_0)]\mid \mathscr{X}_1^{'},\mathscr{D}_0)\Big] d\mathscr{X}_0, d\mathscr{X}[N(s_0)]\pi(\mathscr{X})\pi(\boldsymbol{\lambda}) \\
&= \int\int \frac{1}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}} \exp\Big\{-\frac{1}{2}\Big((\mathscr{X}-\mathscr{X}_0)^T\boldsymbol{\lambda}(\mathscr{X}-\mathscr{X}_0)\Big)\Big\}\times \\
&\quad \frac{1}{det(2\pi\mathbf{D})^{1/2}} \exp\Big\{-\frac{1}{2}\Big((\mathscr{X}_0-\mathbf{c}^T\mathscr{X}^{'}[N(s_0)])^T\mathbf{D}^{-1}(\mathscr{X}_0-\mathbf{c}^T\mathscr{X}^{'}[N(s_0)])\Big)\Big\}\times \\
&\quad \frac{1}{det(2\pi\mathbf{B})^{1/2}} \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^{'}[N(s_0)]-\mathscr{A})^T\mathscr{B}^{-1}(\mathscr{X}^{'}[N(s_0)]-\mathscr{A})\Big)\Big\}\Big] d\mathscr{X}_0, d\mathscr{X}[N(s_0)]\times \\
&\quad \pi(\mathscr{X})\pi(\boldsymbol{\lambda})
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}} \int\int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}-\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}_0-\mathscr{X}_0^T\boldsymbol{\lambda}\mathscr{X}+\mathscr{X}_0^T\boldsymbol{\lambda}\mathscr{X}_0+ \\
&\quad \mathscr{X}_0^T\mathbf{D}^{-1}\mathscr{X}_0-\mathscr{X}_0^T\mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1-(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1}\mathscr{X}_0+(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1}(\mathbf{c}^T\mathscr{X}_1)+(\mathscr{X}_1-\mathscr{A})^T\mathscr{B}^{-1}(\mathscr{X}_1-\mathscr{A})\Big)\Big\} \\
&\quad d\mathscr{X}_0, d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda}) \\
&= \frac{1}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}} \int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}+(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1}(\mathbf{c}^T\mathscr{X}_1)+ \\
&\quad (\mathscr{X}_1-\mathscr{A})^T\mathscr{B}^{-1}(\mathscr{X}_1-\mathscr{A})\Big)\Big\}\int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}_0^T(\boldsymbol{\lambda}+\mathbf{D}^{-1})\mathscr{X}_0-\mathscr{X}_0^T(\boldsymbol{\lambda}\mathscr{X}+\mathbf{D}^{-1}\mathbf{C}^T\mathscr{X}_1)- \\
&\quad (\mathscr{X}^T\boldsymbol{\lambda}+(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1})\mathscr{X}_0\Big\} d\mathscr{X}_0 d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda}) \\
&= \frac{1}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}} \int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}+(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1}(\mathbf{c}^T\mathscr{X}_1)+ \\
&\quad (\mathscr{X}_1-\mathscr{A})^T\mathscr{B}^{-1}(\mathscr{X}_1-\mathscr{A})\Big)\Big\}\int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}_0^T\mathbf{E}^{-1}\mathscr{X}_0-\mathscr{X}_0^T\mathbf{F}-\mathbf{F}^T\mathscr{X}_0\Big\} d\mathscr{X}_0 d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda}) \\
&= \frac{1}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}} \int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}+(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1}(\mathbf{c}^T\mathscr{X}_1)+ \\
&\quad (\mathscr{X}_1-\mathbf{A})^T\mathscr{B}^{-1}(\mathscr{X}_1-\mathscr{A})\Big)\Big\}\int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}_0-\mathbf{EF})^T\mathbf{E}^{-1}(\mathscr{X}_0-\mathbf{EF})-\mathbf{F}^T\mathbf{EF}\} d\mathscr{X}_0 d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda}) \\
&= \frac{det(2\pi\mathbf{E})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}} \int \exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}+(\mathbf{c}^T\mathscr{X}_1)^T\mathbf{D}^{-1}(\mathbf{c}^T\mathscr{X}_1)+ \\
&\quad (\mathscr{X}_1-\mathscr{A})^T\mathscr{B}^{-1}(\mathscr{X}_1-\mathscr{A})-\mathbf{F}^T\mathbf{EF}\Big)\Big\} d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda}),
\end{aligned}$$

where $\mathbf{E}^{-1} = (\boldsymbol{\lambda} + \mathbf{D}^{-1})$ and $\mathbf{F} = (\boldsymbol{\lambda}\mathscr{X} + \mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1)$. Now expanding $\mathbf{F}^T\mathbf{E}\mathbf{F}$, we have

$$\begin{aligned}
&\frac{det(2\pi\mathbf{E})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\int\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}+\mathscr{X}_1^T\mathbf{c}\mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1+\mathscr{X}_1^T\mathscr{B}^{-1}\mathscr{X}_1-\mathscr{X}_1^T\mathscr{B}^{-1}\mathscr{A}-\\
&\mathscr{A}^T\mathscr{B}^{-1}\mathscr{X}_1+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}-\Big[(\boldsymbol{\lambda}\mathscr{X})^T\mathbf{E}\boldsymbol{\lambda}\mathscr{X}+(\boldsymbol{\lambda}\mathscr{X})^T\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1+(\mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1)^T\mathbf{E}\boldsymbol{\lambda}\mathscr{X}+\\
&(\mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1)^T\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathscr{X}_1)\Big]\Big)\Big\}d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda})\\
=&\frac{det(2\pi\mathbf{E})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\int\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\boldsymbol{\lambda}\mathscr{X}+\mathscr{X}_1^T(\mathbf{c}\mathbf{D}^{-1}\mathbf{c}^T+\mathscr{B}^{-1}-\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T)\mathscr{X}_1-\\
&(\mathscr{A}^T\mathscr{B}^{-1}+\mathscr{X}^T\boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T)\mathscr{X}_1-\mathscr{X}_1^T(\mathscr{B}^{-1}\mathscr{A}+\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X})+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}-\mathscr{X}^T\boldsymbol{\lambda}\mathbf{E}\boldsymbol{\lambda}\mathscr{X}\Big)\Big\}d\mathscr{X}_1\\
&\pi(\mathscr{X})\pi(\boldsymbol{\lambda})\\
=&\frac{det(2\pi\mathbf{E})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\int\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}^T\mathbf{E}\boldsymbol{\lambda})\mathscr{X}+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}+\mathscr{X}_1^T\mathbf{K}^{-1}\mathscr{X}_1-\\
&\mathbf{L}^T\mathscr{X}_1-\mathscr{X}_1\mathbf{L}\Big)\Big\}d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda}),
\end{aligned}$$

where $\mathbf{K}^{-1} = (\mathbf{c}\mathbf{D}^{-1}\mathbf{c}^T + \mathscr{B}^{-1} - \mathbf{c}\mathbf{D}^{-1}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T)$ and $\mathbf{L} = (\mathscr{B}^{-1}\mathscr{A} + \mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X})$.

Further,

$$\begin{aligned}
&\frac{det(2\pi\mathbf{E})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}^T\mathbf{E}\boldsymbol{\lambda})\mathscr{X}+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}-\mathbf{L}^T\mathbf{K}\mathbf{L}\Big)\Big\}\times\\
&\int\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}_1-\mathbf{K}\mathbf{L})^T\mathbf{K}^{-1}(\mathscr{X}_1-\mathbf{K}\mathbf{L})\Big)\Big\}d\mathscr{X}_1\pi(\mathscr{X})\pi(\boldsymbol{\lambda})\\
=&\frac{det(2\pi\mathbf{E})^{1/2}det(2\pi\mathbf{K})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}^T\mathbf{E}\boldsymbol{\lambda})\mathscr{X}+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}-\mathbf{L}^T\mathbf{K}\mathbf{L}\Big)\Big\}\times\\
&\pi(\mathscr{X})\pi(\boldsymbol{\lambda}).
\end{aligned}$$

Expanding,

$$\begin{aligned}
=&\frac{det(2\pi\mathbf{E})^{1/2}det(2\pi\mathbf{K})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}^T\mathbf{E}\boldsymbol{\lambda})\mathscr{X}+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}-\\
&\Big[(\mathscr{B}^{-1}\mathscr{A})^T\mathbf{K}(\mathscr{B}^{-1}\mathscr{A})+(\mathscr{B}^{-1}\mathscr{A})^T\mathbf{K}\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X}+(\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X})^T\mathbf{K}\mathscr{B}^{-1}\mathscr{A}+\\
&(\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X})^T\mathbf{K}(\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X})\Big]\Big)\Big\}\pi(\mathscr{X})\pi(\boldsymbol{\lambda})\\
=&\frac{det(2\pi\mathbf{E})^{1/2}det(2\pi\mathbf{K})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T(\boldsymbol{\lambda}-\boldsymbol{\lambda}\mathbf{E}\boldsymbol{\lambda}-\boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathbf{K}\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda})\mathscr{X}-\\
&(\mathscr{B}^{-1}\mathscr{A})^T\mathbf{K}\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda}\mathscr{X}-\mathscr{X}^T\boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathbf{K}\mathscr{B}^{-1}\mathscr{A}-\mathscr{A}^T\mathscr{B}^{-1}\mathbf{K}\mathscr{B}^{-1}\mathscr{A}+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}\Big)\Big\}\pi(\mathscr{X})\pi(\boldsymbol{\lambda}).
\end{aligned}$$

Thus,

$$\begin{aligned}
\pi(\mathscr{X},\boldsymbol{\lambda}\mid\mathscr{X}_1',\boldsymbol{\Lambda},\mathscr{D}_0) =&\frac{det(2\pi\mathbf{E})^{1/2}det(2\pi\mathbf{K})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}det(2\pi\mathbf{D})^{1/2}det(2\pi\mathscr{B})^{1/2}}\\
&\exp\Big\{-\frac{1}{2}\Big(\mathscr{X}^T\mathbf{M}^{-1}\mathscr{X}-\mathbf{W}^T\mathscr{X}-\mathscr{X}^T\mathbf{W}-\mathscr{A}^T\mathscr{B}^{-1}\mathbf{K}\mathscr{B}^{-1}\mathscr{A}+\mathscr{A}^T\mathscr{B}^{-1}\mathscr{A}\Big)\Big\}\pi(\mathscr{X})\pi(\boldsymbol{\lambda}),
\end{aligned}$$

where $\mathbf{M}^{-1} = (\boldsymbol{\lambda} - \boldsymbol{\lambda}\mathbf{E}\boldsymbol{\lambda} - \boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathbf{K}\mathbf{c}\mathbf{D}^{-1}\mathbf{E}\boldsymbol{\lambda})$, $\mathbf{W} = (\boldsymbol{\lambda}\mathbf{E}\mathbf{D}^{-1}\mathbf{c}^T\mathbf{K}\mathscr{B}^{-1}\mathscr{A})$ and $\boldsymbol{\Lambda}$ is the vector of hyperparameters.

