Regional Adaptive Smoothing in Disease Mapping Models

Maqsood Ali

Dissertation
an der Fakultät für Mathematik, Informatik und Statistik
der Ludwig Maximilians Universität München

München, den 07.05.2020

Erstberichterstatter: Prof. Dr. Volker Schmid

Zweitberichterstatter: Prof. Dr. Andrea Riebler

Tag der Disputation: 04.09.2020

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8, Abs. 2 Pkt. .5.)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

Munich, Sep 30, 2020 **Ali, Maqsood**

TO MY PARENTS

WITHOUT THEIR LOVE AND PRAYERS, I AM NOTHING.

Acknowledgments

First and foremost thanks to **Allah Almighty**. All praises to Him for the strengths and His blessing in completing this thesis, without His blessing it is impossible to complete this dissertation. I offer my humblest, sincerest and million Darood to Holy Prophet **Muhammad (Peace Be Upon Him)**, who exhorts His followers to seek for Knowledge from cradle to grave.

I deeply appreciate the complementary perspective of my honorable supervisor **Prof. Dr. Volker Schmid** and all he has contributed to the shape of this thesis. I thank him for his guidance, for great working relationship, and for the grasp of broad concept and attention to details. His invaluable help of constructive comments and suggestions throughout the thesis work have contributed to the success of this research. Without his patience, encouragement and intuitive suggestions, I could not finish my dissertation so smoothly.

I am thankful to the Higher Education Commission (HEC) of Pakistan for providing me scholarship for PhD. I also acknowledge the support of Deutscher Akademischer Austauschdienst (DAAD).

I am heartily grateful and thankful to my friends **Ayaz Ali**, **Tanveer Akhtar** and **Muhammad Shahzad**, who always encouraged me a lot. I am also thankful to the departmental staff for their help in every matter.

Last but not the least, I express my love for my brothers **Arshad Ali**, **Amjad Ali**, **Mehboob Ali**, and my sisters, for their prayers and wishes have been a great source of comfort and ease during my whole study period.

Maqsood Ali, Munich

Abstract

Disease mapping focuses on estimating spatial patterns and evolution of disease risk based on measures of disease effect, incidence ratio, for instance. These measures on maps provide a good visual representation of disease risk, featuring spatial heterogeneities and highlighting units or clusters of high risk. There are a number of factors due to which some units may have higher numbers of disease incidences compared to others, for example, differences in environmental exposures, deprived communities, different administrative structures, lack of awareness about the disease(s), to name a few. To clearly differentiate areas of low or high risk, we must apply some type of smoothing over regions which are presumably similar to each other. The process of such local smoothing increases our ability to clearly discern clusters in the spatial variation. In this thesis, we propose a novel approach to smooth local spatial units based on their larger regional location. The degree of smoothing is constant within each region, but differs between regions. For the purpose of illustration, we consider the spatial structure of German counties as local spatial units with Federal states of Germany as larger regions to apply state-wise adaptive smoothing.

Chapters 4 and 5 propose univariate and multivariate spatial models of regional smoothing. We define that the incidences in each county follow a binomial density with probability of risk linked to spatial correlation matrix in a hierarchical way. The correlation matrix is partitioned into sub-matrices, corresponding to regions (Federal states), and smoothing parameters are introduced into the sub-matrices to locally smooth regions. Appropriate prior assumptions are stated for unknown parameters and samples from full conditional posterior densities are generated using MCMC. In Chapter 5, we adopt coregionalization framework of MacNab (2016) to build multivariate GMRFs as a linear combination of latent independent univariate GMRFs. The smoothing parameters are first applied to each sample separately, in a similar fashion to univariate regionalized spatial model, and then combined in the form of joint correlation matrix.

We use the approach of Anderson et al. (2014) to identify spatial units exhibiting alike disease risks. The approach first elicits configuration of clusters based on past data of disease. In the second step, it fits a Poisson log-linear model using current data to select the best configuration based on deviance information criterion.

The proposed method of smoothing is illustrated using real data sets of Oral cancer (univariate) and Colon, Lung and Pancreatic cancers (multivariate) on spatial structure of German counties. We are able to identify 13 clusters of Oral cancer, 9 of Colon, 6 of Lung and 8 of Pancreatic cancer. The identified clusters are further ranked based on incidence ratios. The analysis of real data and its comparison with simple GMRF (BYM model of Besag et al. (1991)) reveals that the novel method of incorporating smoothing parameters in spatial correlation matrix performs equally well, if not better.

Zusammenfassung

Krankheitskartierung (Dieses Mapping) befasst sich mit der Schätzung räumlicher Muster und Entwicklungen des Krankheitsrisikos auf der Grundlage von Messungen des Krankheitsinzidenz. Die Kartierung von Messungen auf Landkarten bietet eine gute visuelle Darstellung des Krankheitsrisikos, wobei räumliche Heterogenitäten deutlich werden und Einheiten und Cluster mit hohem Risiko sichtbar gemacht werden können. Aufgrund einer Reihe von Faktoren können manche Einheiten im Vergleich zu anderen eine höhere Anzahl von Krankheitsfällen aufweisen, z.B. wegen unterschiedlicher Umweltbelastungen, sozial benachteiligter Bevölkerungsgruppen, unterschiedlicher Verwaltungsstrukturen, fehlendem Bewusstsein für die Krankheit(en), um nur einige Faktoren zu nennen. Um Bereiche mit niedrigem oder hohem Risiko klar zu unterscheiden, muss eine Art Glättung über Regionen vorgenommen werden, die einander ähnlich sein dürften. Der Prozess einer solchen lokalen Glättung verbessert den Prozess der klaren Unterscheidung von Clustern hinsichtlich der räumlichen Variation. In dieser Arbeit schlagen wir einen neuartigen Ansatz zur Glättung lokaler räumlicher Einheiten auf der Grundlage ihrer großräumigen regionalen Lage vor. Der Grad der Glättung ist innerhalb jeder Region konstant, unterscheidet sich jedoch von Region zu Region. Zur Veranschaulichung betrachten wir die räumliche Struktur von deutschen Landkreisen als lokale Raumeinheiten mit Bundesländern als größere Regionen, um die länderspezifische adaptive Glättung anzuwenden.

In den Kapiteln 4 und 5 schlagen wir univariate und multivariate räumliche Modelle zur regionalen Glättung vor. Wir legen fest, dass die Inzidenzen in jedem Kreis einer binomialen Dichte mit Risikowahrscheinlichkeit folgen, die mit einer räumlichen Korrelationsmatrix in hierarchischer Weise verknüpft ist. Die Korrelationsmatrix wird in Untermatrizen unterteilt, die Regionen (Bundesländern) entsprechen, und Glättungsparameter werden in die Untermatrizen eingeführt, um die Regionen lokal zu glätten. Für unbekannte Parameter werden geeignete Vorannahmen aufgestellt und mit Hilfe von MCMC werden Stichproben aus den Posterioridichten generiert. In Kapitel 5 übernehmen wir den Koregionalisierungsrahmen von MacNab (2016), um multivariate GMRFs als Linearkombination von latenten unabhängigen univariaten GMRFs zu erstellen. Die Glättungsparameter werden zunächst auf jede Stichprobe einzeln angewendet, ähnlich wie bei einem univariaten regionalisierten Raummodell, und dann in Form einer gemeinsamen Korrelationsmatrix kombiniert.

Wir verwenden den Ansatz von Anderson et al. (2014), um räumliche Einheiten zu identifizieren, die ähnliche Krankheitsrisiken aufweisen. Der Ansatz eruiert zunächst die Konfiguration von Clustern auf der Grundlage früherer Krankheitsdaten. Im zweiten Schritt passt er ein log-lineares Poisson-Modell unter Verwendung aktueller Daten an, um die beste Konfiguration auf der Grundlage des Abweichungsinformationskriteriums auszuwählen.

Die vorgeschlagene Methode der Glättung wird anhand realer Datensätze von Mundkrebs (univariat) und Dickdarm-, Lungen- und Bauchspeicheldrüsenkrebs (multivariat) auf der räumlichen Struktur deutscher Landkreise veranschaulicht. Wir sind in der Lage, 13 Cluster für Mundkrebs, 9 für Dickdarmkrebs, 6 für Lungenkrebs und 8 für Bauchspeicheldrüsenkrebs zu identifizieren. Die identifizierten Cluster werden anhand von Inzidenzverhältnissen weiter

gereiht. Die Analyse der realen Daten und ihr Vergleich mit dem einfachen GMRF (BYM-Modell von Besag et al. (1991)) zeigt, dass die neuartige Methode der Einbeziehung von Glättungsparametern in der räumlichen Korrelationsmatrix gleich gut, wenn nicht sogar besser abschneidet.

Contents

List of Figures

Chapter 1

Introduction

1.1 Disease Mapping

There are a variety of reasons to study the geographical behavior or spread of a disease. Maps of a disease help quantify its global burden and allow an immediate visualization of the extent and magnitude of the scale and foci of disease. They can support carefully weighted assessments by policy makers on the outcomes of alternative courses of actions. They can be used as a baseline to monitor the success/failure of public health intervention efforts.

The field of disease mapping focuses on estimating spatial patterns and evolution of disease risk based on measures of disease effect, incidence ratio, for instance. These measures are obtained from spatial areas partitioned into non-overlapping spatial units. The main theme is to utilize these measures to obtain statistically precise estimates of disease risk for each spatial unit.

Design based approach through national or state level studies is often infeasible because statistical precision imposes certain restrictions such as a *sufficient* sample size for each spatial unit. This is generally unattainable, there are many reasons to this end, study cost/labor etc. Furthermore, these studies provide estimates on an aggregate scale i.e., national or state level and insufficient sample size for each unit produces poor estimates locally (Schaible (2013)).

On the other hand, model based studies provide the flexibility to borrow *local strength* of units. These models provide a mechanism to smooth extreme values across units to improve local estimates. It is similar to small area estimation problem but only in the context of non-spatial data, because the problem assumes that every unit informs (affects) equally about (to) every other unit. However it is not true in spatial data as the relative location affects units in space and their position is of prime importance in disease mapping models to provide locally smooth estimates. Methods in disease mapping assume positive spatial correlation between observations and local observations borrow additional information from geographical proximity in order to improve (smooth) local geographical mean (chapter 14 of Gelfand et al. (2010)).

1.2 Spatial Clusters

Disease maps provide a good visual representation of disease risk, featuring spatial heterogeneities and highlighting units of high risk. There are a number of factors due to which some units may have higher numbers of disease affected subjects compared to others, for example, differences in environmental exposures, deprived communities, different administrative structures, lack of awareness about the disease(s), to name a few. Cluster detection is focused on such local features of the risk between areas as the location of spatial units may suggest some connection to potential factors associated with disease. Hence, it is important to identify risk clusters and it is another fundamental aspect to know about the spatial characteristics of disease. It is important to note that risk clusters depend on the space properties and not on the individual values at areas.

There are two main approaches to the identification of elevated (reduced) risk areas. The first approach identifies discontinuities or step-changes in disease risk, see for example Lu et al. (2007); Lee and Mitchell (2013). The discontinuities identified with these methods highlight boundaries between areas which do not necessarily complete the entire space. Moreover this approach utilizes scanning methods, originally developed by Kulldorff (1997), which, the literature on disease mapping suggests, is not suitable for identification of spatial patterns, see, for example, Anderson et al. (2014).

The second approach, however, identifies areas or a collection of neighboring areas that exhibit significantly elevated (reduced) relative risk compared to geographical proximity. This approach sews neighboring areas and therefore ensures the close boundary as well as covers entire space (examples include Knorr-Held and Raßer (2000); Anderson et al. (2014)).

1.3 Global vs. Local Smoothing

Disease mapping models govern the global or overall estimation and smoothing of underlying risk assuming homogeneous risk across space. The local properties/heterogeneities of risk such as areas with elevated (reduced) risk are generally not counted for. Section 1.2 highlights several surface factors of space that vary by location and are accountable for varying number of disease incidences across space. The variation in disease, caused by local factors, can be associated with area specific risk factors and requires appropriate local smoothing to local geographic mean in order to produce locally precise estimates. The process of such smoothing is called local or regional smoothing.

Additionally, the population is rarely homogeneous across units and population base measures of risk, such as incidence ratio, assume that the risk is constant across areas. An appropriate local smoothing is required before obtaining these heterogeneous population based measures otherwise they may not inform accurately about the local relative risk for the local regions.

1.4 Bayesian Methods

Many modern statistical applications tend to Bayesian approaches for a variety of reasons. The core of Bayesian approaches lies in the widely applied Markov chain Monte Carlo (MCMC) methods which provide fast computational algorithms to sample from posterior densities of parameters (random variables in Bayesian statistics). Recent studies in epidemiology and pharmaceutical industries have also seen a rise of Bayesian applications. These applications are becoming more popular for evidence based public health practices. This rise is supported by readily available Bayesian methods in form of software such as WinBUGS and in the form of packages in R with applications varying from AIDS clinical trials to population genetics to food safety etc. see Cowles (2004); Lawson (2018); Congdon (2020) and references therein.

Disease data on space has two components, location of disease related event and spatial spread or density of disease. The purpose is to utilize additional information about event available in the form of location. However, much of the disease data often has a complex structure, involving hierarchical nesting of affected subjects e.g. subjects classified by a hospital and by their home location. The Bayesian methods naturally adapt to such complex hierarchy or spatially correlated events using conditionally specified hierarchical priors (Congdon (2020).)

Since Besag et al. (1991), risk estimation has been considered in a variety of Bayesian models. These models utilize the spatial characteristics of data in different ways, such as renowned BYM model in Besag et al. (1991) and its modifications Leroux et al. (2000); Stern and Cressie (2000); Dean et al. (2001).

1.5 Novel Approach of Regional Smoothing

Disease maps provide a good visual representation of disease risk, featuring spatial heterogeneities and highlighting units or clusters of high risk. To clearly differentiate areas of low or high risk, we must apply some type of smoothing over regions which are presumably similar to each other. The process of such local smoothing increases our ability to clearly discern clusters in the spatial variation. In this thesis, we propose a novel approach to smooth local spatial units based on their larger regional location.

For the purpose of illustration, we consider the spatial structure of German counties as local spatial units and Federal states of Germany as larger regions to apply state wise adaptive smoothing. We define the incidences, say y_i, $i = 1, 2, \ldots, n$, in each county to follow a binomial density with probability of risk, say π_i and population size N_i. The number of incidences vary from county to county which creates heterogeneities in the density of π_i, there are many factors to this, some of them have covered in Section 1.2. Our objective is to smooth spatial pattern of π_i or y_i by reducing such heterogeneities which vary from region to region. We suppose the π_i is linked to coefficient β_i through fixed

covariate z_i as follows.

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = z_i\beta_i.$$

Suppose the joint density of $\boldsymbol{\beta}$ is a Gaussian density $N(\boldsymbol{\mu}, \mathbf{Q}^{-1})$ with known $\boldsymbol{\mu}$ and spatial correlation matrix $\mathbf{Q}$ (explained in more detail in Chapter 4). Note that $\boldsymbol{\mu}$ can be considered as unknown and a prior can be assigned to it with spatial correlation matrix in the precision. The correlation matrix represents the spatial dependence of y_i's and we utilize this matrix to apply regional smoothing. The correlation matrix is partitioned into sub-matrices, corresponding to regions (Federal states), and smoothing parameters are introduced into the sub-matrices to locally smooth regions. The degree of smoothing is constant within each region, but differs between regions. Chapters 4 and 5 propose univariate and multivariate spatial models of regional smoothing.

Outline of thesis

Chapter 2 explains necessary concepts related to disease mapping. Section 2.1.1 discusses the nature of spatial data and explains areal data as its sub-type. Characteristics of spatial data (spatial dependence and spatial heterogeneity) are explained in Section 2.1.2 and the ways they can be modeled are discussed. Specification of neighbors in space and construction of their relationship graph are explained in Sections 2.1.3 and 2.1.4. Undirected graphs and their properties are discussed in Section 2.2. Section 2.3 defines Gaussian Markov random fields (GMRFs), proper and improper versions, their expression through conditionally autoregressive (CAR) model, their conditional properties and the construction of linear model of coregionalization from multivariate GMRFs.

Chapter 3 provides implementation of the Bayesian methods to disease mapping. Different types of prior information and their inclusion in posterior density along with observed data is discussed. Section 3.5 explains Markov chain Monte Carlo (MCMC) methods and particularly Metropolis-Hastings algorithm to sample from highly structured posterior densities.

Chapters 4 and 5 propose univariate (single disease) and multivariate (multiple diseases) spatial models of regional smoothing. Chapter 4 starts by describing the probability model of univariate data. The density of disease incidences is assumed to be a binomial density. The spatial structure is partitioned into non-overlapping regions and smoothing is applied on each region to produce spatially smoothed data. The prior densities are proposed and auxiliary mixture sampler from Frühwirth-Schnatter et al. (2009) is used to sample from full conditional posterior densities. The general coregionalization framework of MacNab (2016) to build multivariate GMRFs as a linear combination of latent independent univariate GMRFs is adopted in Chapter 5. The smoothing parameters are applied to each of the sample and then combined in the form of joint precision matrix to get multivariate smoothed data. On the basis of Frühwirth-Schnatter et al. (2009), an algorithm is devised to sample from full conditional posterior densities.

Chapter 6 defines clusters in Section 6.1. It reviews the existing approaches of clustering in Section 6.2. Section 6.3 discusses the adopted approach of identification of risk clusters from Anderson et al. (2014). Section 6.4 describes the method to rank identified clusters based on incidence ratios.

Chapters 7 and 8 provide results from simulation studies and real data, respectively. Chapter 9 discusses the implications of proposed method of regional smoothing and further extensions of proposed work.

Chapter 2

Spatial Data and Its Modeling

2.1 Spatial Data and Neighborhood Graph

2.1.1 Types of Spatial Data

Spatial data is the information about a subject that identifies its geographical location and refers to its shape and size. According to Cressie (1993), the spatial data generally fall into three classes:
— point data e.g., locations of Oral cancer incidences;
— continuous data e.g., temperature measurements across a region; and
— areal data e.g., risk of Oral cancer by counties in some area.
The data of our particular interest is areal data, which are based on non-overlapping partition of a spatial structure into contiguous regions. The location of observations is assumed to be fixed, and the observations are associated with some statistical process. The important characteristic of areal data is the relationship of observations in one region to observations in geographical proximity. The main focus in these data is detecting spatial patterns in order to understand the underlying statistical process and. the analysis begins by first defining the neighboring observations, constructing neighborhood graph (or correlation matrix) of them and then assigning weights to spatially close neighbors.

2.1.2 Characteristics of Spatial Data

Some characteristics of spatial data make it hard to use standard statistical procedures on it, such as spatial dependence and spatial heterogeneity (Anselin (2010)). Below we review these characteristics and appropriate methods to incorporate them into analysis.

Spatial heterogeneity

Each observation in spatial data is attached to a unique (*intrinsic*) location, and being unique it has regional differences depending on the type of data, e.g., observations from deprived communities, spatial heterogeneity simply refers to such regional differences. For example, an explanatory variable may prove influential on dependent variable in one location,

but may not in other locations, which is in contrast to the assumption of homogeneous model coefficients (LeSage and Pace (2009); Anselin (2010)).

Spatial heterogeneity can be handled by fixing the structure of model using standard procedures of Econometrics, either through heteroskedastic error variances or through variable regression coefficients (Anselin (2010)).

Spatial dependence

Spatial dependence implies that observations at one location are correlated with the observations belonging to its geographical proximity, which is supported by Tobler's first law of Geography (Tobler (1970)): "everything is related to everything else, but near things are more related than distant things". Spatial dependence thus breaks the conventional hypothesis of independent observations. Spatial dependence is positive when the neighboring areas are relatively similar and negative when the neighboring areas are relatively dissimilar.

The predominance of spatial dependence in spatial data has important implications for statistical analysis. The observations are spatially clustered and therefore contain less information (in the case of positive spatial auto-correlation) than an independent sample of the same size (Anselin (2010)). Spatial dependence can either be considered *substantive* or *nuisance* depending on the objectives of the study. We, however, utilize it in a hierarchical way in form of neighborhood matrix to substantially decrease the computation time in analyses.

There are three main approaches followed in the literature to model spatial dependence. The first approach specifies a particular functional form for the spatial stochastic process generating the observations, and spatial dependence is incorporated through a neighborhood graph (discussed below in Section 2.1.4). The second approach directly models the spatial dependence as a distance decay function. The third approach is commonly used in panel data, where the form of spatial dependence is left unspecified and estimated using non-parametric methods (chapter 14 of Anselin (2001)).

2.1.3 Neighbors in a Spatial Structure

The extent of spatial dependence classifies the regions as neighbors of each other. These are generally those spatial units which lie in the geographic proximity or are spatially close to each other. There are several ways to define this geographic proximity or spatial closeness, such as based on contiguity, general topography of region, distance between regions etc., for further details please refer to Best et al. (1999); Getis (2009); Gelfand et al. (2010).

In case of regular lattices, such as pixels in an image, neighbors can be defined on grid based contiguity, see Besag et al. (1991). In case of irregular lattices, such as administrative divisions of a land, where the layout of regions is not homogeneous and the regions differ in size and shape, a grid based contiguity approach is not possible and a distance base approach may generate lack/excess of connection among contiguous regions due to their varying size. So, in conclusion, each method of defining neighbors varies given the situation of spatial structure. A careful examination of literature suggests that an adjacency based contiguity approach is a common practice to define neighbors on irregular lattices (Wakefield

(2007); Simões and Natário (2016); Martínez-Beneito and Botella-Rocamora (2019)), and we will proceed with this approach.

Consider an irregular lattice partitioned into N non-overlapping regions with y_i as a study variable for region i. The neighbors of i, denoted by $ne(i)$, are defined as those contiguous regions which share a common border or at least one vertex with i. To express i and $j\,(\neq i)$ to be neighbors of each other, $i \sim j$ is written. The properties of spatial neighbors are discussed with the help of undirected graphs in Section 2.2.

2.1.4 Neighborhood Graph

Once it is defined which regions are spatial neighbors of each other, their connections can be displayed graphically, called a neighborhood graph. If the two units are neighbors they are connected on graph, otherwise they are separated.

Consider Figure 2.1a, the structure has five non overlapping areas A, B, C, D and E. Since A is connected with B and D, therefore, these are neighbors of A. Similarly, C has neighbors B, D and E and so on. The neighborhood graph of this structure is given in Figure 2.1b, which displays all such connections with the arrow symbols. Since neighbors are spatially dependent areas, the neighborhood graph exhibits that observations in, for example, A are directly associated (or spatially dependent) with areas B and D. There is another interesting property of this graph, for example, the areas A and E are not connected directly but through D only, in this case, A and E are said to be *conditionally independent* given D. The conditional independence is further explained in Section 2.2.

(a) Irregular lattice with 5 units (b) Neighborhood graph

Figure 2.1: Construction of neighborhood graph from a spatial structure

2.1.5 Weights of Neighborhood Matrix

Neighborhood graph exhibits the extent of spatial dependence of areas in Section 2.1.4, now we demonstrate how a neighborhood graph can be transformed into a matrix form for analysis. The matrix constructed so is called a neighborhood (or precision) matrix. If the two areas are neighbors of each others, the corresponding matrix entry is non zero, otherwise it is zero. The non zero entries are called weights of the matrix and depend on

the definition of neighbors. The weights represent the strength of relationship between different areas. For identification reasons, the weights (say, w_i) are assigned in such a way that they are both row and column standardized i.e., $\sum_j w_{ij} = ne(i)$ and $\sum_i w_{ij} = ne(j)$.

Various types of weighting methods exist in literature (Best et al. (1999), Getis (2009), Chen (2012), chapter 14 of Gelfand et al. (2010)). Contiguity weights are based on contiguous regions, such as bishop or queen contiguity, distance base weights are based on some distance function such as Euclidean or arc distance etc. Typically the weights are fixed but applications exist where they are estimated from data, see for example, Lu et al. (2007).

A common practice in disease mapping is to consider binary adjacency weights to construct neighborhood matrix, usually denoted by $\mathbf{W}$, where $w_{ii} = 0$ i.e., dependence of a unit to itself does not matter, $w_{ij} = 1$ if $i \sim j$ and otherwise $w_{ij} = 0$ (Rue and Held (2005); Gelfand et al. (2010); Simões and Natário (2016); Lawson (2018)).

2.2 Undirected Graphs with their Key Properties

An undirected graph $\mathcal{G}$ consists of two components, a set of nodes $\mathcal{V} = (1, 2, \ldots, N)^T$, and a set of spatially connected edges $\mathcal{E} = \{i, j\}$; $i, j \in \mathcal{V}$, $i \neq j$. If $i \sim j$, then there is an edge from node i to node j and $\{i, j\} \in \mathcal{E}$. If $ne(i)$ denotes the neighbors of i, then i is connected to all its neighbors on $\mathcal{G}$ (Rue and Held (2005)).

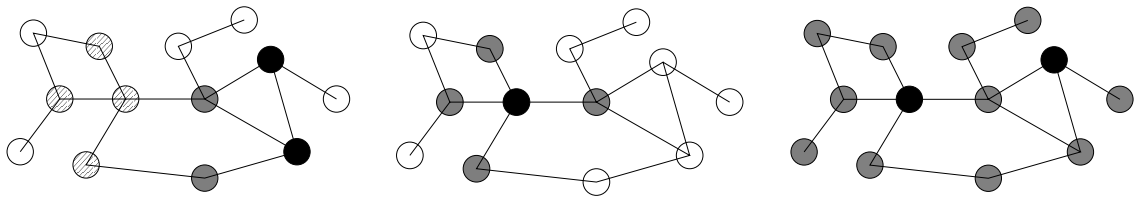
Suppose each node of $\mathcal{G}$ is associated with a random variable, resulting in a set of random variables $\boldsymbol{x} = (x_v)$; $v \in \mathcal{V}$. The variables $\boldsymbol{x}$ on $\mathcal{G}$ satisfy a set of properties and collectively form a Markov random field.

We first present the global Markov property, then a local Markov property and finally a pairwise Markov property is given.

- **Global property:** consider three disjoint sets of variables, A, B and C, such that C separates A and B, then A and B are conditionally independent given C, $x_A \perp x_B \mid x_C$.
- **Local property:** a variable is conditionally independent of all other variables given its neighbors, $x_i \perp x_{-\{i,ne(i)\}} \mid x_{ne(i)}$.
- **Pairwise property:** any two variables not connected spatially become conditionally independent given the rest $\boldsymbol{x}_{-ij} (= \boldsymbol{x}_{-\{i,j\}})$, i.e., $x_i \perp x_j \mid x_{-ij}$.

Consider Figure 2.2a, borrowed from Rue and Held (2005), the black and striped nodes satisfy global Markov property of independence. They are independent given the separating grey nodes. An immediate deduction of global property is the local property in Figure 2.2b, where black node is dependent on its neighbors (grey nodes) but independent of white nodes given its neighbors . From the local Markov property, it is easy to see that the two black nodes in Figure 2.2c are independent given all other nodes.

Note that if the variables on $\mathcal{G}$ satisfy global Markov property then they also satisfy the other two properties but the converse is generally not true and holds only for strictly positive densities. The three Markov properties are equivalent when every property yields



(a) Global independence (b) Local independence (c) Pairwise independence

Figure 2.2: Properties of undirected graphs

the same conditional independence restrictions for variables. Also worth noting is that the pairwise property is weaker than local property which in turn is weaker than the global Markov property (Rue and Held (2005); Matúš (1992)).

2.3 Gaussian Markov Random Fields

Before properly defining a GMRF, we would like to introduce a multivariate Gaussian (or normal) density. The definition of a GMRF is given in Section 2.3.2 with the help of undirected graphs. Their conditional properties are given in Section 2.3.3 and Section 2.3.4 lays the foundation to construct multivariate GMRFs.

2.3.1 Multivariate Gaussian Density

A Gaussian density, also known as a normal density, is a continuous variable density. It has found ways to numerous applications in natural and social sciences due to its nice mathematical properties for simpler theoretical analyses (see for example, Tong (2012) for its properties and applications).

The multivariate Gaussian density of random variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)^T$ with some mean $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)^T$ and a symmetric positive definite precision (inverse of variance) matrix $\mathbf{Q} = (Q_{ij})_{N\times N}$ is defined below.

$$h(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathbf{Q}) = (2\pi)^{-N/2}|\mathbf{Q}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right\}.$$

2.3.2 Proper and Improper GMRFs

Suppose a set of random variables $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)^T$ forms a Markov random field over $\mathcal{G}$ and they follow a multivariate Gaussian density then $\boldsymbol{x}$ is called a GMRF with respect to $\mathcal{G}$. The density of $\boldsymbol{x}$ with Markov restrictions is written as

$$h(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathbf{Q}) = (2\pi)^{-N/2}|\mathbf{Q}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{2.1}$$

$$\text{and} \qquad Q_{ij} \neq 0 \iff \{i,j\} \in \mathcal{E} \;\forall\; i \neq j.$$

The Markov condition states that the element Q_{ij} is non-zero when $i \sim j$ or vice versa.