7.4.4 Derivation of the conditional posterior distribution

7.4.4.1 Under joint spatial power prior

Given the hierarchical model

$$\begin{aligned}\mathbf{Y}_j \mid \mathscr{X}, \boldsymbol{\vartheta} &\sim N(\boldsymbol{\eta}_j, \tau^{-1}\mathbf{I}),\\ \boldsymbol{\eta}_j &= \mathbf{U}_j^T\boldsymbol{\vartheta} + \mathbf{V}\mathscr{X},\\ (\mathscr{X}, \tau) &\sim \pi^p,\\ \boldsymbol{\vartheta} &\sim N(\mathbf{0}, \tau^{-1}\mathbf{R}),\ j = 1,2,...,n.\end{aligned}$$

Hence, the joint posterior distribution follows as

$$\begin{aligned}\pi(\mathscr{X}, \tau, \omega_0, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}, \mathscr{D}_0) \propto{}& \frac{1}{det(2\pi\tau^{-1}\mathbf{I})^{n/2}} \frac{2^{(v+q)/2} k^{v/2} \tau^{(m_0 n_0 \omega_0 + q^*)/2 + a - 1}}{\Gamma(v/2) det(\mathbf{D})^{1/2}} \frac{1}{det(2\pi\tau^{-1}\mathbf{R})^{1/2}} \times\\ &\exp\Big\{ -\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X})^T \tau\mathbf{I}(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta} - \mathbf{V}_j\mathscr{X}) +\\ &\tau(\mathscr{X} - \mathbf{DB})^T\mathbf{D}^{-1}(\mathscr{X} - \mathbf{DB}) + \tau k + \boldsymbol{\vartheta}^T\tau^{-1}\mathbf{R}\boldsymbol{\vartheta}\Big)\Big\},\end{aligned}$$

where $D^{-1} = (\omega_0 \sum_j^{n_0} \mathbf{V_{0j}^T V_{0j}} + \boldsymbol{\Sigma}^{-1}(\boldsymbol{\kappa}))$, $\mathbf{B} = \omega_0 \sum_{j=1}^{n_0} \mathbf{V}_0^T \mathbf{Y}_{0j}$, $k = \omega_0(\sum_{j=1}^{n_0} \mathbf{Y_{0j}^T Y_{0j}}) - \mathbf{B^T D B} + 2b$, $v = 2a + m_0 n_0 \omega_0$, $q^\star = p + r + q$, and $\boldsymbol{\kappa} = \{\sigma_\beta^2, \sigma_\psi^2, \tau_1, \phi, \nu\}$. Consider the exponent, then its expanded form follows as

$$\begin{aligned}&\sum_{j=1}^{n} \mathbf{Y}_j^T\tau\mathbf{I}\mathbf{Y_j} - \sum_{j=1}^{n} Y_j^T\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} - \sum_{j=1}^{n} \mathbf{Y}_j^T\tau\mathbf{I}\mathbf{V_j}\mathscr{X} - \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{Y}_j + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{V}_j\mathscr{X} -\\ &\sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{Y}_j + \sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} + \sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{V}_j\mathscr{X} + \mathscr{X}^T\tau\mathbf{I}\mathbf{D}^{-1}\mathscr{X} - \mathscr{X}^T\tau\mathbf{I}\mathbf{B} - \mathbf{B}^T\tau\mathbf{I}\mathscr{X} + \mathbf{B}^T\tau\mathbf{I}\mathbf{D}\mathbf{B} +\\ &\tau k + \boldsymbol{\vartheta}^T\tau\mathbf{R}^{-1}\boldsymbol{\vartheta}.\end{aligned}$$

The posterior conditional for $\mathscr{X}$ follows as

$$\begin{aligned}\pi(\mathscr{X} \mid .) \propto{}& \exp\Big\{ -\frac{1}{2}\Big(-\sum_{j=1}^{n} \mathbf{Y}_j^T\tau\mathbf{I}\mathbf{V}_j\mathscr{X} + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{V}_j\mathscr{X} - \sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{Y}_j + \sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} +\\ &\sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{V}_j\mathscr{X} + \tau\mathscr{X}^T\mathbf{D}^{-1}\mathscr{X} - \tau\mathscr{X}^T\mathbf{B} - \tau\mathbf{B}^T\mathscr{X}\Big)\Big\}\\ ={}& \exp\Big\{ -\frac{1}{2}\Big(\mathscr{X}^T(\sum_{j=1}^{n} \mathbf{V}_j^T\tau\mathbf{I}\mathbf{V}_j + \tau D^{-1})\mathscr{X} - (\sum_{j=1}^{n} \mathbf{Y}_j^T\tau\mathbf{I}\mathbf{V}_j - \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{V}_j + \tau\mathbf{B}^T)\mathscr{X} -\\ &\mathscr{X}^T(\sum_{j=1}^{n} \mathbf{V}_j^T\tau\mathbf{I}\mathbf{Y}_j - \sum_{j=1}^{n} \mathbf{V}_j^T\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} + \tau\mathbf{B})\Big)\Big\}.\end{aligned}$$

Therefore, $\mathscr{X} \mid . \sim N(\boldsymbol{\mu}_{\mathscr{X}}, \boldsymbol{\Sigma}_{\mathscr{X}})$, where $\boldsymbol{\Sigma}_{\mathscr{X}} = \tau^{-1}(\sum_{j=1}^{n} \mathbf{V}_j^T\mathbf{V}_j + \mathbf{D}^{-1})^{-1}$ and $\boldsymbol{\mu}_{\mathscr{X}} = \tau\boldsymbol{\Sigma}_{\mathscr{X}}(\sum_{j=1}^{n} \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{B})$.

Similarly,

$$\begin{aligned}\pi(\boldsymbol{\vartheta} \mid .) \propto{}& \exp\Big\{ -\frac{1}{2}\Big(-\sum_{j=1}^{n} \mathbf{Y}_j^T\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} - \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{Y}_j + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T\mathbf{U}_j\tau\mathbf{I}\mathbf{V}_j\mathscr{X} +\\ &\sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{U}_j^T\boldsymbol{\vartheta} + \boldsymbol{\vartheta}^T\tau\mathbf{R}^{-1}\boldsymbol{\vartheta}\Big)\Big\}\\ ={}& \exp\Big\{ -\frac{1}{2}\Big(\boldsymbol{\vartheta}^T(\tau\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j\tau\mathbf{I}\mathbf{U}_j^T)\boldsymbol{\vartheta} - (\sum_{j=1}^{n} \mathbf{Y}_j^T\tau\mathbf{I}\mathbf{U}_j^T - \sum_{j=1}^{n} \mathscr{X}^T\mathbf{V}_j^T\tau\mathbf{I}\mathbf{U}_j^T)\boldsymbol{\vartheta} -\\ &\boldsymbol{\vartheta}^T(\sum_{j=1}^{n} \mathbf{U}_j\tau\mathbf{I}\mathbf{Y}_j - \sum_{j=1}^{n} \mathbf{U}_j\tau\mathbf{I}\mathbf{V}_j\mathscr{X})\Big)\Big\}.\end{aligned}$$

Therefore, $\boldsymbol{\vartheta} \mid . \sim N(\boldsymbol{\mu}_{\vartheta}, \boldsymbol{\Sigma}_{\vartheta})$, where $\boldsymbol{\Sigma}_{\vartheta} = \tau^{-1}(\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \mathbf{U}_j^T)^{-1}$ and $\boldsymbol{\mu}_{\vartheta} = \tau \boldsymbol{\Sigma}_{\vartheta} \sum_{j=1}^{n} \mathbf{U}_j (\mathbf{Y}_j - \mathbf{V}_j \mathscr{X})$.

Further,

$$\pi(\tau \mid .) \propto \tau^{\frac{(mn+m_0 n_0 \omega_0 + q^{\star} + p_0 + 2a)}{2} - 1} \exp\Big\{ -\tau\Big(\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + (\mathscr{X} - \mathbf{DB})^T \mathbf{D}^{-1} (\mathscr{X} - \mathbf{DB}) + k + \boldsymbol{\vartheta}^T \mathbf{R}^{-1} \boldsymbol{\vartheta} \Big) \Big\}.$$

Thus, $\tau \mid . \sim Gamma(a_{\tau}, b_{\tau})$, where $a_{\tau} = \frac{(mn + m_0 n_0 \omega_0 + q^{\star} + p_0 + 2a)}{2}$ and $b_{\tau} = \frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + (\mathscr{X} - \mathbf{DB})^T \mathbf{D}^{-1} (\mathscr{X} - \mathbf{DB}) + k + \boldsymbol{\vartheta}^T \mathbf{R}^{-1} \boldsymbol{\vartheta} \Big)$.

The posterior distribution for ω_0 follows as

$$\pi(\omega_0 \mid .) \propto \frac{(2k)^{\frac{\nu}{2}} \tau^{(m_0 n_0 \omega_0)/2}}{\Gamma(\frac{\nu}{2}) \mid \mathbf{D} \mid^{1/2}} \exp\Big\{ -\frac{1}{2}\Big(\tau (\mathscr{X} - \mathbf{DB})^T \mathbf{D}^{-1} (\mathscr{X} - \mathbf{DB}) + \tau k \Big) \Big\} \pi(\omega_0)$$

and for $\boldsymbol{\kappa}$, the posterior distribution follows as

$$\pi(\boldsymbol{\kappa} \mid .) \propto \frac{1}{\mid \mathbf{D} \mid^{1/2}} \exp\Big\{ -\frac{1}{2}\Big(\mathscr{X}^T \tau \Sigma^{-1}(\boldsymbol{\kappa}) \mathscr{X} \Big) \Big\} \pi(\boldsymbol{\kappa}).$$

7.4.4.2 Under joint spatial commensurate prior

Given the hierarchical model

$$\begin{aligned} \mathbf{Y}_j \mid \mathscr{X}, \boldsymbol{\vartheta} &\sim N(\boldsymbol{\eta}_j, \tau^{-1}\mathbf{I}), \\ \boldsymbol{\eta}_j &= \mathbf{U}_j^T \boldsymbol{\vartheta} + \mathbf{V} \mathscr{X}, \\ (\mathscr{X}, \tau) &\sim \pi^c, \\ \boldsymbol{\vartheta} &\sim N(\mathbf{0}, \tau^{-1}\mathbf{R}),\ j = 1, 2, ..., n. \end{aligned}$$

Hence, the joint posterior distribution follows as

$$\begin{aligned} \pi(\mathscr{X}, \tau, \boldsymbol{\lambda}, \boldsymbol{\kappa} \mid \boldsymbol{\Lambda}, \mathscr{D}, \mathscr{D}_0) &\propto \frac{det(2\pi\tau^{-1}\mathbf{D})^{1/2} \pi(\tau) \pi(\boldsymbol{\vartheta}) \pi(\boldsymbol{\kappa}) \pi(\boldsymbol{\lambda})}{det(2\pi\tau^{-1}\mathbf{I})^{n/2} det(2\pi\mathbf{H})^{n_0/2} det(2\pi\tau^{-1}\boldsymbol{\lambda}^{-1})^{1/2} det(2\pi\tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa}))^{1/2}} \times \\ &\exp\Big\{ -\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T \tau \mathbf{I} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + \tau\Big((\mathscr{X} - \mathbf{EF})^T \mathbf{E}^{-1} (\mathscr{X} - \mathbf{EF}) - \mathbf{F}^T \mathbf{EF} - \\ &(\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{Y}_{0j})^T \mathbf{D} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{Y}_{0j}) \Big) \Big) \Big\} \\ =& \frac{\tau^{\frac{mn + m_0 n_0 + q^{\star} + p_0}{2}} det(2\pi\mathbf{D})^{1/2} \pi(\tau) \pi(\boldsymbol{\vartheta}) \pi(\boldsymbol{\kappa}) \pi(\boldsymbol{\lambda})}{det(2\pi\mathbf{I})^{n/2} det(2\pi I)^{n_0/2} det(2\pi\boldsymbol{\lambda}^{-1})^{1/2} det(2\pi\boldsymbol{\Sigma}(\boldsymbol{\kappa}))^{1/2}} \exp\Big\{ -\frac{\tau}{2}\Big(\sum_{j=1}^{n} \Big(Y_j^T Y_j - Y_j^T \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{Y}_j^T \mathbf{V}_j \mathscr{X} - \\ &(\mathbf{U}_j^T \boldsymbol{\vartheta})^T \mathbf{Y}_j + (\mathbf{U}_j^T \boldsymbol{\vartheta})^T (\mathbf{U}_j^T \boldsymbol{\vartheta}) + (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \mathbf{V}_j \mathscr{X} - (\mathbf{V}_j \mathscr{X})^T \mathbf{Y}_j + (\mathbf{V}_j \mathscr{X})^T (\mathbf{U}_j^T \boldsymbol{\vartheta}) + (\mathbf{V}_j \mathscr{X})^T (\mathbf{V}_j \mathscr{X}) \Big) + \\ &\sum_{j=1}^{n} \mathbf{Y}_{0j}^T \mathbf{Y}_0^T + \mathscr{X}^T \mathbf{E}^{-1} \mathscr{X} - \mathscr{X}^T \mathbf{F} - \mathbf{F}^T \mathscr{X} - (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{Y}_{0j})^T \mathbf{D} (\sum_{j=1}^{n_0} \mathbf{V}_{0j}^T \mathbf{Y}_{0j}) + \boldsymbol{\vartheta}^T \mathbf{R}^{-1} \boldsymbol{\vartheta} \Big) \Big\}. \end{aligned}$$

The posterior conditional for $\mathscr{X}$ follows as

$$\begin{aligned} \pi(\mathscr{X} \mid .) \propto& \Big\{ -\frac{\tau}{2}\Big(-\sum_{j=1}^{n} \mathbf{Y}_j^T \mathbf{V}_j \mathscr{X} + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T \mathbf{U}_j \mathbf{V}_j \mathscr{X} - \sum_{j=1}^{n} \mathscr{X}^T \mathbf{V}_j^T \mathbf{Y}_j + \sum_{j=1}^{n} \mathscr{X}^T \mathbf{V}_j^T \mathbf{U}_j^T \boldsymbol{\vartheta} + \sum_{j=1}^{n} \mathscr{X}^T \mathbf{V}_j^T \mathbf{V}_j \mathscr{X} + \\ &\mathscr{X}^T \mathbf{E}^{-1} \mathscr{X} - \mathscr{X}^T \mathbf{F} - \mathbf{F}^T \mathscr{X} \Big) \\ &= \exp\Big\{ -\frac{\tau}{2}\Big(\mathscr{X}^T (\sum_{j=1}^{n} \mathbf{V}_j^T \mathbf{V}_j + \mathbf{E}^{-1}) \mathscr{X} - (\sum_{j=1}^{n} \mathbf{Y}_j^T \mathbf{V}_j - \sum_{j=1}^{n} \boldsymbol{\vartheta}^T \mathbf{U}_j \mathbf{V}_j + \mathbf{F}^T) \mathscr{X} - \mathscr{X}^T (\sum_{j=1}^{n} \mathbf{V}_j^T \mathbf{Y}_j - \\ &\sum_{j=1}^{n} \mathbf{V}_j^T \mathbf{U}_j^T \boldsymbol{\vartheta} + \mathbf{F}) \Big) \Big\}. \end{aligned}$$

Therefore, $\mathscr{X} \mid . \sim N(\boldsymbol{\mu}_{\mathscr{X}}, \boldsymbol{\Sigma}_{\mathscr{X}})$, where $\boldsymbol{\Sigma}_{\mathscr{X}} = \tau^{-1}(\sum_{j=1}^{n} \mathbf{V}_j^T \mathbf{V}_j + \mathbf{E}^{-1})^{-1}$ and $\boldsymbol{\mu}_{\mathscr{X}} = (\sum_{j=1}^{n} \mathbf{V}_j^T \mathbf{V}_j + \mathbf{E}^{-1})^{-1}(\sum_{j=1}^{n} \mathbf{V}_j^T(\mathbf{Y}_j - \mathbf{U}_j^T\boldsymbol{\vartheta}) + \mathbf{F})$.

Similarly,

$$\pi(\boldsymbol{\vartheta} \mid .) \propto \exp\Big\{ -\frac{\tau}{2}\Big(-\sum_{j=1}^{n} \mathbf{Y}_j^T \mathbf{U}_j^T \boldsymbol{\vartheta} - \sum_{j=1}^{n} \boldsymbol{\vartheta}^T \mathbf{U}_j \mathbf{Y}_j + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T \mathbf{U}_j \mathbf{U}_j^T \boldsymbol{\vartheta} + \sum_{j=1}^{n} \boldsymbol{\vartheta}^T \mathbf{U}_j \mathbf{V}_j \mathscr{X} + \sum_{j=1}^{n} \mathscr{X}^T \mathbf{V}_j^T \mathbf{U}_j^T \boldsymbol{\vartheta} + \boldsymbol{\vartheta}^T \mathbf{R}^{-1} \boldsymbol{\vartheta} \Big) \Big\}$$

$$= \exp\Big\{ -\frac{\tau}{2}\Big(\boldsymbol{\vartheta}^T (\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \mathbf{U}_j^T)\boldsymbol{\vartheta}) - (\sum_{j=1}^{n} \mathbf{Y}_j^T \mathbf{U}_j - \sum_{j=1}^{n} \mathscr{X}^T \mathbf{V}_j^T \mathbf{U}_j^T)\boldsymbol{\vartheta} - \boldsymbol{\vartheta}^T (\sum_{j=1}^{n} \mathbf{U}_j \mathbf{Y}_j - \sum_{j=1}^{n} \mathbf{U}_j \mathbf{V}_j \mathscr{X}) \Big) \Big\}.$$

Therefore, $\boldsymbol{\vartheta} \mid . \sim N(\boldsymbol{\mu}_{\vartheta}, \boldsymbol{\Sigma}_{\vartheta})$,
where $\boldsymbol{\Sigma}_{\vartheta} = \tau^{-1}(\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \mathbf{U}_j^T)^{-1}$ and $\boldsymbol{\mu}_{\vartheta} = (\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \mathbf{U}_j^T)^{-1} \sum_{j=1}^{n} \mathbf{U}_j (\mathbf{Y}_j - \mathbf{V}_j \mathscr{X})$.

The posterior conditional distribution of τ, with $\tau \sim Gamma(a,b)$,

$$\pi(\tau \mid .) \propto \tau^{\frac{mn + m_0 n_0 + q^\star + p_0 + 2a}{2} - 1} \exp\Big\{ -\tau \Big(\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + \sum_{j=1}^{n_0} \mathbf{Y}_{0j}^T \mathbf{Y}_{0j} + \mathscr{X}^T \mathbf{E}^{-1} \mathscr{X} - 2\mathscr{X}^T \mathbf{F} - (\sum_{j=1}^{n_0} \mathbf{V}_{oj}^T \mathbf{Y}_{0j})^T \mathbf{D} (\sum_{j=1}^{n_0} \mathbf{V}_{oj}^T \mathbf{Y}_{0j}) + \boldsymbol{\vartheta}^T \mathbf{R}^{-1} \boldsymbol{\vartheta} + 2b \Big) \Big\}.$$

Moreover,

$$\pi(\boldsymbol{\lambda} \mid .) \propto \frac{\pi(\boldsymbol{\lambda}) det(2\pi\mathbf{D})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}} \exp\Big\{ -\frac{\tau}{2}\Big(\mathscr{X}^T \mathbf{E}^{-1} \mathscr{X} - 2\mathscr{X}^T \mathbf{F} - (\sum_{j=1}^{n_0} \mathbf{V}_{oj}^T \mathbf{Y}_{0j})^T \mathbf{D} (\sum_{j=1}^{n_0} \mathbf{V}_{oj}^T \mathbf{Y}_{0j}) \Big) \Big\},$$

and

$$\pi(\boldsymbol{\kappa} \mid .) \propto \frac{\pi(\boldsymbol{\kappa})}{det(2\pi\boldsymbol{\Sigma}(\boldsymbol{\kappa}))^{1/2}} \exp\Big\{ -\frac{\tau}{2} \mathscr{X}^T \boldsymbol{\Sigma}(\boldsymbol{\vartheta}) \mathscr{X} \Big\}.$$

7.4.4.3 Under joint scalable spatial commensurate prior

Given the hierarchical model

$$\begin{aligned} \mathbf{Y}_j \mid \mathscr{X}, \boldsymbol{\vartheta} &\sim N(\boldsymbol{\eta}_j, \tau^{-1}\mathbf{I}), \\ \boldsymbol{\eta}_j &= \mathbf{U}_j^T \boldsymbol{\vartheta} + \mathbf{V} \mathscr{X}, \\ \mathscr{X} &\sim \pi^{sc}, \\ \boldsymbol{\vartheta} &\sim N(\mathbf{0}, \tau^{-1}\mathbf{R}), \\ \tau &\sim \pi(\tau),\ j = 1,2,...,n. \end{aligned}$$

Hence, the joint posterior distribution follows as

$$\pi(\mathscr{X}, \tau, \boldsymbol{\vartheta}, \boldsymbol{\lambda} \mid \mathscr{X}_1', \mathscr{D}, \mathscr{D}_0) = \frac{det(2\pi\mathbf{E})^{1/2} det(2\pi\mathbf{K})^{1/2} \pi(\boldsymbol{\lambda}) \pi(\tau)}{det(2\pi\tau^{-1}\mathbf{I})^{n/2} det(2\pi\boldsymbol{\lambda}^{-1})^{1/2} det(2\pi\mathbf{D})^{1/2} det(2\pi\mathscr{B})^{1/2} det(\tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa}))^{1/2} det(\tau^{-1}\mathbf{R})^{1/2}} \times$$
$$\exp\Big\{ -\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T \tau\mathbf{I} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + \Big(\mathscr{X}^T \mathbf{M}^{-1} \mathscr{X} - \mathbf{W}^T \mathscr{X} - \mathscr{X}^T \mathbf{W} - \mathscr{A}^T \mathscr{B}^{-1} (\mathbf{K}\mathscr{B}^{-1} - \mathbf{I}) \mathscr{A} \Big) + \mathscr{X}^T \tau \boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} \mathscr{X} + \boldsymbol{\vartheta}^T \tau \mathbf{R}^{-1} \boldsymbol{\vartheta} \Big\}$$

$$= \frac{det(2\pi\mathbf{E})^{1/2} det(2\pi\mathbf{K})^{1/2} \pi(\boldsymbol{\lambda})\pi(\tau)}{det(2\pi\tau^{-1}\mathbf{I})^{n/2} det(2\pi\boldsymbol{\lambda}^{-1})^{1/2} det(2\pi\mathbf{D})^{1/2} det(2\pi\mathscr{B})^{1/2} det(\tau^{-1}\boldsymbol{\Sigma}(\boldsymbol{\kappa}))^{1/2} det(\tau^{-1}\mathbf{R})^{1/2}} \times$$

$$\exp\Bigg\{ -\frac{1}{2}\Bigg(\sum_{j=1}^{n} \Big(\mathbf{Y}_j^T \tau\mathbf{I}\mathbf{Y}_j - \mathbf{Y}_j^T \tau\mathbf{I}\mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{Y}_j^T \tau\mathbf{I}\mathbf{V}_j \mathscr{X} - (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}\mathbf{Y}_j + (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}(\mathbf{U}_j^T \boldsymbol{\vartheta}) + (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}\mathbf{V}_j \mathscr{X} -$$

$$(\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}\mathbf{Y}_j + (\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}(\mathbf{U}_j^T \boldsymbol{\vartheta}) + (\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}(\mathbf{V}_j \mathscr{X}) + \mathscr{X}^T \mathbf{M}^{-1} \mathscr{X} - \mathbf{W}^T \mathscr{X} - \mathscr{X}^T \mathbf{W} - \mathscr{A}^T \mathscr{B}^{-1}(\mathbf{K}\mathscr{B}^{-1} - \mathbf{I})\mathscr{A} +$$

$$\mathscr{X}^T \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} \mathscr{X} + \boldsymbol{\vartheta}^T \tau\mathbf{R}^{-1} \boldsymbol{\vartheta} \Big) .$$