Intrinsic or improper GMRF

A special case of a GMRF arises when the precision matrix $\mathbf{Q}$ does not have full rank or some of its eigenvalues are zero, then it is named as *intrinsic* GMRF (iGMRF). Suppose $q\,(>0)$ eigenvalues are zero, then the density of an iGMRF $\boldsymbol{x}$ is written as

$$p(\boldsymbol{x} \mid \boldsymbol{\mu}, \mathbf{Q}) = (2\pi)^{-\frac{N-q}{2}} \left(|\mathbf{Q}|^*\right)^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu})\right\}, \tag{2.2}$$

$$\text{and} \qquad Q_{ij} \neq 0 \iff \{i,j\} \in \mathcal{E} \;\forall\; i \neq j.$$

The rank of $\mathbf{Q}$ is now $N-q$, and $|\mathbf{Q}|^*$ is obtained by the product of only non-zero $(N-q)$ eigenvalues. The rank deficient q is the number of islands on $\mathcal{G}$ (an island is a continuous chain of connected nodes, see chapter 5 of Hodges (2013)). The rank deficiency is caused by the hidden linear constraints $\sum_{j:j\sim i} Q_{ij} = 0 \;\forall\, i$, which leads $\mathbf{1}$ to be the eigenvector of $\mathbf{Q}$ corresponding to each 0 eigenvalue, so that $\mathbf{Q1} = \mathbf{0}$.

Specification through full conditionals

Besides specifying a GMRF through a mean vector and a precision matrix, Besag (1974) specifies each random variable separately through its full conditional, these models are known as *conditionally autoregressive* (CAR) models. Specifically, the CAR formulation replaces (2.2) with a collection of univariate Gaussian conditional densities for each x_i. The full conditional density of x_i given its neighbors (with zero mean) is given by

$$E(x_i|\boldsymbol{x}_{-i}) = \frac{1}{ne(i)} \sum_{j:j\sim i, j\neq i} x_j \qquad \text{and} \qquad Prec(x_i|\boldsymbol{x}_{-i}) = ne(i)\,\kappa_i, \quad \kappa > 0. \tag{2.3}$$

This particular formulation assumes strong positive spatial dependence among variables i.e., the conditional precision directly depends on the number of neighbors ($ne(\cdot)$), higher the number of neighbors, higher the precision, this is intuitive because more information about neighbors lead to less uncertainty about a variable.

Besag (1974) indicates that the mean and precision in (2.3) uniquely identify the following joint density

$$\pi(\boldsymbol{x}) \propto \kappa^{\frac{N-q}{2}} \exp\left\{-\frac{\kappa}{2} \sum_{j:j\sim i, j\neq i} (x_i - x_j)^2\right\}, \tag{2.4}$$

$$\pi(\boldsymbol{x}) \propto \kappa^{\frac{N-q}{2}} \left(|\mathbf{Q}|^*\right)^{1/2} \exp\left\{-\frac{\kappa}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}\right\}, \tag{2.5}$$

with $Q_{ii} = ne(i)$, $Q_{ij} = -1$ when $i \sim j$ and zero otherwise. The formulation (2.4) reveals that the joint density is non identifiable or invariant to addition of any constant to $\boldsymbol{x}$. The density is improper too since $\mathbf{Q}$ is singular by deficiency q and the impropriety is usually addressed in practice by making sure that $\mathbf{Q}$ is diagonally dominant i.e., each diagonal element Q_{ii} is greater than the sum of its neighboring entries (Banerjee et al. (2014)).

2.3.3 Conditional Properties of GMRFs

Suppose the variables $\boldsymbol{x}$ are partitioned into two parts such that $\boldsymbol{x} = \{\boldsymbol{x}_A, \boldsymbol{x}_B\}$ and the interest is to estimate the subset A. Also partition the mean and precision matrix as well, such that

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{pmatrix} \qquad \text{and} \qquad \mathbf{Q} = \begin{pmatrix} \mathbf{Q}_{AA} & \mathbf{Q}_{AB} \\ \mathbf{Q}_{AB} & \mathbf{Q}_{BB} \end{pmatrix}.$$

The conditional density of $\boldsymbol{x_A}$ given $\boldsymbol{x_B}$ is also a GMRF on sub-graph $\mathcal{G}^A = (\mathcal{V}^A, \mathcal{E}^A)$. The sub-graph is easily constructed by removing all those nodes which do not belong to A and removing all those edges whose at least one node does not belong to A. The mean and precision matrix of $\boldsymbol{x}_A|\boldsymbol{x}_B$ on $\mathcal{G}^A$ are

$$E(\boldsymbol{x}_A|\boldsymbol{x}_B) = \boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A - Q_{AA}^{-1} Q_{AB}(\boldsymbol{x}_B - \boldsymbol{\mu}_B),$$
$$\text{and} \qquad Prec(\boldsymbol{x}_A|\boldsymbol{x}_B) = Q_{A|B} = Q_{AA}.$$

It is easy to see the effect of spatial dependence in $\boldsymbol{x}$, the conditional mean depends only on the values which belong to A or are its direct neighbors. The precision matrix is just the principal matrix of partitioned precision matrix which is obtained without additional computation.

2.3.4 Coregionalized Framework of Multivariate GMRFs

Suppose that more than one diseases are simultaneously being observed on the same spatial structure. A first choice is to use a separate GMRF model for each disease, but since the spatial structure is same, there may exist spatial correlation among diseases if they share some common risk factors or are linked by aetiology. Moreover, the presence of one disease may affect/influence the presence of other disease(s). Fitting a separate model to each of these diseases may met with such difficulties as mentioned in Jin et al. (2007). Many multivariate spatial models that allow modeling of such dependence among diseases are proposed in literature, some of which can be studied from Jin et al. (2007), Martínez-Beneito (2013), Botella-Rocamora et al. (2015) and MacNab (2016).

MacNab (2016) presents a general approach of coregionalization to express multivariate GMRFs as a linear combination of latent independent univariate GMRFs. They develop and discuss different classes of coregionalized multivariate GMRF for different choices of cross-variable local interaction parameter $\mathbf{C}$. The framework proposed by MacNab (2016) is adopted in this thesis.

Let $\{x_{ij} : i = 1, 2, \ldots, N; j = 1, 2, \ldots, J\}$ denote J random variables associated with J diseases, each of N observations, on the same spatial structure. Let $\mathbf{X} = (x_{ij})_{N \times J}$ be a matrix of variables x_{ij}, such that its rows represent the spatial domain and its columns represent the variable domain. Suppose variance-covariance matrix of $\mathbf{X}$ is $\boldsymbol{\Sigma}_{J \times J}$ and $\mathbf{A}\mathbf{A}^T$ be its Cholesky factorization. Let $\boldsymbol{\eta} = (\eta_{ij})_{N \times J}$ be another matrix of latent variables η_{ij}, whose rows and columns also represent spatial and variable domains, respectively.

Suppose the variable domain of $\boldsymbol{\eta}$ represents independent GMRFs, then a coregionalization framework can be established by representing variable domain of $\mathbf{X}$ as a linear combination of variable domain in $\boldsymbol{\eta}$ as follows

$$\mathbf{X} = \boldsymbol{\eta}\mathbf{A}^T.$$

Let $\mathbf{P} = (P_{il})_{N\times N}$ be a positive definite precision matrix constructed from the neighborhood structure in such a way that its diagonal entries are zero i.e., $P_{ii} = 0$, and $P_{il} = 1$ if $i \sim l$, otherwise $P_{il} = 0$. The joint precision matrix for J random variables is defined as

$$\begin{aligned} \boldsymbol{\Omega}_{\mathbf{X}} &= (\boldsymbol{\Lambda} \otimes \mathbf{I}_N)\,(\mathbf{I}_{N\times J} - \mathbf{C} \otimes \mathbf{P})\,(\boldsymbol{\Lambda} \otimes \mathbf{I}_N)^T, \\ \text{or} \qquad &= \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_N - \boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\Lambda}^T \otimes \mathbf{P}, \\ \text{where} \quad P_{ii} = 0, \qquad P_{il} = 1 \quad &\text{if} \quad i \sim l \qquad \text{and} \qquad P_{il} = 0 \quad \text{elsewhere.} \end{aligned}$$

The matrix $\boldsymbol{\Lambda} = \left(\mathbf{A}^{-1}\right)^T$ and $\mathbf{C}$ is a $J \times J$ diagonal matrix of cross-variable local interaction parameters.

Chapter 3

Bayesian Implementation in Disease Mapping

The Bayesian approach provides a coherent framework for combining complex data models and any unknowns (parameters) as random variables through external knowledge. In addition to specifying the observed data model, this approach assigns a prior density to quantify unknown parameters.

Bayesian statistics is a combination of likelihood and prior information. A likelihood is considered as a joint density describing the dependence of parameter(s) on observed data while the prior density translates external knowledge about unknowns in probabilistic form. Below, we present likelihood in Section 3.1 and different forms of prior densities in Section 3.2. The construction of posterior density is given in Section 3.3. Bayesian modeling of hierarchy is reviewed in Section 3.4. The numerical approximation methods to sample from posterior densities are provided in Section 3.5.

3.1 Likelihood Function

The likelihood principle assumes that the individual sample values given parameter(s) are independent and thus can be multiplied together to get a likelihood function. Consider $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$, be a sample from some probability density defined by $\pi(\boldsymbol{x}|\theta)$ with parameter θ, the likelihood for this data is defined as

$$L(\boldsymbol{x}|\theta) = \prod_{i=1}^{N} \pi(x_i|\theta).$$

Since each sample observation represents the individual contribution so when combined in the from of likelihood function they are assumed to be conditionally independent. This assumption of conditional independence is fundamental to many disease mapping applications.

3.2 Prior Density

The Bayesian deduction requires appropriate choice of priors for the parameters. We discuss various types of prior information and their selection under certain circumstances.

A prior density is a central part of any Bayesian analysis and represents information about unknown parameter (random variable in Bayesian) before the observed data are examined. One may see prior density as additional "data" to improve estimation of parameters or sometimes for their identification. A prior density should include all plausible values of the parameter because if the prior assigns a zero probability to some value of parameter then the posterior will as well (Berger (2013)).

Many things are considered when choosing a prior density. Some of them include the following.

— Informative or subjective prior: some information about the parameters are available and we wish to combine that in the analysis.

— Non-informative or objective prior: there is no prior knowledge about the parameters and we wish not to influence the analysis by some subjective choice.

— Conjugate prior: this prior is chosen for mathematical convenience (mentioned in Agarwal and Daumé (2010) among many others). Since the posterior density has the same distributional form as the prior density, it becomes easy to analyze the posterior density as the characteristics of prior density are already known.

3.2.1 Non-informative Priors

It is a common perception that Bayesian analysis is primarily a subjective theory. This perception is true neither historically nor in practice. Bayes (1763) and Laplace (1812) performed Bayesian analysis using a constant prior density for the unknown parameters. This approach was then called inverse probability and was highly influential in the early part of the 19th century. Jeffreys (1961) made significant refinements to end criticism of the use of constant prior density.

Non-informative priors generally make flat preferences over the range of variable yielding close to uniform preferences for all the values. They are sometimes referred as objective, neutral, flat or reference priors (Lawson (2018)). These priors have little impact compared to likelihood of observed data. A simple example of such a prior is a uniform density which assigns equal probability to all possible values of parameter. For a continuous parameter λ of N possible values between a and b, the uniform prior density is written as

$$\pi(\lambda) = \frac{1}{b-a}.$$

The parameters of prior density, a and b for λ, are called hyperparameters.

The choice of non-informative priors is made with some general understanding of the range and behavior of the parameter. For example, the variance (or precision) of Gaussian density is strictly positive, therefore, its prior density must have range over positive real line, such as a gamma density or a uniform density over some positive range. For instance,

a $Gamma(0.001, 0.001)$ with small precision (0.001) is relatively flat over a large range of values as compared to $Gamma(0.1, 0.1)$ with precision (0.1). When the parameters have an infinite range, such as regression parameters, a Gaussian density with very small precision is seen in practice in many applications (Lawson (2018)). Non-informative priors can be proper or improper. A proper prior is defined above for λ while a prior of $\pi(\theta)$ with range $\{-\infty, \infty\}$ is said to be improper, if

$$\int_{-\infty}^{\infty} \pi(\theta) \mathrm{d}\theta = \infty.$$

i.e., a prior density is improper if its normalizing constant is not finite. It is often the case that an improper prior yields a proper posterior, such as iGMRFs discussed in Section 2.3 are improper priors, they define value of each x_i relative to the values of the others i.e., by contrasts between pairs $x_i - x_j$, but they generate proper posterior densities (see chapter 14 of Gelfand et al. (2010)).

3.2.2 Informative Priors

In many problems, the use of informative prior is clearly essential, and in other cases it is readily available. No doubt, few experts would argue against if the subjective information in form of informative prior produces accurate results. A subjective prior is constructed on probability bounds or chosen from well known probability densities based on the range of parameter e.g. a Gaussian density for real numbers or a Poisson density for non-negative integers.

Identifiability is another issue related to the ability to distinguish between parameters (Bernardo and Smith (2009)). If a restricted range is assumed for variables to be identified then the prior densities should be assigned in an informative way to support such assumptions. Ultimately if the separation of identification of variables is not apparent from the likelihood then the identification comes only from prior assignment. An example of identification or separation in disease mapping arises when a linear predictor defines the variation as a sum of two random components, one with small variability and the other with greater variability.

3.2.3 Conjugate Priors

A prior is said to be a conjugate prior if the prior and posterior densities are from the same family. For example, if we are sampling from a Bernoulli density having parameter p, the conjugate prior for p is the beta density; it follows that the posterior density of p is also a beta density.

The development of conjugate priors is inspired by a desire for algebraic convenience (see for example Agarwal and Daumé (2010)). Conjugate priors always guarantee a proper posterior density, however, simple conjugacy is less likely in hierarchical models of disease mapping (Lawson (2018)).

3.3 Posterior Density

Likelihood function and prior densities are two forms of information about the parameter; likelihood expresses the information of sample values while the prior density translates prior believes or assumptions. The Bayes rule combines likelihood with prior density to yield a posterior density, which is then used to carry out all inferences about parameter. A combination of prior belief and likelihood becomes posterior belief and this posterior belief may also be used as prior information in the future. Analytically, the posterior density of a parameter μ given data $\boldsymbol{x}$ is

$$\pi(\mu \mid \boldsymbol{x}) = \frac{1}{C} L(\boldsymbol{x} \mid \mu)\pi(\mu), \tag{3.1}$$
$$\text{and} \qquad C = \int_{\mu} L(\boldsymbol{x} \mid \mu)\pi(\mu)\mathrm{d}\mu.$$

$\pi(\mu)$ is the prior density of μ and C is constant of proportionality. It is a common practice to omit C and posterior density is written in proportionality form.

A simple example in context of disease mapping is where the disease incidences, $\{x_i : i = 1, 2, \ldots, n\}$, are generated from a Bernoulli density with probability p. The prior density of p is assumed to follow a beta density with known hyperparameters a and b. The posterior density turns out to be a beta density as follows.

$$\begin{aligned} \pi(x_i|p) &\sim Bernoulli(p) \\ \pi(p) &\sim p^{a-1}(1-p)^{b-1} \\ \pi(p|\boldsymbol{x}) &\sim p^{\alpha-1}(1-p)^{\beta-1}. \end{aligned}$$

The parameters of posterior density are $\alpha = a + \sum_i x_i$ and $\beta = b + n - \sum_i x_i$.

3.4 Bayesian Hierarchical Modeling

Suppose $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}$ denote the number of persons affected by some disease on N locations and θ_i be the risk of location i. The data follows some density $\pi(x_i|\theta_i)$ and prior density of θ_i is denoted by $\pi(\theta_i|\lambda)$ where λ controls how the risk varies across locations. If λ is known, the posterior density of $\boldsymbol{\theta}$ is given by

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \lambda) \propto L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\lambda). \tag{3.2}$$

The parameter λ can be fixed by some constant value but since the fundamental feature of Bayesian methodology is that the values of parameters could arise from densities, therefore it naturally leads to the use of models within hierarchies. Banerjee et al. (2014) and Lawson (2018) discuss hierarchical models commonly found in disease mapping. In above example, λ in practice is generally not known, therefore, a prior density is assigned at second stage to govern its behavior. The prior densities assigned to hyperparameters are known as hyperpriors. If the hyperprior is denoted by $\pi(\boldsymbol{\lambda})$, then the posterior (3.2) can be written as

$$\pi(\boldsymbol{\theta}|\boldsymbol{x}, \boldsymbol{\lambda}) \propto L(\boldsymbol{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}|\boldsymbol{\lambda})\pi(\boldsymbol{\lambda}). \tag{3.3}$$

3.5 Markov Chain Monte Carlo

The results from the product of prior density and the likelihood are not always very generous and often lead to complex posterior densities. Markov chain Monte Carlo (MCMC) techniques come handy in such situations. The MCMC algorithm generates dependent realizations from the posterior density by running a Markov chain which has the posterior as its equilibrium density.

The MCMC algorithm has seen dramatic growth in applications of Bayesian methods after Gelfand and Smith (1990). Its implementation is easy even in complex structures such as hierarchical models. An excellent monograph on the implementation of MCMC in practice is Brooks et al. (2011).

The Markov chain is usually constructed by sampling from the conditional density of the selected variable(s) or from a density close to it, given the rest. For this to be computationally tractable, we often need a Markov property in the posterior, implying that density of one variable conditioned on the rest i.e., the conditional posterior does not depend on all other variables but only those in the Markov neighborhood.

3.5.1 Metropolis-Hastings Algorithm

The Metropolis-Hastings algorithm is used in this thesis which is the basis of most MCMC algorithms. The basic idea is to propose θ' based on current value θ with some probability function $g(\theta)$, where $g(\cdot)$ is called the proposal density or kernel. The Metropolis-Hastings algorithm is quite a flexible algorithm in the sense that any kind of proposal can be chosen under certain conditions, the reader can consult Brooks et al. (2011) and Murphy (2012) for details.

Different choices exist for the proposal kernels, among them are the two forms of proposals which are extensively used in practice; a proposal kernel based on Gaussian or uniform density centered on the present value, referred to as a random walk Metropolis algorithm. The other type of proposal kernel is that in which the proposed value does not depend on the present value, referred to as an independent sampler.

After generating proposal from proposal kernel, the next step is to decide whether to accept the proposed value on the basis of some formula, known as acceptance probability. If the proposed value is accepted then we repeat the process of generating proposals considering the accepted value as present value.

Suppose the density of θ is $\pi(\theta)$ and we are interested in generating Metropolis-Hastings based Markov chain with $g(\theta)$ as a symmetric proposal kernel, then acceptance probability is given by

$$\alpha = \min\left(1, \frac{\pi(\theta')\, g(\theta)}{\pi(\theta)\, g(\theta')}\right).$$

A common practice is to occasionally allow the "downhill" moves by rejecting the proposed values between 25% to 40% of the times. This is done in an attempt to explore the entire space of θ, more on this can be studied from Brooks et al. (2011) and Murphy (2012).

A special case of Metropolis-Hastings algorithm is the Gibbs sampler, in which each conditional posterior density is a proposal density and the acceptance probability is always one, because we always accept the proposal from conditional density. The advantage of using this sampler is that it is easy to sample from a conditional density instead of a marginal density because we only need to integrate (sum, in discrete case) over the selective dependent variables to find conditional density.

3.5.2 Auxiliary Variable Approach for Binomial Data

Consider a set of n variables $\{y_1, y_2, \ldots, y_n\}$ where each y_i follows a binomial density with population N_i and probability of success p_i. The probability p_i depends on unknown coefficient β_i through some covariate x_i (continuous or discrete) of logit function as $p_i = \{1+\exp(-x_i\beta)\}^{-1}$. Let $\pi(\boldsymbol{\beta})$ represents the prior density of $\boldsymbol{\beta}$ and the interest is to estimate its posterior density $\pi(\boldsymbol{\beta}|\boldsymbol{y})$. The expression for posterior density is given by

$$\pi(\boldsymbol{\beta} \mid \boldsymbol{y}) = \frac{1}{C}\pi(\boldsymbol{\beta}) \prod_{i=1}^{n} \left\{ \frac{1}{1+\exp(-\boldsymbol{x}^T\boldsymbol{\beta})} \right\}^{i} \left\{ 1 - \frac{1}{1+\exp(-\boldsymbol{x}^T\boldsymbol{\beta})} \right\}^{N_i - i}. \tag{3.4}$$

C is the constant of proportionality as in (3.1), whose $\int$ being replaced by $\sum$. The posterior density is intractable and direct sampling from this density is not possible (Albert and Chib (1993); Frühwirth-Schnatter et al. (2009)). Moreover, the coefficient β_i is not linearly related with p_i, which causes non-linearity of β_i and non-normality of logit errors. Frühwirth-Schnatter et al. (2009) provide an auxiliary mixture sampling approach to handle these problems using the idea of data augmentation. The advantage of their approach is that the posterior density of $\boldsymbol{\beta}$ has an explicit form conditional on some latent variables as well as the MCMC procedure is simplified and Gibbs sampling can be used to obtain posterior sample for $\boldsymbol{\beta}$. The details for our specific case is given in Section 4.4.1.

Chapter 4

Regionally Smoothed Univariate Spatial Model

This chapter starts by describing the study data and its density model. Section 4.2 explains the novel approach of incorporating smoothing parameters into neighborhood matrix. The prior densities of $\boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{w}$, regional smoothing parameters $\boldsymbol{r}$ and global smoothing parameter s are presented in Section 4.3. The full conditional posterior densities obtained from joint posterior density are given in Section 4.4 along with the MCMC algorithm. Explicit expressions for posterior densities of smoothing parameters could not be derived, therefore, suitable proposal kernels along with their acceptance probabilities are provided in Section 4.4.2.

4.1 Data Structure

Consider a spatial structure partitioned into N non-overlapping areas or units. Let y_i be the number of subjects affected by some acute disease in i^{th} spatial unit having population N_i, such that $\boldsymbol{y} = (y_1, y_2, \ldots, y_N)^T$ forms a sequence of conditionally independent observations. Consider a binomial model for y_i with parameters N_i and $\pi_i\,(0 < \pi_i < 1)$, where π_i is the probability of risk. Assume that π_i depends on covariate z_i through coefficient β_i.

$$\begin{aligned} y_i \mid N_i, \pi_i &\sim Bin(N_i, \pi_i), \\ \text{and} \quad \ln\left(\frac{\pi_i}{1-\pi_i}\right) &= \ln \lambda_i(\beta_i) = z_i\beta_i, \end{aligned} \tag{4.1}$$

where $Bin(\cdot, \cdot)$ denotes a binomial density.

Suppose $z_i = 1$. The coefficients $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_N)^T$ follow a multivariate Gaussian density with some mean $\boldsymbol{x}$ and precision matrix $\boldsymbol{w}^T\mathbf{I}_N : \boldsymbol{w} > \mathbf{0}$ and $\mathbf{I}_N$ is identity matrix of order $N \times N$. In next stage, $\boldsymbol{x}$ follow an iGMRF (details on iGMRF in Section 2.3) with parameters $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q}$ (based on neighborhood).

4.2 Regionalized Smoothing of Spatial Structure

In this section, we explain the novel approach of smoothing a spatial structure by dividing it into various non-overlapping regions. Suppose the structure has N units and each one of them is spatially categorized into one of non-overlapping M regions. We introduce M smoothing parameters, namely $\boldsymbol{r} = (r_1, r_2, \ldots, r_M)^T$, to smooth each region locally. Additionally, a global smoothing parameter is also applied and is denoted by s.

The application of global smoothing parameter to neighborhood matrix $\mathbf{Q}$ is straight forward, just multiply s with its each entry to produce $s \cdot \mathbf{Q}$. To apply $\boldsymbol{r}$, first split $\mathbf{Q}$ into M block-diagonal sub-matrices, one sub-matrix for each region. Then every regional smoothing parameter is multiplied with its corresponding regional sub-matrix, i.e., r_1 is multiplied with $\mathbf{Q}_1$ and so on, then regionally smoothed matrix $\mathbf{Q}_{\boldsymbol{r}}$ is constructed as

$$\mathbf{Q}_{\boldsymbol{r}} = \begin{pmatrix} r_1 \cdot \mathbf{Q}_1 & . & . & . \\ . & r_2 \cdot \mathbf{Q}_2 & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & r_M \cdot \mathbf{Q}_M \end{pmatrix}.$$

The complete precision matrix after applying regional and global smoothing is defined as $\mathbf{Q}_{\boldsymbol{r}s} = \mathbf{Q}_{\boldsymbol{r}} + s \cdot \mathbf{Q}$.

4.3 Prior Assumptions

The subjective prior densities are assigned to $\boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{r}$ and s for Bayesian analysis.
Prior Density of $\boldsymbol{\beta}$: The coefficients $\boldsymbol{\beta} = \{\beta_1, \beta_2, \ldots, \beta_N\}^T$ follow a multivariate Gaussian density with mean $\boldsymbol{x}$ and precision matrix $\boldsymbol{w}^T\mathbf{I}_N$, as stated earlier.

$$p(\boldsymbol{\beta} \mid \boldsymbol{x}, \boldsymbol{w}) \propto |\boldsymbol{w}^T\mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{x})^T\boldsymbol{w}^T\mathbf{I}_N(\boldsymbol{\beta} - \boldsymbol{x})\right\}; -\infty < \boldsymbol{\beta} < \infty. \tag{4.2}$$

Prior Density of $\boldsymbol{x}$: The prior density of $\boldsymbol{x} = \{x_1, x_2, \ldots, x_N\}^T$ is an iGMRF with parameters $\boldsymbol{\mu} = \mathbf{0}$ and spatially smoothed precision matrix $\mathbf{Q}_{\boldsymbol{r}s}$.

$$p(\boldsymbol{x} \mid \boldsymbol{r}, s) \propto (|\mathbf{Q}_{\boldsymbol{r}s}|^*)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\mathbf{Q}_{\boldsymbol{r}s}\boldsymbol{x}\right); -\infty < \boldsymbol{x} < \infty. \tag{4.3}$$

Prior Density of w_i : The parameter $w_i : i = (1, 2, \ldots N)^T$ follows a gamma density with common and fixed parameters $a_w > 0$ and $b_w > 0$.

$$p(w_i) \propto w_i^{a_w - 1} \exp(-b_w w_i); \qquad w_i > 0. \tag{4.4}$$

Prior Densities of $\boldsymbol{r}$ *and* s : Independent prior gamma densities are assumed for smoothing parameters. Let $a_{r_k} > 0$ and $b_{r_k} > 0$ be hyperparameters for regional smoothing parameter $r_k : k = (1, 2, \ldots M)^T$. The prior density of r_k is given by

$$p(r_k) \propto r_k^{a_{r_k} - 1} \exp(-b_{r_k} r_k); \qquad r_k > 0. \tag{4.5}$$

Let $a_s > 0$ and $b_s > 0$ be hyperparameters for global smoothing parameter s.

$$p(s) \propto s^{a_s-1} \exp(-b_s s); \qquad s > 0. \tag{4.6}$$

4.4 Implementation Through MCMC Algorithm

4.4.1 Auxiliary Variable Approach

The auxiliary sampler discussed in Section 3.5.2 introduces a latent variable y_i^* for each y_i to facilitate in MCMC via Gibbs sampler. The relation between $\boldsymbol{y}$ and $\boldsymbol{\beta}$ is given in (4.1), while $\boldsymbol{y}^* = (y_1^*, y_2^* \ldots, y_N^*)^T$ and $\boldsymbol{\beta}$ are connected through $\lambda_i(\boldsymbol{\beta}_i)$ in the following model.

$$y_i^* = \ln \lambda_i(\boldsymbol{\beta}_i) + \mu_{d_i}(N_i) + \epsilon_i, \qquad \epsilon_i \mid d_i \sim \mathcal{MIXN}(0, \sigma_{d_i}^2(N_i)). \tag{4.7}$$

The actual density of $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, \ldots, \epsilon_N)^T$ is a negative log-gamma density but it is approximated by a mixture of normal densities, $\mathcal{MIXN}(\cdot, \cdot)$. The variable d_i has mean $\mu_{d_i}(N_i)$ and variance $\sigma_{d_i}^2(N_i)$, both depend on population size of i^{th} spatial unit. The $\boldsymbol{y}^*$ follow a multivariate Gaussian density with mean $\boldsymbol{\beta} + \boldsymbol{\mu}_d$ and precision matrix $\left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N$.

$$p(\boldsymbol{y}^* \mid \boldsymbol{\beta}) \propto |\left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}^* - \boldsymbol{\beta} - \boldsymbol{\mu}_d)^T \left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N (\boldsymbol{y}^* - \boldsymbol{\beta} - \boldsymbol{\mu}_d)\right\}; \\ -\infty < \boldsymbol{y}^* < \infty. \tag{4.8}$$

4.4.2 MCMC Algorithm

The posterior density is derived by multiplying data (4.8) and priors (4.2) to (4.6). The joint posterior density of $\boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{r}, s$ given $\boldsymbol{y}^*$ is:

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{x}, \boldsymbol{w}, \boldsymbol{r}, s \mid \boldsymbol{y}^*) \propto\ & |\boldsymbol{w}^T \mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{x})^T \boldsymbol{w}^T \mathbf{I}_N (\boldsymbol{\beta} - \boldsymbol{x})\right\} \\ & \times |\left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}^* - \boldsymbol{\beta} - \boldsymbol{\mu}_d)^T \left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N (\boldsymbol{y}^* - \boldsymbol{\beta} - \boldsymbol{\mu}_d)\right\} \\ & \times |\mathbf{Q}_{\boldsymbol{r}s}|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T \mathbf{Q}_{\boldsymbol{r}s} \boldsymbol{x}\right) \prod_{i=1}^{N} w_i^{a_w - 1} \exp(-b_w w_i) \\ & \times \prod_{k=1}^{M} r_k^{a_{r_k} - 1} \exp(-b_{r_k} r_k)\, s^{a_s - 1} \exp(-b_s s). \end{aligned} \tag{4.9}$$

Posterior estimates are obtained by following the procedure: generate initial values of $\boldsymbol{x}$ from prior density (4.3) by assuming initial values of $\boldsymbol{r}$ and s in $\mathbf{Q}_{\boldsymbol{r}s}$. Obtain probability of disease risk π_i using incidences y_i and population N_i. Use π_i to obtain initial values of β_i from (4.1) and y_i^* from step 4 below. Initial values of w_i are generated using β_i and x_i. With these initial values, the following algorithm provides an MCMC chain for the posterior estimates.