The posterior conditional for $\mathscr{X}$ follows as

$$\pi(\mathscr{X} \mid .) \propto \exp\Bigg\{ -\frac{1}{2}\Bigg(-\sum_{j=1}^{n} \mathbf{Y}_j^T \tau\mathbf{I}\mathbf{V}_j \mathscr{X} + \sum_{j=1}^{n} (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}\mathbf{V}_j \mathscr{X} - \sum_{j=1}^{n} (\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}\mathbf{Y}_j + \sum_{j=1}^{n} (\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}(\mathbf{U}_j^T \boldsymbol{\vartheta}) +$$

$$\sum_{j=1}^{n} (\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}(\mathbf{V}_j \mathscr{X}) + \mathscr{X}^T \mathbf{M}^{-1} \mathscr{X} - \mathbf{W}^T \mathscr{X} - \mathscr{X}^T \mathbf{W} + \mathscr{X}^T \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} \mathscr{X} \Bigg) \Bigg\}$$

$$= \exp\Bigg\{ -\frac{1}{2}\Bigg(\mathscr{X}^T (\mathbf{M}^{-1} + \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} + \sum_{j=1}^{n} \mathbf{V}_j^T \tau\mathbf{I}\mathbf{V}_j) \mathscr{X} - (\sum_{j=1}^{n} \mathbf{Y}_j^T \tau\mathbf{I}\mathbf{V}_j - \sum_{j=1}^{n} \boldsymbol{\vartheta}^T \mathbf{U}_j \tau\mathbf{I}\mathbf{V}_j + \mathbf{W}^T) \mathscr{X} -$$

$$\mathscr{X}^T (\sum_{j=1}^{n} \mathbf{V}_j^T \tau\mathbf{I}\mathbf{Y}_j - \sum_{j=1}^{n} \mathbf{V}_j^T \tau\mathbf{I}\mathbf{U}_j^T \boldsymbol{\vartheta} + \mathbf{W}) \Bigg) \Bigg\} .$$

Thus, the posterior conditional $\mathscr{X} \mid . \sim N(\boldsymbol{\mu}_{\mathscr{X}}, \boldsymbol{\Sigma}_{\mathscr{X}})$, where $\boldsymbol{\Sigma}_{\mathscr{X}} = (\mathbf{M}^{-1} + \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} + \sum_{j=1}^{n} \mathbf{V}_j^T \tau\mathbf{I}\mathbf{V}_j)^{-1}$, and $\boldsymbol{\mu}_{\mathscr{X}} = \boldsymbol{\Sigma}_{\mathscr{X}} (\tau \sum_{j=1}^{n} \mathbf{V}_j^T (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta}) + \mathbf{W})$.

Similarly,

$$\pi(\boldsymbol{\vartheta} \mid .) \propto \exp\Bigg\{ -\frac{1}{2}\Bigg(-\sum_{j=1}^{n} \mathbf{Y}_j^T \tau\mathbf{I}\mathbf{U}_j^T \boldsymbol{\vartheta} - \sum_{j=1}^{n} (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}\mathbf{Y}_j + \sum_{j=1}^{n} (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}(\mathbf{U}_j^T \boldsymbol{\vartheta}) + \sum_{j=1}^{n} (\mathbf{U}_j^T \boldsymbol{\vartheta})^T \tau\mathbf{I}\mathbf{V}_j \mathscr{X} +$$

$$\sum_{j=1}^{n} (\mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}(\mathbf{U}_j^T \boldsymbol{\vartheta}) + \boldsymbol{\vartheta}^T \tau\mathbf{R}^{-1} \boldsymbol{\vartheta} \Bigg) \Bigg\}$$

$$= \exp\Bigg\{ -\frac{1}{2}\Bigg(\boldsymbol{\vartheta}^T (\tau\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \tau\mathbf{I}\mathbf{U}_j^T) \boldsymbol{\vartheta} - (\sum_{j=1}^{n} \mathbf{Y}_j^T \tau\mathbf{I}\mathbf{U}_j^T - \mathscr{X}^T \sum_{j=1}^{T} \mathbf{V}_j^T \tau\mathbf{I}\mathbf{U}_j^T) \boldsymbol{\vartheta} - \boldsymbol{\vartheta}^T (\sum_{j=1}^{n} \mathbf{U}_j \tau\mathbf{I}\mathbf{Y}_j -$$

$$\sum_{j=1}^{n} \mathbf{U}_j \tau\mathbf{I}\mathbf{V}_j \mathscr{X}) \Bigg) \Bigg\} .$$

Therefore, the posterior conditional distribution $\boldsymbol{\vartheta} \mid . \sim N(\boldsymbol{\mu}_{\vartheta}, \boldsymbol{\Sigma}_{\vartheta})$, where $\boldsymbol{\Sigma}_{\vartheta} = (\tau\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \tau\mathbf{I}\mathbf{U}_j^T)^{-1}$ and $\boldsymbol{\mu}_{\vartheta} = (\mathbf{R}^{-1} + \sum_{j=1}^{n} \mathbf{U}_j \mathbf{U}_j^T)^{-1} \sum_{j=1}^{n} \mathbf{U}_j (\mathbf{Y}_j - \mathbf{V}_j \mathscr{X})$.

Further, assuming $\pi(\tau) \sim \text{Gamma}(a, b)$ the posterior conditional follows as

$$\pi(\tau \mid .) \propto \tau^{\frac{mn+q^\star+p_0+2a}{2}-1} \exp\Bigg\{ -\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T \tau\mathbf{I}(\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + \mathscr{X}^T \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} \mathscr{X} +$$

$$\boldsymbol{\vartheta}^T \tau\mathbf{R}^{-1} \boldsymbol{\vartheta} + 2\tau b \Bigg\}$$

$$= \tau^{\frac{mn+q^\star+p_0+2a}{2}-1} \exp\Bigg\{ -\tau\Bigg(\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + \mathscr{X}^T \tau\boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} \mathscr{X} +$$

$$\boldsymbol{\vartheta}^T \tau\mathbf{R}^{-1} \boldsymbol{\vartheta} + 2b \Big) \Bigg) \Bigg\} .$$

The posterior conditional distribution $\tau \mid . \sim \text{Gamma}(a_\tau, b_\tau)$, where $a_\tau = \frac{mn+q^\star+p_0+2a}{2}$ and $b_\tau = \left(\frac{1}{2}\Big(\sum_{j=1}^{n} (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X})^T (\mathbf{Y}_j - \mathbf{U}_j^T \boldsymbol{\vartheta} - \mathbf{V}_j \mathscr{X}) + \mathscr{X}^T \boldsymbol{\Sigma}(\boldsymbol{\kappa})^{-1} \mathscr{X} + \boldsymbol{\vartheta}^T \mathbf{R}^{-1} \boldsymbol{\vartheta} + 2b \Big) \right)$.

The posterior conditional distribution for $\boldsymbol{\lambda}$ follows as

$$\pi(\boldsymbol{\lambda} \mid .) \propto \frac{det(2\pi\mathbf{E})^{1/2} det(2\pi\mathbf{K})^{1/2}}{det(2\pi\boldsymbol{\lambda}^{-1})^{1/2}} \exp\left\{ -\frac{1}{2}\Big(\mathscr{X}^T \mathbf{M}^{-1} \mathscr{X} - \mathbf{W}^T \mathscr{X} - \mathscr{X}^T \mathbf{W} - \mathscr{A}^T \mathscr{B}^{-1} (\mathbf{K}\mathscr{B}^{-1} - \mathbf{I}) \mathscr{A} \Big) \right\} \pi(\boldsymbol{\lambda}).$$

and

$$\pi(\phi, \nu \mid .) \propto \frac{1}{det(\boldsymbol{\Sigma}(\boldsymbol{\kappa}))^{1/2}} \exp\Big\{ -\frac{\tau}{2} (\mathscr{X}^T \boldsymbol{\Sigma}(\boldsymbol{\kappa}) \mathscr{X}) \Big\} \pi(\phi, \nu).$$

7.5 Appendix of Chapter 6

7.5.1 Model prediction

This section gives a theoretical detail of the posterior predictive distribution of the response for the proposed model in the case where $\varepsilon_j = \varepsilon, \forall j$.

7.5.1.1 Prediction of replicates of the response

The goal of spatial modeling is often to develop a technique to make predictions on unobserved spatial locations on each j spatial window. Suppose the interest is to make predictions of a new replication $\mathbf{Y}_{j0}$ for study j. That is $\mathbf{Y}_{j0} = (Y_{j10}, Y_{j20}, ..., Y_{jI_j0})^T$ with corresponding spatial field replicate $\boldsymbol{\theta}_{j0} = (\theta_{j10}, \theta_{j20}, ..., \theta_{jL_j0})^T$. Let $\mathbf{X}_{j0}$ and $\mathbf{B}_{j0}$ be the corresponding design and basis matrices for the linear and nonlinear effects, and D be the observed data used in modeling. In addition, let $\Gamma_j^{(b)} = (\mathscr{N}^{(b)}, \mathscr{N}_j^{c(b)}, \{\boldsymbol{\theta}_{jt}^{(b)}\}_t)$ and the collection $(\boldsymbol{\beta}_j^{(b)}, \boldsymbol{\phi}_j^{(b)}, \Gamma_j^{(b)}, \varepsilon^{(b)}, \boldsymbol{\Lambda}^{(b)}), b = 1, 2, ..., B$ be the post-convergence posterior samples through the MCMC technique detailed in the next section. The posterior predictive distribution given the observed data follows as

$$\begin{aligned}
\mathbf{Y}_{j0} \mid \mathbf{X}_{j0}, \mathbf{B}_{j0}, D \sim & \int \pi(\mathbf{Y}_{j0} \mid \mathbf{X}_{j0}, \mathbf{B}_{j0}, \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda}, D) \times \\
& \qquad \pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda} \mid D) d[(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda})] \\
= & \sum_{w_{j0} \in \{0,1\}} \int \int \pi(\mathbf{Y}_{j0} \mid \mathbf{X}_{j0}, \mathbf{B}_{j0}, \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \boldsymbol{\theta}_{j0}, \tau_{j0}) \times \\
& \qquad \pi(\boldsymbol{\theta}_{j0}, w_{j0}, \tau_{j0}, \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda} \mid D) d[(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda})] d[(\boldsymbol{\theta}_{j0}, \tau_0)] \\
= & \sum_{w_{j0} \in \{0,1\}} \int \int \pi(\mathbf{Y}_{j0} \mid \mathbf{X}_{j0}, \mathbf{B}_{j0}, \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \boldsymbol{\theta}_{j0}, \tau_{j0}) \times \\
& \qquad \pi(\boldsymbol{\theta}_{j0} \mid w_{j0}, \Gamma_j, \boldsymbol{\Lambda}) \pi(w_{j0} \mid \varepsilon) \pi(\tau_{j0} \mid \boldsymbol{\Lambda}) \times \\
& \qquad \pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda} \mid D) d[(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda})] d[(\boldsymbol{\theta}_{j0}, \tau_0)]
\end{aligned} \tag{7.2}$$

where $\pi(w_{j0} \mid \varepsilon)$ is a Bernoulli distribution with success probability ε, and a discrete uniform over the space $\{\tau_{jt}\}_t$ for $\pi(\tau_{j0} \mid \boldsymbol{\Lambda})$. $\pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j, \varepsilon, \boldsymbol{\Lambda} \mid D)$ is the posterior distribution and $\pi(\boldsymbol{\theta}_{j0} \mid w_{j0}, \Gamma_j, \boldsymbol{\Lambda})$ is the conditional prior

distribution given as

$$\boldsymbol{\theta}_{j0} \mid \Gamma_j, \boldsymbol{\Lambda}, w_{j0} = 0 \sim \frac{1}{n_j + \alpha_j} \sum_{k=1} \delta_{\boldsymbol{\theta}_{jk}} + \frac{\alpha_j}{n_j + \alpha_j} G_j, (jk) \in \mathscr{N}_j^c, \tag{7.3a}$$

$$\boldsymbol{\theta}_{j0} \mid \Gamma_j, \boldsymbol{\Lambda}, w_{j0} = 1 \sim \frac{1}{n_0 + \alpha_0} \sum_{uk} \delta_{\boldsymbol{\theta}_{uk}} + \frac{\alpha_0}{n_0 + \alpha_0} G_0, (uk) \in \mathscr{N}, \tag{7.3b}$$

and $\pi(\mathbf{Y}_{j0} \mid \mathbf{X}_{j0}, \mathbf{B}_{j0}, \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \boldsymbol{\theta}_{j0}, \tau_{j0})$ is the data likelihood.

It is easy to draw samples of the replicate $\mathbf{Y}_{j0}$ from (7.2). w_{j0} is first drawn from a Bernoulli distribution, which determines whether the specific spatial effect or shared spatial effect of the posterior samples is to be used in the prediction. Based on the MCMC posterior samples, $\boldsymbol{\theta}_{j0}$ is drawn from (7.3a) for specific spatial effect or (7.3b) for shared spatial effect. Given $\boldsymbol{\theta}_{j0}$, $\mathbf{Y}_{j0}$ is drawn from $N(\boldsymbol{\mu}_{j0}, \tau_{j0}\mathbf{I})$, $\boldsymbol{\mu}_{j0} = \mathbf{X}_{j0}^T\boldsymbol{\beta}_j + \mathbf{B}_{j0}\boldsymbol{\phi}_j + \mathbf{A}_{j0}\boldsymbol{\theta}_{j0}$, using the posterior samples of $\boldsymbol{\beta}_j$ and $\boldsymbol{\phi}_j$ and drawing τ_{j0} from $\pi(\tau_{j0} \mid \boldsymbol{\Lambda})$.

7.5.1.2 Prediction of replicates of the response on new spatial locations

The interest is to make predictions on new spatial data locations $(\bar{s}_1, \bar{s}_2, ..., \bar{s}_{\bar{l}_0})$. Suppose that the corresponding random field at chosen spatial knots with respect to the new locations is denoted by $\bar{\boldsymbol{\theta}}_{jt} = (\bar{\theta}_{j1t}, \bar{\theta}_{j2t}, ..., \bar{\theta}_{jL_ot})^T$ and the corresponding predicted response for the jth data be $\bar{\mathbf{Y}}_{jt}$. Again, let $(\boldsymbol{\beta}_j^{(b)}, \boldsymbol{\phi}_j^{(b)}, \Gamma_j'^{(b)}, \varepsilon^{(b)}, \boldsymbol{\Lambda}^{(b)}), b = 1, 2, ..., B$ be the B post-convergence posterior draws from the joint posterior distribution. For any bth sample, the configuration of the spatial effects in $\Gamma_j'^{(b)}$ is known and contained in $\mathscr{R}_j^{(b)} = (\{\mathscr{R}_{jh}^{(b)}\}_{jh}, \{\mathscr{R}_{0h}^{(b)}\}_h)$ as detailed in the MCMC algorithm. For emphasis, $\mathscr{R}_{jh} = \{(jt) \in \mathscr{N}_j^c : \boldsymbol{\theta}_{jt} = \boldsymbol{\theta}_{jh}^\star\}$ and $\mathscr{R}_{0h} = \{(jt) \in \mathscr{N} : \boldsymbol{\theta}_{jt} = \boldsymbol{\theta}_{0h}^\star\}$, where $\boldsymbol{\theta}_{jh}^\star$ is the hth cluster value for the specific spatial effect j and $\boldsymbol{\theta}_{0h}^\star$ is the hth cluster value for the shared effect. It implies that $\Gamma_j' \equiv \{\mathscr{R}_j, \boldsymbol{\theta}_j^\star\}$, where $\boldsymbol{\theta}_j^\star = (\boldsymbol{\theta}_{j1}^\star, \boldsymbol{\theta}_{j2}^\star, ..., \boldsymbol{\theta}_{jn_j^\star}^\star, \boldsymbol{\theta}_{01}^\star, \boldsymbol{\theta}_{02}^\star, ..., \boldsymbol{\theta}_{0n_0^\star}^\star)^T$. Given the observed data $\mathscr{D}$, the prediction of $\bar{\mathbf{Y}}_{jt}$ is of interest. Thus a pair $(\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star)$ is formed, where $\tilde{\boldsymbol{\theta}}_j^\star = (\tilde{\boldsymbol{\theta}}_{j1}^\star, \tilde{\boldsymbol{\theta}}_{j2}^\star, ..., \tilde{\boldsymbol{\theta}}_{jn_j^\star}^\star, \tilde{\boldsymbol{\theta}}_{01}^\star, \tilde{\boldsymbol{\theta}}_{02}^\star, ..., \tilde{\boldsymbol{\theta}}_{0n_0^\star}^\star)^T, j = 1, 2, ..., J$ are vectors of $n_j^\star + n_0^\star$ unique cluster effects. Conditioned on the configuration $\mathscr{R}_j$, $(\boldsymbol{\theta}_{jh}^\star, \tilde{\boldsymbol{\theta}}_{jh}^\star) \perp (\boldsymbol{\theta}_{jh'}^\star, \tilde{\boldsymbol{\theta}}_{jh'}^\star), h \neq h'$ and assumes the base Gaussian prior distribution $N(\mathbf{0}_{(L_j+L_o)}, \lambda_j^{-1}\boldsymbol{\Sigma}_{j(L_j+L_o)}), j = 1, ..., J$, where $\boldsymbol{\Sigma}_j$ is the NNGP sparse representation of the full covariance matrix constructed from the Matérn covariance function while using the posterior estimates of the hyperparameters. Thus, the conditional distribution $\tilde{\boldsymbol{\theta}}_{jh}^\star \mid \boldsymbol{\theta}_{jh}^\star$ is Gaussian for all h and j. Following from (7.2), let D be the observed data and $\tilde{\mathbf{X}}_{j0}$ and $\tilde{\mathbf{B}}_{j0}$ be the design and basis matrices related to the new replicate in the new data locations, the posterior predictive of a new replication $(\mathbf{Y}_{j0}, \bar{\mathbf{Y}}_{j0})$ follows as

$$\begin{aligned}
&(\mathbf{Y}_{j0}, \tilde{\mathbf{Y}}) \mid (\mathbf{X}_{j0}, \tilde{\mathbf{X}}_{j0}), (\mathbf{B}_{j0}, \tilde{\mathbf{B}}_{j0}), D \sim \\
&\quad \sum_{w_{j0} \in \{0,1\}} \int\int \pi((\mathbf{Y}_{j0}, \tilde{\mathbf{Y}}) \mid (\mathbf{X}_{j0}, \bar{\mathbf{X}}_{j0}), (\mathbf{B}_{j0}, \tilde{\mathbf{B}}_{j0}), \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, (\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}), \tau_{j0}) \times \\
&\qquad \pi((\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}) \mid w_{j0}, (\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star), \mathscr{R}_j, \boldsymbol{\Lambda}) \pi((\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star) \mid \mathscr{R}_j, \boldsymbol{\Lambda}) \times \\
&\qquad \pi(\tau_{j0} \mid \boldsymbol{\Lambda}) \pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \mathscr{R}_j, \varepsilon, \boldsymbol{\Lambda}, w_{j0} \mid D) d[(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma_j', \varepsilon, \boldsymbol{\Lambda})] d[(\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}, \tau_0)].
\end{aligned} \tag{7.4}$$

Observe that conditioning on $\mathscr{R}_j$,

$$\begin{aligned}
\pi((\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star) \mid \mathscr{R}_j, \boldsymbol{\Lambda}) &= \prod_h \pi((\boldsymbol{\theta}_{jh}^\star, \tilde{\boldsymbol{\theta}}_{jh}^\star) \mid \mathscr{R}_j, \boldsymbol{\Lambda}) \prod_{h'} \pi((\boldsymbol{\theta}_{0h}^\star, \tilde{\boldsymbol{\theta}}_{0h}^\star) \mid \mathscr{R}_j, \boldsymbol{\Lambda}) \\
&= \prod_h \pi(\tilde{\boldsymbol{\theta}}_{jh}^\star \mid \boldsymbol{\theta}_{jh}^\star, \mathscr{R}_j, \boldsymbol{\Lambda}) \pi(\boldsymbol{\theta}_{jh}^\star \mid \mathscr{R}_j, \boldsymbol{\Lambda}) \times \\
&\quad \prod_{h'} \pi(\tilde{\boldsymbol{\theta}}_{0h}^\star \mid \boldsymbol{\theta}_{0h}^\star, \mathscr{R}_j, \boldsymbol{\Lambda}) \pi(\boldsymbol{\theta}_{0h}^\star \mid \mathscr{R}_j, \boldsymbol{\Lambda}) \\
&= \pi(\boldsymbol{\theta}_j^\star \mid \mathscr{R}_j, \boldsymbol{\Lambda}) \prod_h \pi(\tilde{\boldsymbol{\theta}}_{jh}^\star \mid \boldsymbol{\theta}_{jh}^\star, \mathscr{R}_j, \boldsymbol{\Lambda}) \prod_{h'} \pi(\tilde{\boldsymbol{\theta}}_{0h}^\star \mid \boldsymbol{\theta}_{0h}^\star, \mathscr{R}_j, \boldsymbol{\Lambda})
\end{aligned} \tag{7.5}$$