1. Sample $\boldsymbol{\beta}$ conditional on $\boldsymbol{y}^*, \boldsymbol{x}$, $\boldsymbol{w}$ and $\boldsymbol{d}$. The full conditional density of $\boldsymbol{\beta}$ is a multivariate Gaussian density with $\boldsymbol{\mu_\beta} = (\boldsymbol{\sigma}_d^2\boldsymbol{w}+1)^{-1}(\boldsymbol{\sigma}_d^2\boldsymbol{x}\boldsymbol{w}+\boldsymbol{y}^*-\boldsymbol{\mu}_d)$ and precision matrix $\boldsymbol{\Omega_\beta} = \boldsymbol{\sigma}_d^{-2}(\boldsymbol{\sigma}_d^2\boldsymbol{w}+1)$.

$$
\begin{aligned}
p(\boldsymbol{\beta} \mid \boldsymbol{y}^*, \boldsymbol{x}, \boldsymbol{w}) &\propto |\boldsymbol{w}^T\mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{x})^T\boldsymbol{w}^T\mathbf{I}_N(\boldsymbol{\beta}-\boldsymbol{x})\right\} \\
&\quad \times |\left(\boldsymbol{\sigma}_d^{-2}\right)^T\mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{y}^*-\boldsymbol{\beta}-\boldsymbol{\mu}_d)^T\left(\boldsymbol{\sigma}_d^{-2}\right)^T\mathbf{I}_N(\boldsymbol{y}^*-\boldsymbol{\beta}-\boldsymbol{\mu}_d)\right\} \\
p(\boldsymbol{\beta} \mid \boldsymbol{y}^*, \boldsymbol{x}, \boldsymbol{w}) &\propto |\boldsymbol{\Omega_\beta}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{\mu_\beta})^T\boldsymbol{\Omega_\beta}(\boldsymbol{\beta}-\boldsymbol{\mu_\beta})\right\}; -\infty < \boldsymbol{\beta} < \infty, \quad (4.10)
\end{aligned}
$$

2. Sample $\boldsymbol{x}$ conditional on $\boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{r}$ and s. The full conditional density of $\boldsymbol{x}$ is a GMRF with $\boldsymbol{\mu_x} = \boldsymbol{w}^T\boldsymbol{\Omega}_{\boldsymbol{x}}^{-1}\boldsymbol{\beta}$ and precision structure $\boldsymbol{\Omega_x} = \mathbf{Q}_{\boldsymbol{r}s} + \boldsymbol{w}^T\mathbf{I}_N$.

$$
\begin{aligned}
p(\boldsymbol{x} \mid \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{r}, s) &\propto |\mathbf{Q}_{\boldsymbol{r}s}|^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^T\mathbf{Q}_{\boldsymbol{r}s}\boldsymbol{x}\right) \\
&\quad \times |\boldsymbol{w}^T\mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{x})^T\boldsymbol{w}^T\mathbf{I}_N(\boldsymbol{\beta}-\boldsymbol{x})\right\} \\
p(\boldsymbol{x} \mid \boldsymbol{\beta}, \boldsymbol{w}, \boldsymbol{r}, s) &\propto |\boldsymbol{\Omega_x}|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu_x})^T\boldsymbol{\Omega_x}(\boldsymbol{x}-\boldsymbol{\mu_x})\right\}; -\infty < \boldsymbol{x} < \infty, \quad (4.11)
\end{aligned}
$$

3. Generate $w_i\,;\, i = (1, 2, \ldots, N)^T$ from full conditional gamma density with parameters $a_w + 0.5$ and $b_w + 0.5(\beta_i - x_i)^2$.

$$
\begin{aligned}
p(w_i \mid \boldsymbol{\beta}, \boldsymbol{x}) &\propto w_i^{a_w-1} \exp(-b_w w_i) w_i^{1/2} \exp\left\{-\frac{1}{2}(\beta_i - x_i)^2 w_i\right\} \\
p(w_i \mid \boldsymbol{\beta}, \boldsymbol{x}) &\propto w_i^{a_w-1/2} \exp\left\{-\left(b_w + \frac{1}{2}(\beta_i - x_i)^2\right) w_i\right\}; w_i > 0. \quad (4.12)
\end{aligned}
$$

4. Generate $U_i^{(\boldsymbol{y}^*)}$ and $V_i^{(\boldsymbol{y}^*)}$ from gamma densities with parameters $(N_i, 1)$ and $(N_i - y_i, 1)$, respectively. Calculate y_i^* from the following.

$$
y_i^* = -\log\left(\frac{U_i^{(\boldsymbol{y}^*)}}{1+\lambda_i(\beta_i)} + \frac{V_i^{(\boldsymbol{y}^*)}}{\lambda_i(\beta_i)}\right).
$$

5. Sample d_i, conditional on y_i^* and β_i, from weighted sum of $G(N_i)$ component mixture of normal densities with weights $h_g(N_i)\,;\, g = (1, 2, \ldots, G(N_i))^T$.

$$
p(d_i = g \mid y_i^*, \beta_i) \propto h_g(N_i)\, p(y_i^* - \log\lambda_i(\beta_i)\,;\, \mu_d(N_i), \sigma_d^2(N_i)).
$$

6. The posterior densities of smoothing parameters are given below.

$$p(r_k) \propto |\mathbf{Q}_{\boldsymbol{r}s}|^{1/2}\, r_k^{a_{r_k}-1} \exp\left\{-r_k\left(b_{r_k} + \frac{1}{2}\boldsymbol{x}_{r_k}^T \mathbf{Q}_k \boldsymbol{x}_{r_k}\right)\right\}. \tag{4.13}$$

$$p(s) \propto |\mathbf{Q}_{\boldsymbol{r}s}|^{1/2}\, s^{a_s-1} \exp\left\{-s\left(b_s + \frac{1}{2}\boldsymbol{x}^T \mathbf{Q} \boldsymbol{x}\right)\right\}. \tag{4.14}$$

The $\boldsymbol{x}_{r_k}$ are only those $\boldsymbol{x}$'s which belong to k^{th} region. Since the parameters do not have explicit forms of well known densities, therefore, the estimates $\boldsymbol{r}$ and s are obtained using Metropolis-Hastings algorithm (Section 3.5.1). A proposal is generated using a uniform density at each iteration and then accepted or rejected based on the acceptance probability. Proposal kernel of r_k and its acceptance probability are

$$r_k^* = r_k^{(0)} + U(L_{r_k}, U_{r_k}),$$

and

$$\alpha_{r_k} = \left(\frac{r_k^*}{r_k^{(0)}}\right)^{a_{r_k}-1} \left(\frac{|\mathbf{Q}_{r_k^*} + s\cdot\mathbf{Q}|}{|\mathbf{Q}_{r_k^{(0)}} + s\cdot\mathbf{Q}|}\right)^{1/2} \times \exp\left\{-\left(r_k^* - r_k^{(0)}\right)\left(b_{r_k} + \frac{1}{2}\boldsymbol{x}_{r_k}^T \mathbf{Q}_k \boldsymbol{x}_{r_k}\right)\right\}.$$

The $U(\cdot,\cdot)$ denotes uniform density with its lower and upper limits. The values with * denote proposed values against the previous value $^{(0)}$. The matrices $\mathbf{Q}_{r_k^*}$ and $\mathbf{Q}_{r_k^{(0)}}$ are obtained by replacing only the k^{th} regional smoothing parameter in $\mathbf{Q}_{\boldsymbol{r}}$, and are given below.

$$\mathbf{Q}_{r_k^*} = \begin{pmatrix} r_1\cdot\mathbf{Q}_1 & . & . & . & . \\ . & . & . & . & . \\ . & . & r_k^*\cdot\mathbf{Q}_k & . & . \\ . & . & . & . & . \\ . & . & . & . & r_M\cdot\mathbf{Q}_M \end{pmatrix},$$

and

$$\mathbf{Q}_{r_k^{(0)}} = \begin{pmatrix} r_1\cdot\mathbf{Q}_1 & . & . & . & . \\ . & . & . & . & . \\ . & . & r_k^{(0)}\cdot\mathbf{Q}_k & . & . \\ . & . & . & . & . \\ . & . & . & . & r_M\cdot\mathbf{Q}_M \end{pmatrix}.$$

Proposal kernel of s with its acceptance probability are given by.

$$s^* = s^{(0)} + U(L_s, U_s).$$

$$\alpha_s = \left(\frac{s^*}{s^{(0)}}\right)^{a_s-1} \left(\frac{|\mathbf{Q}_{\boldsymbol{r}} + s^*\cdot\mathbf{Q}|}{|\mathbf{Q}_{\boldsymbol{r}} + s^{(0)}\cdot\mathbf{Q}|}\right)^{1/2} \exp\left\{-\left(s^* - s^{(0)}\right)\left(b_s + \frac{1}{2}\boldsymbol{x}^T\mathbf{Q}\boldsymbol{x}\right)\right\}.$$

The matrix $\mathbf{Q}_{\boldsymbol{r}}$ is obtained by updating all the regional smoothing parameters in $\mathbf{Q}_{\boldsymbol{r}}$.

Chapter 5

Regionally Smoothed Multivariate Spatial Model

The construction of a linear coregionalized model from multivariate GMRFs is already discussed, please refer to Section 2.3.4. Section 5.1 describes the data generation from multivariate samples (multiple diseases) and probability model. The novel approach to apply spatial smoothing on non-overlapping regions is given in Section 5.2. The smoothing is applied to neighborhood matrix separately on each sample and then combined in the form of precision matrix. The prior densities are proposed in Section 5.3 and the estimation of posterior densities through MCMC is carried out in Section 5.4.

5.1 Data Structure

Let $\{y_{ij} : i = 1, 2, \ldots, N; j = 1, 2, \ldots, J\}$ denote conditionally independent count observations from J random variables (diseases), each of N observations, on some spatial structure. Let the subjects affected by j^{th}, disease are denoted by $\boldsymbol{y}_{\cdot j} = (y_{1j}, y_{2j}, \ldots, y_{Nj})^T$. Suppose the population of i^{th} spatial unit is N_i and probability of risk for j^{th} disease is $\pi_{ij}\,(0 < \pi_{ij} < 1)$, then we suppose that $y_{ij} \mid N_i, \pi_{ij}$ follows a binomial density with parameters N_i and π_{ij}.

$$\begin{aligned} y_{ij} \mid N_i, \pi_{ij} &\sim Bin(N_i, \pi_{ij}), \\ \text{and} \quad \ln\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) &= \ln \lambda_{ij}(\beta_{ij}) = z_{ij}\beta_{ij}, \end{aligned} \tag{5.1}$$

where z_{ij} are covariates. Suppose the density of $\boldsymbol{\beta}_{\cdot j} = (\beta_{1j}, \beta_{2j}, \ldots, \beta_{Nj})^T$, is a multivariate Gaussian density with mean $\boldsymbol{x}_{\cdot j} = (x_{1j}, x_{2j}, \ldots, x_{Nj})^T$ and precision matrix $\boldsymbol{w}_{\cdot j}^T \mathbf{I}_N : \boldsymbol{w}_{\cdot j} = (w_{1j}, w_{2j}, \ldots, w_{Nj})^T$, and $\boldsymbol{w}_{\cdot j} > \mathbf{0}$. The mean $\boldsymbol{x}_{\cdot j}$ follow an iGMRF with parameters $\boldsymbol{\mu}_{\cdot j} = (\mu_{1j}, \mu_{2j}, \ldots, \mu_{Nj})^T$, and a neighborhood matrix $\mathbf{P}$, defined earlier in Section 2.3.4.

5.2 Regionalized Smoothing of Spatial Structure

For the spatial smoothing on regional levels we divide the spatial structure $\mathbf{P}$ of N spatial units into M non-overlapping regions. We introduce M regional smoothing parameters $\{r_k^{(j)} : k = 1, 2, \ldots, M, j = 1, 2, \ldots, J\}$ for each of J variables (diseases). The smoothing parameter for j^{th} disease are denoted by $\boldsymbol{r}^{(j)} = (r_1^{(j)}, r_2^{(j)}, \ldots, r_M^{(j)})^T$. The global smoothing parameter for j^{th} disease is denoted by $s^{(j)}$.

The construction of $\mathbf{P}$ from neighborhood structure is provided in Section 2.3.4 i.e., $P_{ii} = 0$, $P_{il} = 1$ if $i \sim l$, $i \neq l$ and $P_{il} = 0$, otherwise. The parameter $s^{(j)}$ is multiplied with each entry of $\mathbf{P}$ to produce $s^{(j)} \cdot \mathbf{P}$. To apply $\boldsymbol{r}^{(j)}$, we first split $\mathbf{P}$ into M block-diagonal sub-matrices, $\mathbf{P}_1, \mathbf{P}_2, \ldots, \mathbf{P}_M$, one sub-matrix for each region. Each regional smoothing parameter is multiplied with its corresponding neighborhood sub-matrix, $r_1^{(j)}$ is multiplied with $\mathbf{P}_1$ and so on. The precision matrix for j^{th} disease after applying regional smoothing is constructed as

$$\mathbf{P}_{\boldsymbol{r}}^{(j)} = \begin{pmatrix} r_1^{(j)} \cdot \mathbf{P}_1 & . & . & . \\ . & r_2^{(j)} \cdot \mathbf{P}_2 & . & . \\ . & . & . & . \\ . & . & . & . \\ . & . & . & r_M^{(j)} \cdot \mathbf{P}_M \end{pmatrix}.$$

The spatially smoothed complete precision matrix for j^{th} disease is defined as $\mathbf{P}_{\boldsymbol{rs}}^{(j)} = \mathbf{P}_{\boldsymbol{r}}^{(j)} + s^{(j)} \cdot \mathbf{P}$. Let $\boldsymbol{\Sigma}_{J\times J}$ denotes the variance-covariance matrix of $\{\boldsymbol{y}_{\cdot j}; j = 1, 2, \ldots, J\}$ and $\mathbf{A}\mathbf{A}^T$ be its Cholesky factorization, then following MacNab (2016), the joint precision matrix for J samples is given by.

$$\boldsymbol{\Omega} = \boldsymbol{\Sigma}^{-1} \otimes \mathbf{I}_N - \boldsymbol{\Lambda}\mathbf{C}\boldsymbol{\Lambda}^T \otimes \begin{pmatrix} \mathbf{P}_{\boldsymbol{rs}}^{(1)} & . & . & . & . \\ . & . & . & . & . \\ . & . & \mathbf{P}_{\boldsymbol{rs}}^{(j)} & . & . \\ . & . & . & . & . \\ . & . & . & . & \mathbf{P}_{\boldsymbol{rs}}^{(J)} \end{pmatrix}.$$

The matrix $\boldsymbol{\Lambda} = (\mathbf{A}^{-1})^T$ and $\mathbf{C}$ is a $J \times J$ diagonal matrix of within-component local interaction parameters.