The equality in (7.5) indicates that for each cluster, the field value on the new spatial location can be drawn from a Gaussian distribution conditioning on the field value on the locations used in the modeling. The implication of (7.4)

and (7.5) is that $\pi(\boldsymbol{\theta}_j^\star \mid \mathscr{R}_j, \boldsymbol{\Lambda})\pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \mathscr{R}_j, \varepsilon, \boldsymbol{\Lambda}, w_{j0} \mid D) = \pi(w_{j0} \mid \varepsilon)\pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma'_j, \varepsilon, \boldsymbol{\Lambda} \mid D)$, which is the product of a Bernoulli distribution and the posterior distribution. Therefore, the full posterior predictive follows as

$$
\begin{aligned}
&(\mathbf{Y}_{j0}, \tilde{\mathbf{Y}}) \mid (\mathbf{X}_{j0}, \tilde{\mathbf{X}}_{j0}), (\mathbf{B}_{j0}, \tilde{\mathbf{B}}_{j0}), D \sim \\
&\quad \sum_{w_{j0} \in \{0,1\}} \int\int \pi((\mathbf{Y}_{j0}, \tilde{\mathbf{Y}}) \mid (\mathbf{X}_{j0}, \bar{\mathbf{X}}_{j0}), (\mathbf{B}_{j0}, \tilde{\mathbf{B}}_{j0}), \boldsymbol{\beta}_j, \boldsymbol{\phi}_j, (\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}), \tau_{j0}) \times \\
&\qquad \pi((\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}) \mid w_{j0}, (\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star), \mathscr{R}_j, \boldsymbol{\Lambda}) \prod_h \pi(\tilde{\boldsymbol{\theta}}_{jh}^\star \mid \boldsymbol{\theta}_{jh}^\star, \mathscr{R}_j, \boldsymbol{\Lambda}) \prod_{h'} \pi(\tilde{\boldsymbol{\theta}}_{0h}^\star \mid \boldsymbol{\theta}_{0h}^\star, \mathscr{R}_j, \boldsymbol{\Lambda}) \times \\
&\qquad \pi(\tau_{j0} \mid \boldsymbol{\Lambda}) \pi(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma'_j, \varepsilon, \boldsymbol{\Lambda} \mid D) \pi(w_{j0} \mid \varepsilon) d[(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma'_j, \varepsilon, \boldsymbol{\Lambda})] d[(\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}, \tau_0)].
\end{aligned}
\tag{7.6}
$$

where

$$
\begin{aligned}
&\pi((\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}) \mid (\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star), \mathscr{R}_j, \boldsymbol{\Lambda}, w_{j0} = 0) \\
&\qquad = \frac{1}{n_j + \alpha_j} \sum_{h=1}^{n_j^\star} n_{jh} \delta_{(\boldsymbol{\theta}_{jh}^\star, \tilde{\boldsymbol{\theta}}_{jh}^\star)} + \frac{\alpha_j}{n_j + \alpha_j} N(\mathbf{0}_{(L_j + L_o)}, \lambda_j^{-1} \boldsymbol{\Sigma}_{j(L_j + L_o)}), \text{and} \\
&\pi((\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0}) \mid (\boldsymbol{\theta}_j^\star, \tilde{\boldsymbol{\theta}}_j^\star), \mathscr{R}_j, \boldsymbol{\Lambda}, w_{j0} = 1) \\
&\qquad = \frac{1}{n_0 + \alpha_0} \sum_{h=1}^{n_0^\star} n_{0h} \delta_{(\boldsymbol{\theta}_{0h}^\star, \tilde{\boldsymbol{\theta}}_{0h}^\star)} + \frac{\alpha_0}{n_0 + \alpha_j} N(\mathbf{0}_{(L_j + L_o)}, \lambda_0^{-1} \boldsymbol{\Sigma}_{0(L_j + L_o)}).
\end{aligned}
\tag{7.7}
$$

n_{jh} and n_{0h} are the total number of replicates of the jth specific and shared effects in cluster h.

Equation 7.6 shows how samples can be drawn from the posterior predictive distribution. The idea is to (1) draw $(\boldsymbol{\beta}_j, \boldsymbol{\phi}_j, \Gamma', \varepsilon, \boldsymbol{\Lambda})$ from the posterior distribution and draw w_{j0} from a Bernoulli distribution with success probability ε. (2) Once w_{j0} is determined, draw from the conditional $\tilde{\boldsymbol{\theta}}_{jh}^\star \mid \boldsymbol{\theta}_{jh}^\star$ if $w_{j0} = 0$ or $\tilde{\boldsymbol{\theta}}_{0h}^\star \mid \boldsymbol{\theta}_{0h}^\star$ if $w_{j0} = 1$ for all h based on the posterior samples. This will gives rise to the collection of smaples $\{(\boldsymbol{\theta}_{jh}^\star, \tilde{\boldsymbol{\theta}}_{jh}^\star)\}_h$ for $w_{j0} = 0$ and $\{(\boldsymbol{\theta}_{0h}^\star, \tilde{\boldsymbol{\theta}}_{0h}^\star)\}_h$ for $w_{j0} = 1$. (3) Draw $(\boldsymbol{\theta}_{j0}, \tilde{\boldsymbol{\theta}}_{j0})$ from the conditional in (7.7) and draw $(\mathbf{Y}_{j0}, \bar{\mathbf{Y}}_{j0})$ from the data distribution.

One benefit of using spatial knots in continuous spatial modeling is that it facilitates projection. This implies that rather than selecting new spatial knots, the same spatial knots used in the modeling are used for projection, which consequently eliminates $\tilde{\boldsymbol{\theta}}^\star$ from the projection steps. Thus step (2) above in the projection steps is eliminated, and projection of the spatial effect on the new location is obtained through the projection matrix $\mathbf{A}_j$ and the estimated spatial field.

7.5.2 Details of the MCMC algorithm

This section describes the Markov Chain Monte Carlo method for drawing samples from the joint posterior distribution, $\pi(\boldsymbol{\theta}, \boldsymbol{\phi}, \boldsymbol{\beta}, \boldsymbol{\Lambda} \mid \mathbf{y})$ assuming equal rates of sharing. Begin by randomly assigning $w_{jt} \in \{0, 1\}$ with probability $1/2$ and determine $\mathscr{N}$ and $\mathscr{N}_j^c$ for all j. The initial values of the model parameters are drawn from their prior distributions. More specifically, $\{\boldsymbol{\theta}_{jt}\}$ are drawn from a multivariate normal distribution centered at zero with the corresponding prior covariance matrix $\{\lambda_j^{-1} \boldsymbol{\Sigma}_j\}$.

Step I

Draw from the posterior conditional distribution of the spatial effect using the following posterior conditional distributions. On one hand,

$$
[\boldsymbol{\theta}_{jt}, w_{jt} = 0 \mid \mathbf{y}, \boldsymbol{\Delta}_j^{-t}, \mathscr{N}_j, .] \sim \sum_{k \in \{t^\star : (jt^\star) \in \mathscr{N}_j^c \wedge t^\star \neq t\}} q_{jt_k} \delta_{\boldsymbol{\theta}_{jk}} + q_{jt_0} \mathscr{H}_{jt}, \tag{7.8}
$$

where $q_{jt_k} = \frac{v(1-\varepsilon)}{n_j-1+\alpha_j} N(\mathbf{y}_{jt}; \boldsymbol{\mu}_{jt_k}, \tau_{jt}^{-1}\mathbf{I})$, $\boldsymbol{\mu}_{jt_k} = \mathbf{X}_{jt}^T\boldsymbol{\beta}_j + \mathbf{B}_{jt}\boldsymbol{\phi}_j + \mathbf{A}_{jk}\boldsymbol{\theta}_{jk}$, $\mathbf{w}_{-jt}$ is a vector of all latent variables except w_{jt}. Moreover,

$$q_{jt_0} = \frac{v\alpha_j(1-\varepsilon)}{n_j-1+\alpha_j} \frac{det(2\pi\mathbf{E}_{jt})^{1/2}}{det(2\pi\tau_{jt}^{-1}\mathbf{I})^{1/2} det(2\pi\lambda_j^{-1}\boldsymbol{\Sigma}_j)^{1/2}} \exp\left(-\frac{1}{2}\left(d_{jt} - \mathbf{W}_{jt}^T\mathbf{E}_{jt}\mathbf{W}_{jt}\right)\right) \text{ and}$$

$$d_{jt} = \tau_{jt}\left(\mathbf{y}_{jt}^T(\mathbf{y}_{jt} - 2\mathbf{X}_{jt}^T\boldsymbol{\beta}_j - 2\mathbf{B}_{jt}\boldsymbol{\phi}_j) + (\mathbf{X}_{jt}^T\boldsymbol{\beta}_j)^T(\mathbf{X}_{jt}^T\boldsymbol{\beta}_j) + 2(\mathbf{X}_{jt}^T\boldsymbol{\beta}_j)^T(\mathbf{B}_{jt}\boldsymbol{\phi}_j) + (\mathbf{B}_{jt}\boldsymbol{\phi}_j)^T(\mathbf{B}_{jt}\boldsymbol{\phi}_j)\right).$$

$\mathscr{H}_{jt}$ is a posterior distribution assuming G_j as the prior, and it is given as $\mathscr{H}_{jt} = N(\boldsymbol{\theta}_{jt}; \mathbf{E}_{jt}\mathbf{W}_{jt}, \boldsymbol{E}_{jt})$, where $\mathbf{W}_{jt} = \tau_{jt}\mathbf{A}_{jt}^T\left(\mathbf{y}_{jt} - \mathbf{B}_{jt}\boldsymbol{\phi}_j - \mathbf{X}_{jt}^T\boldsymbol{\beta}_j\right), \mathbf{E}_{jt}^{-1} = \tau_{jt}\mathbf{A}_{jt}^T\mathbf{A}_{jt} + \lambda_j\boldsymbol{\Sigma}_j^{-1}$. On the other hand,

$$[\boldsymbol{\theta}_{jt}, w_{jt} = 1 \mid \mathbf{y}, \boldsymbol{\Delta}_0^{-jt}, \mathscr{N}, .] \sim \sum_{(uk)\in\{(jt)^\star\in\mathscr{N}:(jt)^\star\neq(jt)\}} q_{ut_k}\delta_{\boldsymbol{\theta}_{uk}} + q_{0t_0}\mathscr{H}_{0t}, \tag{7.9}$$

where $q_{ut_k} = \frac{v\varepsilon}{n_0-1+\alpha_0} N(\mathbf{y}_{jt}; \boldsymbol{\mu}_{jt_{uk}}, \tau_{jt}^{-1}\mathbf{I}), \boldsymbol{\mu}_{jt_{uk}} = \mathbf{X}_{jt}^T\boldsymbol{\beta}_j + \mathbf{B}_{jt}\boldsymbol{\phi}_j + \mathbf{A}_{uk}\boldsymbol{\theta}_{uk}$. $\mathscr{H}_{0t}$ is a posterior distribution given as $\mathscr{H}_{0t} = N(\boldsymbol{\theta}_{jt}; \mathbf{E}_{0t}\mathbf{W}_{jt}, \mathbf{E}_{0t})$, where $\mathbf{E}_{0t}^{-1} = (\tau_{jt}\mathbf{A}_{jt}^T\mathbf{A}_{jt} + \lambda_0\boldsymbol{\Sigma}_0^{-1})$ and

$$q_{0t_0} = \frac{v\alpha_0\varepsilon}{n_0-1+\alpha_0} \frac{det(2\pi\mathbf{E}_{0t})^{1/2}}{det(2\pi\tau_{jt}^{-1}\mathbf{I})^{1/2} det(2\pi\lambda_0^{-1}\boldsymbol{\Sigma}_0)^{1/2}} \exp\left(-\frac{1}{2}\left(d_{jt} - \mathbf{W}_{jt}^T\mathbf{E}_{0t}\mathbf{W}_{jt}\right)\right).$$

v is obtained such that $\sum_{k\in\{t^\star:(jt^\star)\in\mathscr{N}_j^c\wedge t^\star\neq t\}} q_{jt_k} + q_{jt_0} + \sum_{(uk)\in\{(jt)^\star\in\mathscr{N}:(jt)^\star\neq(jt)\}} q_{ut_k} + q_{0t_0} = 1$. That is, for any chosen pair $(\boldsymbol{\theta}_{jt}, w_{jt} = 0)$, q_{jt_k} is the probability that $\boldsymbol{\theta}_{jt}$ belongs to the specific data j and takes up the value of $\boldsymbol{\theta}_{jk}, k \neq t$ and with probability q_{jt_0}, it takes up a new value from $\mathscr{H}_{jt}$. Similar for the pair $(\boldsymbol{\theta}_{jt}, w_{jt} = 1)$, q_{ut_k} is the probability that $\boldsymbol{\theta}_{jt}$ belongs to the shared component and takes up the value $\boldsymbol{\theta}_{uk}$, which belongs to the shared set, and with probability q_{0t_0}, $\boldsymbol{\theta}_{jt}$ takes up a new value drawn from the posterior $\mathscr{H}_{0t}$ based on the shared prior distribution G_0. After sampling $\boldsymbol{\theta}_{jt}$, update the set $\mathscr{N}$ and repeat the process for all $j = 1,2,...,J$ and $t = 1,2,...,n_j$.