5.3 Prior Assumptions

Prior Density of $\boldsymbol{\beta}_{\cdot j}$: $\boldsymbol{\beta}_{\cdot j} = (\beta_{1j}, \beta_{2j}, \ldots, \beta_{Nj})^T$ follow a multivariate Gaussian density with mean $\boldsymbol{x}_{\cdot j}$ and precision $\boldsymbol{w}_{\cdot j}^T \mathbf{I}_N$.

$$p(\boldsymbol{\beta}_{\cdot j} \mid \boldsymbol{x}_{\cdot j}, \boldsymbol{w}_{\cdot j}) \propto |\boldsymbol{w}_{\cdot j}^T \mathbf{I}_N|^{1/2} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}_{\cdot j} - \boldsymbol{x}_{\cdot j})^T \boldsymbol{w}_{\cdot j}^T \mathbf{I}_N (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{x}_{\cdot j})\right\}; \\ -\infty < \boldsymbol{\beta}_{\cdot j} < \infty. \tag{5.2}$$

Prior Density of $\boldsymbol{x}_{\cdot j}$: The prior density of $\boldsymbol{x}_{\cdot j} = (x_{1j}, x_{2j}, \ldots, x_{Nj})^T$ is an iGMRF with parameters $\boldsymbol{\mu}_{\cdot j} = \mathbf{0}$ and spatially smoothed precision matrix $\mathbf{P}_{\boldsymbol{rs}}^{(j)}$.

$$p(\boldsymbol{x}_{\cdot j} \mid \boldsymbol{r}^{(j)}, s^{(j)}) \propto \left(|\mathbf{P}_{\boldsymbol{rs}}^{(j)}|^*\right)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}_{\cdot j}^T \mathbf{P}_{\boldsymbol{rs}}^{(j)} \boldsymbol{x}_{\cdot j}\right); -\infty < \boldsymbol{x}_{\cdot j} < \infty. \tag{5.3}$$

Prior Density of w_{ij} : w_{ij} follows a gamma density with common and fixed parameters $a_w > 0$ and $b_w > 0$ for each variable (disease).

$$p(w_{ij}) \propto w_{ij}^{a_w - 1} \exp\left(-b_w w_{ij}\right); \quad w_{ij} > 0. \tag{5.4}$$

Prior Densities of $\boldsymbol{r}^{(j)}$ *and* $s^{(j)}$: For simplicity of notations, we omit writing j for j^{th} sample (disease) and it is understood that the prior densities are written for j^{th} sample. The prior density of r_k is a gamma density with parameters $a_{r_k} > 0$ and $b_{r_k} > 0$.

$$p(r_k) \propto r_k^{a_{r_k} - 1} \exp(-b_{r_k} r_k); \qquad r_k > 0. \tag{5.5}$$

The prior density of s is also a gamma density with parameters $a_s > 0$ and $b_s > 0$.

$$p(s) \propto s^{a_s - 1} \exp(-b_s s); \qquad s > 0. \tag{5.6}$$

5.4 Implementation Through MCMC Algorithm

The auxiliary mixture sampler for j^{th} variable (disease) is defined on the basis of Section 4.4.1.

$$y_{ij}^* = \ln \lambda_{ij}(\beta_{ij}) + \mu_{d_i}(N_i) + \epsilon_{ij}; \qquad \epsilon_{ij} \mid d_i \sim \mathcal{MIXN}(0, \sigma_{d_i}^2(N_i)). \tag{5.7}$$

Since the size of spatial unit, N_i, is fixed for all the variables (diseases), the indicator d_i which depends on N_i, is also fixed for all the variables. The mean and variance of d_i are $\mu_{d_i}(N_i)$ and $\sigma_{d_i}^2(N_i)$, respectively. The density of $\boldsymbol{y}_{\cdot j}^* = (\boldsymbol{y}_{1j}^*, \boldsymbol{y}_{2j}^*, \ldots, \boldsymbol{y}_{Nj}^*)^T$ is a multivariate Gaussian density with mean $\boldsymbol{\beta}_{\cdot j} + \boldsymbol{\mu}_d$ and precision matrix $\left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N$.

$$p(\boldsymbol{y}_{\cdot j}^* \mid \boldsymbol{\beta}_{\cdot j}) \propto \exp\left\{-\frac{1}{2}(\boldsymbol{y}_{\cdot j}^* - \boldsymbol{\beta}_{\cdot j} - \boldsymbol{\mu}_d)^T \left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N (\boldsymbol{y}_{\cdot j}^* - \boldsymbol{\beta}_{\cdot j} - \boldsymbol{\mu}_d)\right\}; \\ -\infty < \boldsymbol{y}_{\cdot j}^* < \infty. \tag{5.8}$$

Proceed with the following steps to implement MCMC algorithm: generate initial values of $\boldsymbol{r}^{(j)}$ and $s^{(j)}$ to get smoothed precision matrix $\mathbf{P}_{\boldsymbol{rs}}^{(j)}$ and then take a sample from (5.3) to get initial values of $\boldsymbol{x}_{\cdot j}$. Obtain probability of disease risk π_{ij} from observed y_{ij} and population N_i. Use π_{ij} to obtain initial values of y_{ij}^* through $\ln \lambda_{ij}(\beta_{ij})$ from step 4 below. Use β_{ij} and x_{ij} to sample w_{ij} from prior gamma density. Use these initial values in the following algorithm to construct an MCMC chain for the posterior estimates.

1. Sample $\boldsymbol{\beta}_{\cdot j}$ conditional on $\boldsymbol{y}^*_{\cdot j}, \boldsymbol{x}_{\cdot j}$, $\boldsymbol{w}_{\cdot j}$ and $\boldsymbol{d}$. The full conditional density of $\boldsymbol{\beta}_{\cdot j}$ is a multivariate Gaussian density with $\boldsymbol{\mu}_{\boldsymbol{\beta}_{\cdot j}} = (\boldsymbol{\sigma}_d^2 \boldsymbol{w}_{\cdot j} + 1)^{-1}(\boldsymbol{\sigma}_d^2 \boldsymbol{x}_{\cdot j} \boldsymbol{w}_{\cdot j} + \boldsymbol{y}^*_{\cdot j} - \boldsymbol{\mu}_d)$ and precision matrix $\boldsymbol{\Omega}_{\boldsymbol{\beta}_{\cdot j}} = \boldsymbol{\sigma}_d^{-2}(\boldsymbol{\sigma}_d^2 \boldsymbol{w}_{\cdot j} + 1)$.

$$
\begin{aligned}
p(\boldsymbol{\beta}_{\cdot j} \mid \boldsymbol{y}^*_{\cdot j}, \boldsymbol{x}_{\cdot j}, \boldsymbol{w}_{\cdot j}) &\propto |\boldsymbol{w}_{\cdot j}^T \mathbf{I}_N|^{1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{x}_{\cdot j})^T \boldsymbol{w}_{\cdot j}^T \mathbf{I}_N (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{x}_{\cdot j}) \right\}, \\
&\quad \times \exp\left\{ -\frac{1}{2} (\boldsymbol{y}^*_{\cdot j} - \boldsymbol{\beta}_{\cdot j} - \boldsymbol{\mu}_d)^T \left(\boldsymbol{\sigma}_d^{-2}\right)^T \mathbf{I}_N (\boldsymbol{y}^*_{\cdot j} - \boldsymbol{\beta}_{\cdot j} - \boldsymbol{\mu}_d) \right\} \\
p(\boldsymbol{\beta}_{\cdot j} \mid \boldsymbol{y}^*_{\cdot j}, \boldsymbol{x}_{\cdot j}, \boldsymbol{w}_{\cdot j}) &\propto |\boldsymbol{\Omega}_{\boldsymbol{\beta}_{\cdot j}}|^{1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{\cdot j}})^T \boldsymbol{\Omega}_{\boldsymbol{\beta}_{\cdot j}} (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{\mu}_{\boldsymbol{\beta}_{\cdot j}}) \right\}; \\
&\quad -\infty < \boldsymbol{\beta}_{\cdot j} < \infty. \qquad (5.9)
\end{aligned}
$$

2. Sample $\boldsymbol{x}_{\cdot j}$ conditional on $\boldsymbol{\beta}_{\cdot j}, \boldsymbol{w}_{\cdot j}, \boldsymbol{r}^{(j)}$ and $s^{(j)}$. The full conditional density of $\boldsymbol{x}_{\cdot j}$ is a GMRF with $\boldsymbol{\mu}_{\boldsymbol{x}_{\cdot j}} = \boldsymbol{w}_{\cdot j}^T \boldsymbol{\Omega}_{\boldsymbol{x}_{\cdot j}}^{-1} \boldsymbol{\beta}_{\cdot j}$ and precision structure $\boldsymbol{\Omega}_{\boldsymbol{x}_{\cdot j}} = \mathbf{P}^{(j)}_{\boldsymbol{rs}} + \boldsymbol{w}_{\cdot j}^T \mathbf{I}_N$.

$$
\begin{aligned}
p(\boldsymbol{x}_{\cdot j} \mid \boldsymbol{\beta}_{\cdot j}, \boldsymbol{w}_{\cdot j}, \boldsymbol{r}^{(j)}, s^{(j)}) &\propto \left(|\mathbf{P}^{(j)}_{\boldsymbol{rs}}|^*\right)^{1/2} \exp\left(-\frac{1}{2} \boldsymbol{x}_{\cdot j}^T \mathbf{P}^{(j)}_{\boldsymbol{rs}} \boldsymbol{x}_{\cdot j} \right) \\
&\quad \times |\boldsymbol{w}_{\cdot j}^T \mathbf{I}_N|^{1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{x}_{\cdot j})^T \boldsymbol{w}_{\cdot j}^T \mathbf{I}_N (\boldsymbol{\beta}_{\cdot j} - \boldsymbol{x}_{\cdot j}) \right\}, \\
p(\boldsymbol{x}_{\cdot j} \mid \boldsymbol{\beta}_{\cdot j}, \boldsymbol{w}_{\cdot j}, \boldsymbol{r}^{(j)}, s^{(j)}) &\propto |\boldsymbol{\Omega}_{\boldsymbol{x}_{\cdot j}}|^{1/2} \exp\left\{ -\frac{1}{2} (\boldsymbol{x}_{\cdot j} - \boldsymbol{\mu}_{\boldsymbol{x}_{\cdot j}})^T \boldsymbol{\Omega}_{\boldsymbol{x}_{\cdot j}} (\boldsymbol{x}_{\cdot j} - \boldsymbol{\mu}_{\boldsymbol{x}_{\cdot j}}) \right\}; \\
&\quad -\infty < \boldsymbol{x}_{\cdot j} < \infty, \qquad (5.10)
\end{aligned}
$$

3. Generate w_{ij} from full conditional gamma density with parameters $a_w + 0.5$ and $b_w + 0.5\,(\beta_{ij} - x_{ij})^2$.

$$
\begin{aligned}
p(w_{ij} \mid \beta_{ij}, x_{ij}) &\propto w_{ij}^{a_w - 1} \exp(-b_w w_{ij}) w_{ij}^{1/2} \exp\left\{ -\frac{1}{2} (\beta_{ij} - x_{ij})^2 w_{ij} \right\} \\
p(w_{ij} \mid \beta_{ij}, x_{ij}) &\propto w_{ij}^{a_w - 1/2} \exp\left\{ -\left(b_w + \frac{1}{2} (\beta_{ij} - x_{ij})^2 \right) w_{ij} \right\}; w_{ij} > 0. \qquad (5.11)
\end{aligned}
$$

4. Generate $U_{ij}^{(\boldsymbol{y}^*_{\cdot j})}$ and $V_{ij}^{(\boldsymbol{y}^*_{\cdot j})}$ from gamma densities with parameters $(N_i, 1)$ and $(N_i - y_{ij}, 1)$, respectively. Calculate y^*_{ij} from the following.

$$
y^*_{ij} = -\log\left(\frac{U_{ij}^{(\boldsymbol{y}^*_{\cdot j})}}{1 + \lambda_{ij}(\beta_{ij})} + \frac{V_{ij}^{(\boldsymbol{y}^*_{\cdot j})}}{\lambda_{ij}(\beta_{ij})} \right).
$$

5. Sample d_i, conditional on y^*_{ij} and β_{ij}, from weighted sum of $G(N_i)$ component mixture of normal densities with weights $h_g(N_i)\,;\, g = (1, 2, \ldots, G(N_i))^T$.

$$
p(d_i = g \mid y^*_{ij}, \beta_{ij}) \propto h_g(N_i)\, p(y^*_{ij} - \log \lambda_{ij}(\beta_{ij})\,;\, \mu_d(N_i), \sigma_d^2(N_i)).
$$

6. The posterior densities of smoothing parameters do not have explicit forms, therefore, the estimates $\boldsymbol{r}^{(j)}$ and $s^{(j)}$ are obtained using Metropolis-Hastings algorithm (Section 3.5.1). A proposal is generated using a uniform density at each iteration and then accepted or rejected based on the acceptance probability. For simplicity of notations, we omit writing j for j^{th} disease on $\boldsymbol{r}$ and s and it is understood that the following expressions are written for them corresponding to j^{th} disease. Proposal kernel of r_k and its acceptance probability are

$$
\begin{aligned}
& r_k^* = r_k^{(0)} + U(L_{r_k}, U_{r_k}), \\
\text{and} \quad & \alpha_{r_k} = \left(\frac{r_k^*}{r_k^{(0)}}\right)^{a_{r_k}-1} \left(\frac{|\mathbf{P}^{(j)}_{r_k^*} + s \cdot \mathbf{P}|}{|\mathbf{P}^{(j)}_{r_k^{(0)}} + s \cdot \mathbf{P}|}\right)^{1/2} \\
& \qquad \times \exp\left\{-\left(r_k^* - r_k^{(0)}\right)\left(b_{r_k} + \frac{1}{2}\left(\boldsymbol{x}^{(j)}_{r_k}\right)^T \mathbf{P}_k \boldsymbol{x}^{(j)}_{r_k}\right)\right\}.
\end{aligned}
$$

The values with * denote proposed values against the previous value $^{(0)}$ and $\boldsymbol{x}^{(j)}_{r_k}$ are only those $\boldsymbol{x}$'s which belong to k^{th} region in j^{th} sample. The matrices $\mathbf{P}^{(j)}_{r_k^*}$ and $\mathbf{P}^{(j)}_{r_k^{(0)}}$ are obtained by replacing only the k^{th} regional smoothing parameter in $\mathbf{P}_{\boldsymbol{r}}$, and are given below.

$$
\mathbf{P}^{(j)}_{r_k^*} = \begin{pmatrix}
r_1 \cdot \mathbf{P}_1 & . & . & . & . \\
. & . & . & . & . \\
. & . & r_k^* \cdot \mathbf{P}_k & . & . \\
. & . & . & . & . \\
. & . & . & . & r_M \cdot \mathbf{P}_M
\end{pmatrix},
$$

$$
\text{and} \qquad \mathbf{P}^{(j)}_{r_k^{(0)}} = \begin{pmatrix}
r_1 \cdot \mathbf{P}_1 & . & . & . & . \\
. & . & . & . & . \\
. & . & r_k^{(0)} \cdot \mathbf{P}_k & . & . \\
. & . & . & . & . \\
. & . & . & . & r_M \cdot \mathbf{P}_M
\end{pmatrix}.
$$

Proposal kernel of s with its acceptance probability are given by.

$$
\begin{aligned}
& s^* = s^{(0)} + U(L_s, U_s). \\
& \alpha_s = \left(\frac{s^*}{s^{(0)}}\right)^{a_s - 1} \left(\frac{|\mathbf{P}^{(j)}_{\boldsymbol{r}} + s^* \cdot \mathbf{P}|}{|\mathbf{P}^{(j)}_{\boldsymbol{r}} + s^{(0)} \cdot \mathbf{P}|}\right)^{1/2} \exp\left\{-\left(s^* - s^{(0)}\right)\left(b_s + \frac{1}{2}\left(\boldsymbol{x}^{(j)}\right)^T \mathbf{P}\boldsymbol{x}^{(j)}\right)\right\}.
\end{aligned}
$$

The matrix $\mathbf{P}^{(j)}_{\boldsymbol{r}}$ is obtained by updating all the regional smoothing parameters in $\mathbf{P}^{(j)}_{\boldsymbol{r}}$.

Chapter 6

Identification of Risk Clusters

Relative risk estimation (or disease mapping) models govern the global or overall smoothing and estimation of underlying risk level, whereas local properties/heterogeneities of risk such as sharp boundaries between individual areas or elevated (reduced) risk areas are generally not counted for. Cluster detection is focused on such local features of risk. These features depend on the space properties and not on the individual values at areas. Further, relative risk for the overall geographic area may not represent the local risk accurately since population is rarely homogeneous across the study area. Hence, cluster detection is another fundamental aspect to know about the disease characteristics at local levels.

As already discussed that a number of factors contribute to spatial heterogeneities in risk such as heterogeneous population, different administrative structures, etc. therefore, it becomes customary to smooth extreme observations locally to flatten the heterogeneities. To this end, we propose a novel way to regionally smooth neighborhood structure. The effect of such smoothing is that previously the areas with extreme values may otherwise not be properly identified are now adjusted according to risk levels of their neighbors. The regional smoothing utilizes neighborhood structure and accordingly penalizes small extreme values and shrinks large extreme values so that areas with similar risk can be identified together. We will see in the results that clusters identified after regional smoothing make much more sense than from the raw data without smoothing.

6.1 Spatial Clusters

There are several ways to define clusters but a general definition in the context of disease mapping on irregular lattices can be that any spatially bounded area (or a collection of contiguous areas) showing significant elevated (reduced) risk relative to the geographical proximity is called a cluster. The units within a cluster area have a common relative measure of risk throughout that area which really helps in interpretation and comes handy when comparing clusters.

A simple yet widely used criterion is the use of *hot spot* clustering (Richardson et al. (2004)), famous in epidemiology, in which each individual area is considered a cluster itself.

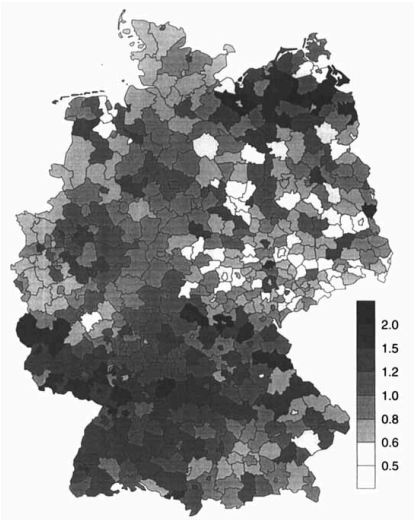
The hot spot clustering is useful when no prior knowledge about the disease is available and a preliminary screening of data could reveal certain characteristics about the disease.

Since spatial data generally has dependence on neighboring areas, the hot spot clustering approach does not quantify such dependence and hence a more general approach which allows the inclusion of neighboring areas in clusters is required.

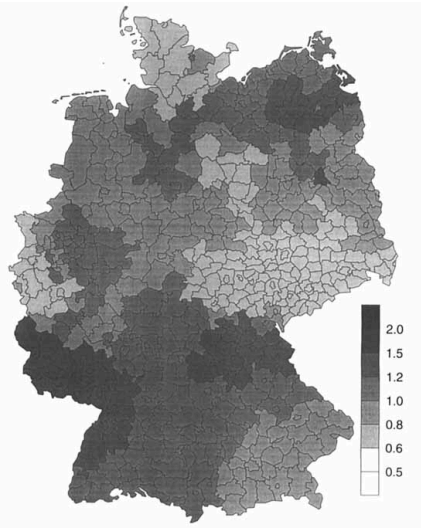
6.1.1 Graphical Explanation

Consider standard mortality ratios (SMRs) of Oral cavity cancer in males from 1986–1990 displayed in Figure 6.1a for 544 counties of Germany. Summary statistics of SMRs show that they range from 0.15 to 2.40 with a standard deviation of 0.39 for log values of SMRs. The variation of SMRs is reduced if we consider only highly populated counties which have at least 50 expected incidences, now they range from 0.53 to 1.60 with a standard deviation of 0.26 for log SMRs. The SMRs appear to be more extreme in less populated counties which shows risk is not equally distributed over spatial counties of Germany.

Consider second graph of median relative risks (taken from Knorr-Held and Raßer (2000)) in Figure 6.1b. The posterior medians range from 0.65 to 1.42, of particular interest are the three clusters having an elevated risk of 1.20 or above; one in north east in Mecklenburg-West Pomerania, second in south west near Saarland along the border of France and third in south east in Bavaria. The elevated risk in these clusters is generally linked to higher usage of alcohol and tobacco smoking, see Knorr-Held and Raßer (2000) and references therein. The discussion of the two graphs clearly show that there are unknown



(a) Standard mortality ratios (SMRs)



(b) Median relative risk

Figure 6.1: Standard mortality ratios and estimated median relative risks for Oral cavity cancer for the period 1986–1990 (taken from Knorr-Held and Raßer (2000)).

clusters in Oral cavity cancer data and therefore these clusters should be properly detected to reveal elevated (reduced) risks areas.

6.2 Identification Methods

There are two main approaches for identification of risk heterogeneities on irregular lattices. The first approach identifies discontinuities or step-changes in disease risk between geographically adjacent areas, see for example Lu et al. (2007); Lee and Mitchell (2013). The discontinuities identified with these methods do not produce collectively exhaustive borders i.e., the areas or the group of identified units do not complete the entire space. The second approach, such as Charras-Garrido et al. (2012), however, identifies areas or a collection of neighboring areas that exhibit significant relative risk compared to geographical proximity. This approach sews neighboring areas and therefore ensures the close boundary as well as covers entire space.

One prime method in the first approach is scan statistics developed by Kulldorff (1997), which identifies elevated risk areas of disease. It is readily available in form of the SaTScan software for easy implementation. However, the literature on disease mapping suggests that scan statistics is not suitable for identification of spatial patterns, see for example Anderson et al. (2014).

Knorr-Held and Raßer (2000) identifies clusters using Besag et al. (1991) models in a hierarchical framework, they force the clusters to be spatially contiguous and assume the risk to be constant within a cluster. The constant risk within a cluster has an advantage that it partitions the relative risk into various risk classes which is useful for easy interpretation and comparison. However, their approach requires computationally complex estimation in form of reversible jump MCMC algorithm.

6.3 Agglomerative Approach to Clustering

We utilize the cluster identification approach proposed by Anderson et al. (2014). They propose a two-stage solution for identification of clusters which exhibit alike disease risks. In the first stage, a spatially adjusted hierarchical agglomerative clustering algorithm is applied on the disease data, prior to study period. The first stage produces a set of potential cluster structures based on one of the three linkage measures, single linkage measure, centroid linkage measure and Ward's linkage measure. The second stage fits a Bayesian Poisson log-linear model to study data only, to pick best cluster structure elicited at first stage. The best cluster structure is finalized by observing the deviance information statistic.

In this approach, each of the N non-overlapping areas is considered a cluster itself, making a total of N clusters in the start. The two clusters with least Euclidean distance join together, reducing the number of clusters by one at each iteration. This process continues until all the clusters are merged together to form only one giant cluster. The second stage selects the best configuration of clusters.

The following algorithm produces the clusters whose subjects within a single cluster are equally affected by the disease.

1. Construct $C_n = \{C_n(1), C_n(2), \ldots, C_n(N)\}$, an initial cluster structure where each unit is a cluster itself.

2. Repeat the following process $N - 1$ times.
   — Compute $N \times N$ matrix based on Euclidean distance among the neighboring clusters.
   — Find the two least dissimilar clusters based on Euclidean distance and merge them.
   — Update the cluster structure from step 1 and update the data where two clusters have been merged.

Choose the cluster configuration which minimizes the deviance information statistic.

6.4 Ranking of Clusters

The clusters identification approach defined in Section 6.3 only identifies a group of contiguous areas which have alike risk but does not identify the order in which they can be ranked from least affected to most affected.

We utilize the measure of IR and rank the identified clusters i.e., we combine incidences y_i's and population N_i's of all the areas within each cluster and obtain IR ($\Sigma y_i / \Sigma N_i$) of the cluster. The incidence rate ratio (IRR) of two clusters is computed by dividing IR of one cluster to IR of the other cluster. Let us say, if K clusters are identified, then there will be $K \cdot (K-1)/2$ possible IRRs (all possible combinations of 2-clusters from K clusters). A cluster is ranked least affected if the ratio of its IRR when compared with all other IRRs is less than 1. On the other hand, it is ranked most affected if this ratio is greater than 1.

Chapter 7

Simulation Study

7.1 Univariate Regionally Smoothed Model

7.1.1 Simulation Parameters

Consider the spatial structure of German administrative counties, there are 544 spatial units scattered throughout Germany, the study variable has 544 observations $y_1, y_2, \ldots, y_{544}$. The population size in each county is fixed to a large number, $N_1 = 40,000 = N_2 = \cdots = N_{544}$ and probability of risk π_i is obtained by the ratio $y_i/40,000$. The precision matrix $\mathbf{Q}_{544\times544}$ is constructed from neighborhood matrix as $\mathbf{Q} = diag(ne(i)) - \mathbf{W}$, where $diag(ne(i))$ is a diagonal only matrix with number of neighbors on the diagonal and neighborhood matrix $\mathbf{W}$ is defined in Section 2.1.5 ($w_{ii} = 0, w_{ij} = 1$ if $i \sim j$, otherwise $w_{ij} = 0$).

Suppose Federal states of Germany are local regions to apply regional smoothing. There are sixteen Federal states, therefore, $M = 16$. The precision matrix is partitioned into sixteen block-diagonal submatrices as described in Section 4.2, $\mathbf{Q}_1$ to $\mathbf{Q}_{16}$ and smoothing parameters are $\boldsymbol{r} = \{r_1, r_2, \ldots, r_{16}\}$.

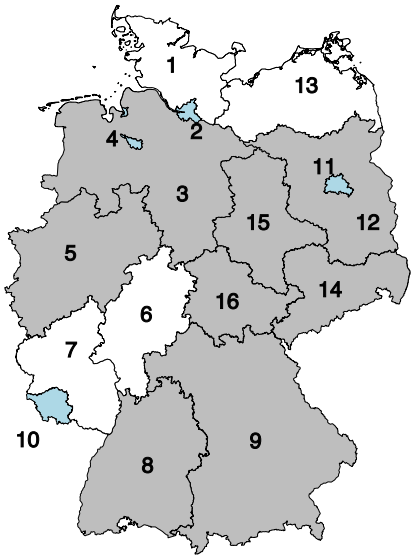


Figure 7.1: Classification of states for regional smoothing

To select numerical values of r_k ; $k = 1, 2, \ldots, 16$, we classify states into three categories based on the number of counties (Figure 7.1). First category contains four states (2, 4, 10, 11), they have less than 12 counties each. Second category also has four states (1, 6, 7, 13)

in which each state contains between 12 and 36 counties. The rest are categorized into third category (3, 5, 8, 9, 12, 14, 15, 16), they have at least 37 counties each. We use same value of r_k for states within each category e.g., ($r_2 = r_4 = r_{10} = r_{11} = 0.1, r_1 = r_6 = r_7 = r_{13} = 0.2$, $r_3 = r_5 = r_8 = r_9 = r_{12} = r_{14} = r_{15} = r_{16} = 0.5$).

The simulation of smoothing parameters is done by using various sets of $\boldsymbol{r}$ and s with three possible scenarios in mind. Scenario-I: $\boldsymbol{r}$ and s both change simultaneously; Scenario-II: s changes with fixed $\boldsymbol{r}$; and Scenario-III: $\boldsymbol{r}$ changes with fixed s. Different values of s are considered between 0.8 and 4.0, the minimum and the maximum values of $\boldsymbol{r}$ in each state are shown in the following table.

Category	states	$\boldsymbol{r}$
I	(2, 4, 10, 11)	(0.1,0.5)
II	(1, 6, 7, 13)	(0.2,1.0)
III	(3, 5, 8, 9, 12, 14, 15, 16)	(0.5,2.5)

The hyperparameters of prior densities of r_k and s in (4.5) and (4.6) are chosen in such a way that the mean of prior density becomes equivalent to true population parameter. Take, for example, $s = 0.8$, the hyperparameters for this value are $a_s = 800$ and $b_s = 1,000$. Small value of s requires small value of scale hyperparameter ($b_s^{-1} = 0.001$) so that its posterior density does not move far away from true value and explores proposals only in the vicinity.

The precision of $\boldsymbol{\beta}$ depends on tuning parameter $\boldsymbol{w}$, which in turn is governed by hyperprior (4.4) with hyperparameters a_w and b_w. The optimal choice of these values is $a_w = 100$ and $b_w = 10$ which produces good mixing in posterior density of $\boldsymbol{w}$ and in turn provide convergence of $\boldsymbol{\beta}$ to its stationary density.