Step II

Given the updated set of all pairs $(\boldsymbol{\theta}, \mathbf{w}) = \{(\boldsymbol{\theta}_{jt}, w_{jt}), j = 1,2,...,J, t = 1,2,...,n_j\}$, obtain the clusters and remix the cluster effects. Let $\boldsymbol{\theta}_{jh}^\star$ be the jth data specific spatial effect of cluster h and let $\boldsymbol{\theta}_{0h}^\star$ be the shared spatial effect for cluster h, and $\mathscr{R}_{jh} = \{(jt) \in \mathscr{N}_j^c : \boldsymbol{\theta}_{jt} = \boldsymbol{\theta}_{jh}^\star\}$ and $\mathscr{R}_{0h} = \{(jt) \in \mathscr{N} : \boldsymbol{\theta}_{jt} = \boldsymbol{\theta}_{0h}^\star\}$ are the corresponding index sets. Given $\mathscr{R}_{jh}$, update the cluster spatial effect from the posterior distribution based on the base prior distribution G_j with those responses whose indexes are in $\mathscr{R}_{jh}$. The posterior distributions follow as

$$\begin{aligned} \boldsymbol{\theta}_{jh}^\star \mid \mathbf{y}, \mathscr{R}_{jh}, . &\sim N(\boldsymbol{\mu}_{jh}, \boldsymbol{\Sigma}_{jh}), \\ \boldsymbol{\theta}_{0h}^\star \mid \mathbf{y}, \mathscr{R}_{0h}, . &\sim N(\boldsymbol{\mu}_{0h}, \boldsymbol{\Sigma}_{0h}), \end{aligned} \tag{7.10}$$

where $\boldsymbol{\mu}_{jh} = \boldsymbol{\Sigma}_{jh}\mathbf{W}_{jh}, \boldsymbol{\Sigma}_{jh}^{-1} = \left(\sum_{(jt)\in\mathscr{R}_{jh}} \tau_{jt}\mathbf{A}_{jt}^T\mathbf{A}_{jt} + \lambda_j\boldsymbol{\Sigma}_j^{-1}\right)$ and $\mathbf{W}_{jt} = \sum_{(jt)\in\mathscr{R}_{jh}} \tau_{jt}\mathbf{A}_{jt}^T\left(\mathbf{y}_{jt} - \mathbf{X}_{jt}^T\boldsymbol{\beta}_j - \mathbf{B}_{jt}\boldsymbol{\phi}_j\right)$, $\boldsymbol{\mu}_{0h} = \boldsymbol{\Sigma}_{0h}\mathbf{W}_{0h}, \boldsymbol{\Sigma}_{0h}^{-1} = \left(\sum_{(jt)\in\mathscr{R}_{0h}} \tau_{jt}\mathbf{A}_{jt}^T\mathbf{A}_{jt} + \lambda_0\boldsymbol{\Sigma}_0^{-1}\right)$ and $\mathbf{W}_{0h} = \sum_{(jt)\in\mathscr{R}_{0h}} \tau_{jt}\mathbf{A}_{jt}^T\left(\mathbf{y}_{jt} - \mathbf{X}_{jt}^T\boldsymbol{\beta}_j - \mathbf{B}_{jt}\boldsymbol{\phi}_j\right)$. Based on these updates of the cluster values, the elements of each cluster are assigned their cluster values to proceed in the estimation process.

Step III

Update the hyperparameters α_j. For a given j, let $n_j^\star$ be the number of clusters for the specific spatial effect and $n_0^\star$ be the number of clusters in the shared spatial effect. That is, for a fixed j, $n_j^\star$ is the number of unique spatial effect in $\{\boldsymbol{\theta}_{jt}\}_t$ for which the latent variable $w_{jt} = 0$ and $n_0^\star$ is the number of unique spatial effect in $\{\boldsymbol{\theta}_{jt}\}_{jt}$ for which the latent variable $w_{jt} = 1$, and consequently $\mathscr{R}_{jh}, h = 1,2,...,n_j^\star$ and $\mathscr{R}_{0h}, h = 1,2,...,n_0^\star$. Let γ_j be a latent variable such that given the current value of α_j and $n_j^\star$, $\gamma_j \sim Beta(\alpha_j + 1, n_j^\star)$ and α_j is then updated from the mixture

$$\begin{aligned} \alpha_j \mid \gamma_j, n_j^\star &\sim \pi_{\alpha_j} Gamma(a_j + n_j^\star, b_j - \log\gamma_j) + (1 - \pi_{\alpha_j}) Gamma(a_j + n_j^\star - 1, b_j - \log\gamma_j), \\ \pi_{\alpha_j} &= c_{\alpha_j}/(1 + c_{\alpha_j}), c_{\alpha_j} = \frac{a_j + n_j^\star - 1}{n_j(b_j - \log\gamma_j)}, \; j = 1,2,...,J, \end{aligned} \tag{7.11}$$

which also applies for α_0.

Step IV

Next, update the linear and nonlinear effects. Conditioning on the spatial effects, the response variables for each study are independent of those in a different study, and thus, the posterior conditional distributions of $\boldsymbol{\beta}_j$ and $\boldsymbol{\phi}_j$ only consider the responses from the jth study which are given as

$$\begin{aligned}\boldsymbol{\beta}_j \mid \mathbf{y}, . &\sim N(\boldsymbol{\mu}_{\beta_j}, \boldsymbol{\Sigma}_{\beta_j}),\\ \boldsymbol{\phi}_j \mid \mathbf{y}, . &\sim N(\boldsymbol{\mu}_{\phi_j}, \boldsymbol{\Sigma}_{\phi_j}),\end{aligned} \tag{7.12}$$

where $\boldsymbol{\mu}_{\beta_j} = \boldsymbol{\Sigma}_{\beta_j}\mathbf{W}_{\beta_j}, \boldsymbol{\Sigma}_{\beta_j}^{-1} = \left(\sum_{t=1}^{n_j}\tau_{jt}\mathbf{X}_{jt}\mathbf{X}_{jt}^T + Q_{\beta_j}\right)$ and $\mathbf{W}_{\beta_j} = \sum_{t=1}^{n_j}\tau_{jt}\mathbf{X}_{jt}\left(\mathbf{y}_{jt} - \mathbf{B}_{jt}\boldsymbol{\phi}_j - \mathbf{A}_{jt}\boldsymbol{\theta}_{jt}\right)$, $\boldsymbol{\mu}_{\phi_j} = \boldsymbol{\Sigma}_{\phi_j}\mathbf{W}_{\phi_j}, \boldsymbol{\Sigma}_{\phi_j}^{-1} = \left(\sum_{t=1}^{n_j}\tau_{jt}\mathbf{B}_{jt}^T\mathbf{B}_{jt} + Q_{\phi_j}\right)$, $\mathbf{W}_{\phi_j} = \sum_{t=1}^{n_j}\tau_{jt}\mathbf{B}_{jt}^T\left(\mathbf{y}_{jt} - \mathbf{X}_{jt}^T\boldsymbol{\beta}_j - \mathbf{A}_{jt}\boldsymbol{\theta}_{jt}\right), Q_{\phi_j}^{-1} = \tau_{\phi_j}^{-1}\mathbf{R}\mathbf{R}^{\mathbf{T}}$, and $Q_{\beta_j}^{-1} = \tau_{\beta_j}^{-1}\mathbf{I}$, with fixed hyperparameters τ_{β_j} and τ_{ϕ_j}. Due to the rank deficiency of the nonlinear effects, a sum to zero constraint, $\sum_r \phi_{jr} = 0$ is imposed for each study effect.

Step V

Similarly, update the precision hyperparameters τ_{jt} from their corresponding posterior conditional distribution. Given the clustering indices, the prior distribution for each cluster spatial field is known, and thus, the posterior conditional follows as $\tau_{jt} \mid \mathbf{y}_{jt} . \sim Gamma(c_{\tau_{jt}}, d_{\tau_{jt}})$, where $c_{\tau_{jt}} = 0.5(I_j + 2a_{\tau_{jt}})$ and $d_{\tau_{jt}} = 0.5\left((\mathbf{y}_{jt} - \boldsymbol{\mu}_{jt})^T(\mathbf{y}_{jt} - \boldsymbol{\mu}_{jt}) + 2b_{\tau_{jt}}\right)$. Similarly, update $\lambda_j, j = 1,2,...,J$ and λ_0 from the posterior conditional distribution. That is, $\lambda_j \mid \mathbf{y}, \{\mathscr{R}_{jh}\}_h . \sim Gamma(\lambda_{a_j}, \lambda_{b_j})$, where $\lambda_{a_j} = 0.5(n_j^\star L_j + 2a_{\lambda_j})$ and $\lambda_{b_j} = 0.5(\sum_{h=1}^{n_j^\star}\boldsymbol{\theta}_{jh}^{\star T}\boldsymbol{\Sigma}_j\boldsymbol{\theta}_{jh}^\star + 2b_{\lambda_j})$. This also applies to λ_0. Update $\tau_{\beta_j} \mid . \sim Gamma(0.5p_j + a_{\beta_j}, 0.5\boldsymbol{\beta}_j^T\boldsymbol{\beta}_j + b_{\beta_j})$ and $\tau_{\phi_j} \mid . \sim Gamma(0.5r_j + a_{\phi_j}, 0.5\boldsymbol{\phi}_j^T(\mathbf{R}\mathbf{R}^T)^{-1}\boldsymbol{\phi}_j + b_{\phi_j})$, where p_j and r_j are the corresponding vector length of the fixed and nonlinear effects.

Step VI

Next, update the weight parameter ε from the posterior conditional. Let $N_1 = \sum_{j=1}^{J}\sum_{t=1}^{n_j} w_{jt}$ and $N_0 = n - N_1, n = \sum_j n_j$. Based on the prior distribution given in Equation (6.10) for all t and j and the Beta prior for ε, the posterior conditional distribution follows as

$$\begin{aligned}\varepsilon \mid \mathscr{N}, \mathscr{N}^c . &\propto (1-\varepsilon)^{N_0}\varepsilon^{N_1}\left(Beta(\varepsilon; a_\varepsilon, b_\varepsilon)\right)\\ &\sim Beta(\varepsilon; N_1 + a_\varepsilon, N_0 + b_\varepsilon),\end{aligned} \tag{7.13}$$

That is, ε is drawn from a Beta distribution. Note that the posterior conditional given in Equation 7.13 depends on the data through N_0 and N_1, which are updated at every step in the chain.

Step VII

Recall that for each base prior distribution of the spatial component, there are associated parameters σ_j^2, ν_j, and κ_j that control the variance (partial sill), smoothness, and range of the stationary spatial field, which may be difficult to elicit. The Metropolis-Hastings algorithm is used to update each parameter. By construction, in this work, $\sigma_j^2 = \lambda_j^{-1}\tau_\theta^{-1}$, where τ_θ is a fixed hyperparameter, and (κ_j, ν_j) are updated from their joint conditional posterior distribution. Conditioning on the latent variable and assigning an appropriate joint prior distribution, the posterior conditional distribution follows as

$$\begin{aligned}(\kappa_j, \nu_j) \mid \{\mathscr{R}_{jh}\}_h, . &\propto \prod_{h=1}^{n_j^\star}\prod_{(jt),\in\mathscr{R}_{jh}} N\left(\boldsymbol{\theta}_{jt}; \mathbf{0}, \lambda_j^{-1}\boldsymbol{\Sigma}_j\right)\pi(\kappa_j, \nu_j),\ j = 1,2,...,J \text{ and}\\ (\kappa_0, \nu_0) \mid \{\mathscr{R}_{0h}\}_h, . &\propto \prod_{h=1}^{n_0^\star}\prod_{(jt)\in\mathscr{R}_{0h}} N\left(\boldsymbol{\theta}_{jt}; \mathbf{0}, \lambda_0^{-1}\boldsymbol{\Sigma}_0\right)\pi(\kappa_0, \nu_0).\end{aligned} \tag{7.14}$$

An independent lognormal prior distribution for $\pi(\kappa_j, \nu_j), j = 1,2,...,J$ and $\pi(\kappa_0, \nu_0)$ were assumed. The parameters are then updated using the Metropolis-Hastings algorithm. An independent truncated normal proposal distribution was used. Suppose at iteration $\mathtt{t}$ the current value are $(\kappa_j^\mathtt{t}, \nu_j^\mathtt{t})$, proposals are drawn from $q(\kappa_j^\star, \nu_j^\star \mid \kappa_j^\mathtt{t}, \nu_j^\mathtt{t})$ and subjected to a constraint $det(\boldsymbol{\Sigma}_j) > 0$. The acceptance probability is then computed as

$$\zeta_j = min\left(1, \frac{\pi\left(\kappa_j^\star, \nu_j^\star \mid \{\mathscr{R}_{jh}\}_h, .\right) q(\kappa_j^t, \nu_j^t \mid \kappa_j^\star, \nu_j^\star)}{\pi\left(\kappa_j^t, \nu_j^t \mid \{\mathscr{R}_{jh}\}_h, .\right) q(\kappa_j^\star, \nu_j^\star \mid \kappa_j^t, \nu_j^t)}\right). \tag{7.15}$$

Draw $U \sim Uniform(0,1)$ and set $(\kappa_j^{t+1}, \nu_j^{t+1}) = (\kappa_j^\star, \nu_j^\star)$ if $U < \zeta_j$, otherwise, set $(\kappa_j^{t+1}, \nu_j^{t+1}) = (\kappa_j^t, \nu_j^t)$.

The `R` code used for the estimation is deposited on a GitHub repository, accessible through the link https://github.com/eosafu/mixtureDPspatial.

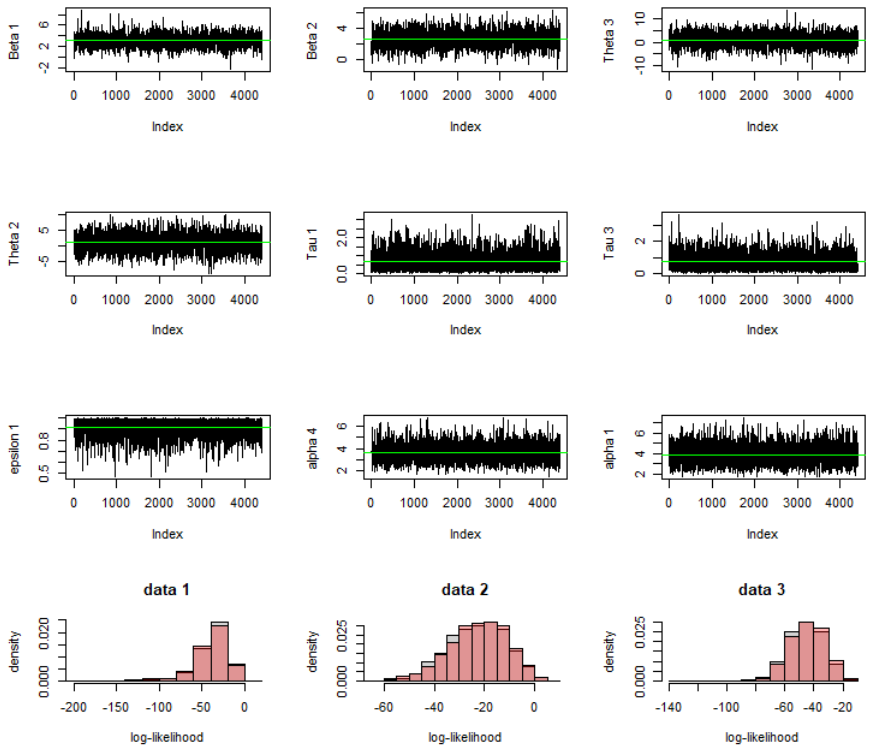


Figure 51 – TMS data analysis trace plot of some selected parameters and the histogram of the MCMC chain of the log-likelihood. The histogram shows the first chain of the log-likelihood overlapped on the second chain, indicating an overall convergence of the chain.

7.6 Correcting for spatial confounding

A situation where the observed fixed effect covariates in $\mathbf{X}$ are correlated with the spatial effect can lead to estimation bias, especially when the spatial effect has a highly effective range of spatial autocorrelation (PACIOREK, 2010). This problem is referred to as spatial confounding. In the presence of spatial confounding, the inference on the importance of the fixed effect covariates based on $\boldsymbol{\beta}$ estimates can be misleading (HANKS *et al.*, 2015). The restricted spatial regression (RSR) technique has been used to alleviate spatial confounding. See for example, Hanks *et al.* (2015) and Azevedo, Prates and Bandyopadhyay (2022). The RSR technique constrains the spatial effect to be orthogonal to the fixed effects in $\mathbf{X}$ (i.e. $\mathbf{XA}\boldsymbol{\theta} = \mathbf{0}$). Suppose the nonlinear effect $\mathbf{B}\boldsymbol{\phi}$ is not confounding with the spatial component in the linear predictor $\boldsymbol{\mu} = \mathbf{X}^T\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\phi} + \mathbf{A}\boldsymbol{\theta}$. Within the modeling framework developed in this work, the conditional posterior distribution of the fixed effect follows as

$$\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\theta}, . \sim N(\boldsymbol{\mu}_\beta, \boldsymbol{\Sigma}_\beta),$$

$\boldsymbol{\mu}_\beta = \boldsymbol{\Sigma}_\beta \mathbf{W}_\beta, \boldsymbol{\Sigma}_\beta^{-1} = \left(\tau \mathbf{X}\mathbf{X}^T + \tau_\beta\right)$ and $\mathbf{W}_\beta = \tau \mathbf{X}\Big(\mathbf{y} - \mathbf{B}\boldsymbol{\phi} - \mathbf{A}\boldsymbol{\theta})\Big)$. The "." is used to denote other parameters. Using the law of total expectation and variance, the marginal mean and variance are

$$
\begin{aligned}
\mathbb{E}(\boldsymbol{\beta} \mid \mathbf{y}, .) &= \mathbb{E}(\mathbb{E}(\boldsymbol{\beta} \mid \mathbf{y}, \boldsymbol{\theta}, .) \mid \mathbf{y}) \\
&= \mathbb{E}\left(\left(\tau \mathbf{X}\mathbf{X}^T + \tau_\beta \mathbf{I}\right)^{-1} \tau \mathbf{X}\Big(\mathbf{y} - \mathbf{B}\boldsymbol{\phi} - \mathbf{A}\boldsymbol{\theta}\Big) \mid \mathbf{y} \right) \\
&= \left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1} \mathbf{X}\Big(\mathbf{y} - \mathbf{B}\boldsymbol{\phi} - \mathbf{A}\mathbb{E}(\boldsymbol{\theta} \mid \mathbf{y})\Big) \\
&= \boldsymbol{\beta}_R - \mathbf{U}\boldsymbol{\phi} - \mathbf{V}\hat{\boldsymbol{\theta}}, \\
Var(\boldsymbol{\beta} \mid \mathbf{y}, .) &= \mathbb{E}(Var(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mathbf{y}) \mid \mathbf{y}) + Var(\mathbb{E}(\boldsymbol{\beta} \mid \boldsymbol{\theta}, \mathbf{y}) \mid \mathbf{y}) \\
&= \tau^{-1}\left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1} + \left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1} \mathbf{X}\mathbf{A} Var(\boldsymbol{\theta} \mid \mathbf{y}) \mathbf{A}^T \mathbf{X}^T \left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1}
\end{aligned}
$$

where $\tau_\beta^\star = \tau_\beta / \tau$, $\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta} \mid \mathbf{y}, .)$, $\mathbf{U} = \left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1} \mathbf{X}\mathbf{B}$, $\mathbf{V} = \left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1} \mathbf{X}\mathbf{A}$, and $\boldsymbol{\beta}_R$ is the ridge regression estimate. By adding the spatial effect to the model, the Bayes estimator (under mean square error) is a shift on the estimator without the effect and the marginal variance increases. Carefully examining the equation above, the prediction of $\mathbf{y}$ based on the fixed effect is $\mathbf{X}^T\left(\boldsymbol{\beta}_R - \mathbf{U}\boldsymbol{\phi} - \mathbf{V}\hat{\boldsymbol{\theta}}\right)$. When the spatial effect is correlated with the fixed effect, then the quantity $\mathbf{X}^T\mathbf{V}\hat{\boldsymbol{\theta}} = \mathbf{P}(\tau_\beta^\star)\mathbf{A}\hat{\boldsymbol{\theta}}$, where $\mathbf{P}(\tau_\beta^\star) = \mathbf{X}^T\left(\mathbf{X}\mathbf{X}^T + \tau_\beta^\star \mathbf{I}\right)^{-1}\mathbf{X}$ in the prediction of $\mathbf{y}$ indicates a duplicated information. To ensure orthogonality, define the new adjusted spatial effect as $\mathbf{u} = (\mathbf{I} - \mathbf{P}(\tau_\beta^\star))\mathbf{A}\boldsymbol{\theta}$, implying that as $\tau_\beta \to 0 \implies \mathbf{X}\mathbf{u} \to \mathbf{0}$, and it is exact for a flat prior for $\boldsymbol{\beta}$ ($\tau_\beta = 0$). Thus, the adjusted linear predictor associated with $\mathbf{y}$ is given as

$$
\boldsymbol{\mu} = \mathbf{X}^T\boldsymbol{\beta}^\star + \mathbf{B}\boldsymbol{\phi} + \mathbf{A}^\star\boldsymbol{\theta}, \tag{7.16}
$$

where $\mathbf{A}^\star = (\mathbf{I} - \mathbf{P}(\tau_\beta^\star))\mathbf{A}$. Here, $\mathbf{P}(0)$ is the projection of the spatial effect onto the space of $\mathbf{X}$. The linear predictor 7.16 is similar to those used in Chapter 4, 5, and 6, and hence can be implemented in a straightforward manner for the developed models. For the developed interface in Chapter 4, the predictor was achieved as follows: obtain the orthogonal matrix $\mathbf{X}^\star$ from the QR decomposition of $\mathbf{X}$. Then constrain the spatial effect such that $\mathbf{X}^\star\mathbf{A}\boldsymbol{\theta} = \mathbf{0}$. The constraint is easily achieved in `inla()` function by using the argument `extraconstr` and `e` in the function `f()` that defined the spatial model. That is,

```
f(s, model, extraconstr =list(A=G),e=rep(0,p)),
```
(7.17)

where `s` is the observed spatial covariate, “model” is the spatial model, $\texttt{G} = \mathbf{X}^{\star T}\mathbf{A}$, `rep(0,p)` is a vector of zeroes with size p, which is the number of fixed effects covariates including the intercept.


ICMC
USP
SÃO CARLOS