Since $N = 40,000$, the density of d in (4.7) reduces to a univariate Gaussian density instead of a mixture of Gaussians. The mean and variance of d are -10.59662 and $2.5e^{-0.05}$, respectively (see appendix of Frühwirth-Schnatter et al. (2009)).

The model parameters are simulated 30 times and each time a Markov chain of 10,000 values is generated. Since y_i's are correlated (with neighbors) with strong positive spatial autocorrelation, therefore, to obtain an independent sample from simulations, every 10^{th} value is collected with the assumption that y_i^{th} and $(y_i + 10)^{\text{th}}$ are independent (Albert and Chib (1993)). Further, first 100 values are burn-in from Markov chain of 1,000 values to reduce any initial value effects, posterior estimates are then obtained from further reduced chain of 900 values.

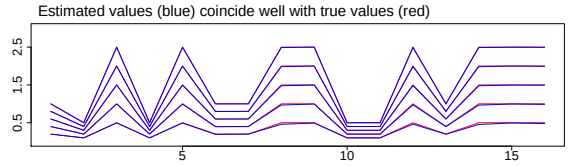
7.1.2 Results

We use different sets of $\boldsymbol{r}$ and s to compare the amount of smoothing. Their true values with posterior means and variances are shown in Figures 7.2 to 7.4. Each line, for example, in Figure 7.2b connects sixteen regional smoothing parameters (true and estimated). The line closest to horizontal axis has lowest set of $\boldsymbol{r}$ values, we call it Level-1 of regional smoothing and use it as a base level to compare with other levels. Figure 7.2b shows other four levels as well (moving from bottom to top). Similarly, Figure 7.2a shows various levels of s, and

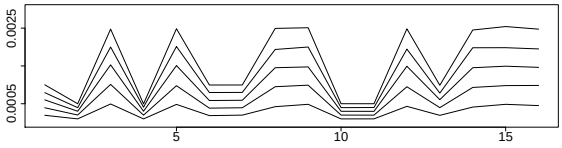
$s = 0.8$ against Level-1 is used as base level for comparison with other levels of s. Figure 7.2 displays the results under first scenario when both $\boldsymbol{r}$ and s change simultaneously from Level-1 to Level-5 of smoothing. In the second scenario, Figure 7.3 displays them when $\boldsymbol{r}$ is fixed at base level and Figure 7.4 displays them under third scenario when s is fixed at base level.

(a) True values of s, posterior means ($\hat{s}$) and posterior variances of $\hat{s}$

Levels	1	2	3	4	5
s	0.8	1.5	2.5	3.2	4.0
$\hat{s}$	0.7113	1.4535	2.4779	3.1918	3.9994
$Var(\hat{s})$	0.0006	0.0014	0.0024	0.0032	0.0040



(b) True values ($\boldsymbol{r}$) and their posterior means

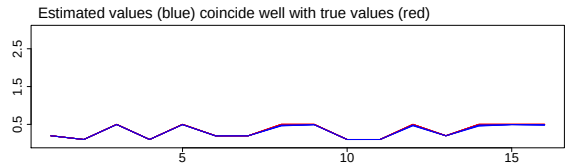


(c) Posterior variances of $\hat{\boldsymbol{r}}$

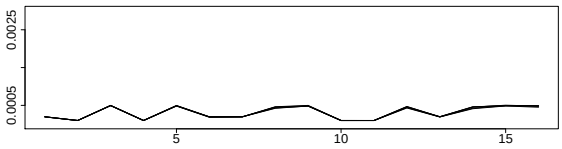
Figure 7.2: Smoothing parameters under Scenario-I

(a) True values of s, posterior means ($\hat{s}$) and posterior variances of $\hat{s}$

Levels	1	2	3	4	5
s	0.8	1.5	2.5	3.2	4.0
$\hat{s}$	0.7113	1.4111	2.4129	3.1121	3.9211
$Var(\hat{s})$	0.0006	0.0013	0.0023	0.0030	0.0038



(b) True values ($\boldsymbol{r}$) and their posterior means



(c) Posterior variances of $\hat{\boldsymbol{r}}$

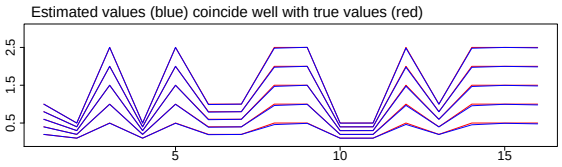
Figure 7.3: Smoothing parameters under Scenario-II

Level-1 or base level indicates low amount of smoothing while Level-5 indicates its strong degree. It is obvious from the figures that the parameters are estimated accurately no matter what amount of smoothing is used. Posterior variances of s vary from 0.0006 to 0.0040 in comparison to posterior variances or $\boldsymbol{r}$ which vary between 0.0001 and 0.0026. Posterior variances are low when values are low (e.g. 0.0006 for $s = 0.8$) and higher for

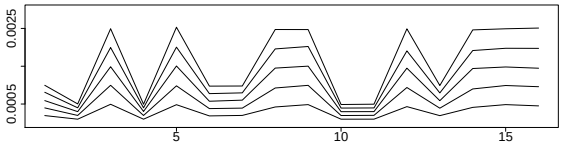
high values, which is expected as higher values mean Markov chain explores a wide range around true values.

(a) True values of s, posterior means ($\hat{s}$) and posterior variances of $\hat{s}$

Levels	1	2	3	4	5
s	0.8	0.8	0.8	0.8	0.8
$\hat{s}$	0.7113	0.7483	0.7657	0.7750	0.7822
$Var(\hat{s})$	0.0006	0.0007	0.0007	0.0007	0.0007



(b) True values ($\boldsymbol{r}$) and their posterior means



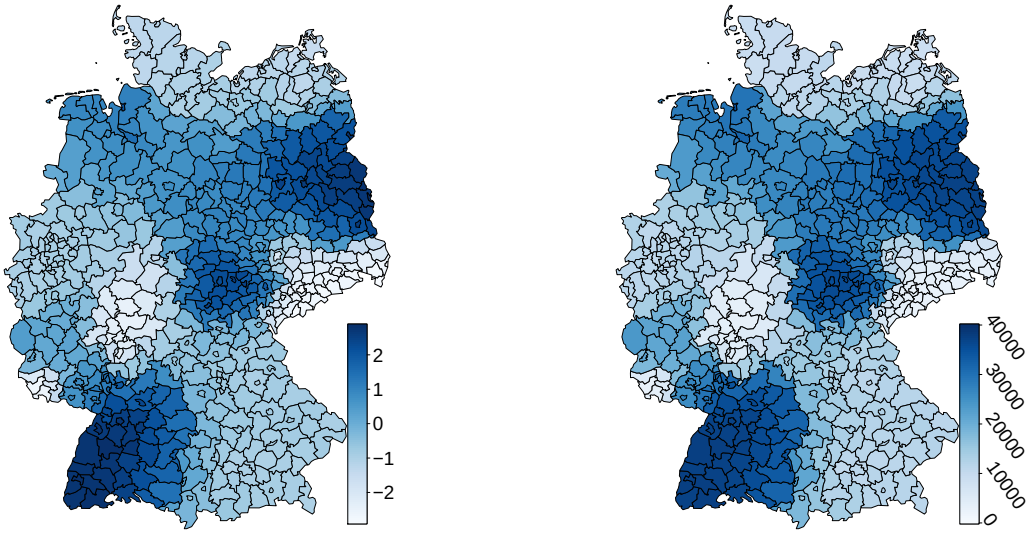
(c) Posterior variances of $\hat{\boldsymbol{r}}$

Figure 7.4: Smoothing parameters under Scenario-III

Figure 7.5a shows simulated GMRF (prior mean of $\boldsymbol{\beta}$) from the smoothed data of neighborhood structure of 544 German counties. The conditional dependence of counties on their neighbors can be read from map. This map is produced with state wise mean values $(-1,-3,1,3,-1,-3,1,3,-1,-3,1,3)$ and smoothed precision matrix $\mathbf{Q}_{\boldsymbol{r}s}$, for which both the smoothing parameters are set at base levels i.e., $r_2 = r_4 = r_{10} = r_{11} = 0.1, r_1 = r_6 = r_7 = r_{13} = 0.2, r_3 = r_5 = r_8 = r_9 = r_{12} = r_{14} = r_{15} = r_{16} = 0.5$ and $s = 0.8$. The Figure 7.5b displays simulated incidences based on the simulated GMRF. The number of incidences per county vary from 2,000 to 38,000 with π_i ranging from 0.05 to 0.95. The dark regions have high probabilities of risk such as Baden-Württemberg, Brandenburg and Thüringen states. The counties in these regions have relatively elevated incidences.

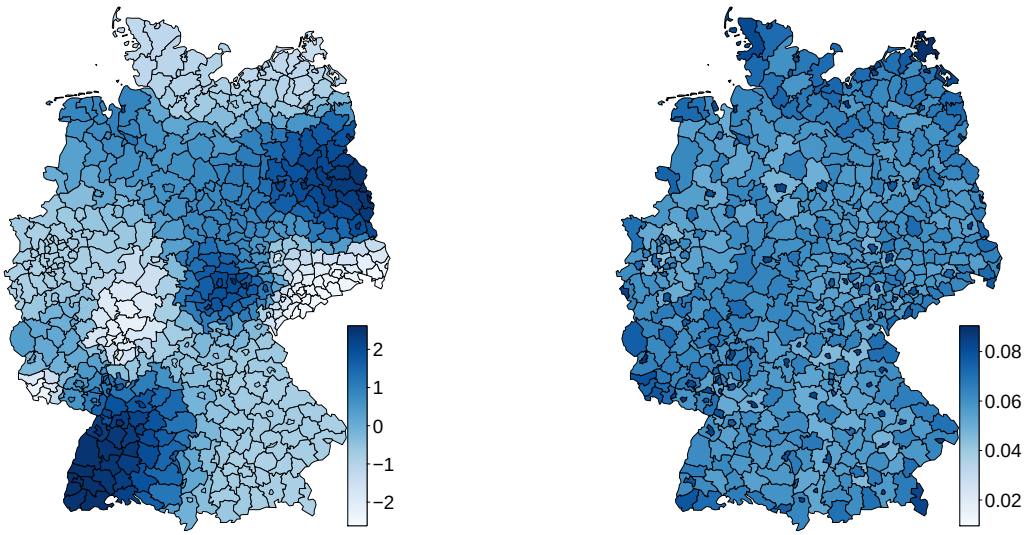
The map similar to Figure 7.5a is simulated for all three scenarios of smoothing parameters and the posterior estimates $\hat{\boldsymbol{x}}$ under just first scenario are presented here in Figures 7.6 and 7.7, while for the other two scenarios are given in appendix in Figures A.2 and A.3. Look at the Figure 7.6a and move from Figure 7.7a to Figure 7.7d, it is clear that the proposed method of local smoothing helps reduce local (regional) variations. The regions with extreme risk probabilities are now smoothed toward their regional mean risk values. If we compare posterior variances of base level (Figure 7.6b) to Level-5 (Figure 7.7h), we see that variances are generally high when there is much variation in the form of heterogeneities in underlying risk but when the risk becomes relatively smooth these posterior variances also become small.

The simulation of incidences is done in such a way that y_i per county is allowed to have extreme values as well, such as some counties have as little as just 2,000 incidences and some have as more as 38,000 incidences (out of 40,000 population size per county). In 30 simulated data sets, the minimum, maximum and mean values of y_1 are approximately 2,500, 34,000 and 7,000, respectively. Figure A.1 simply depicts such variation in all the



(a) $\boldsymbol{x}$ (b) $\boldsymbol{y}$

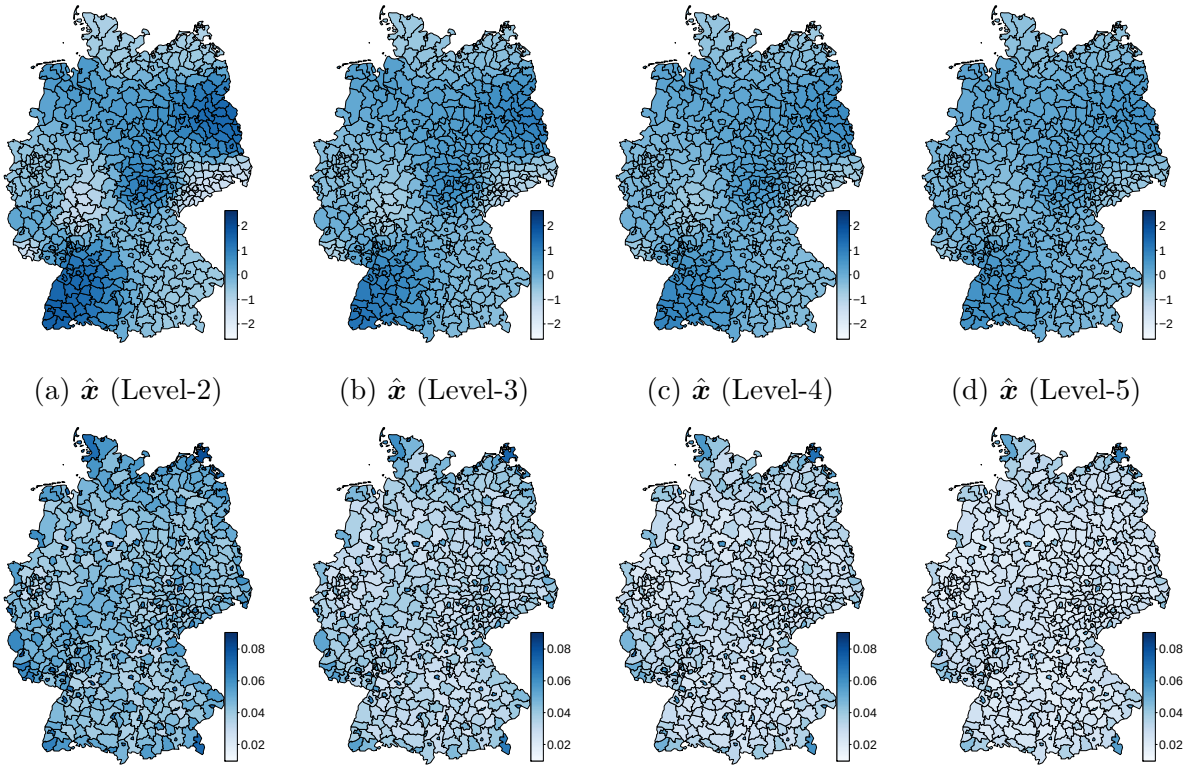
Figure 7.5: Simulated GMRF with regional smoothing and corresponding incidences (base level)



(a) $\hat{\boldsymbol{x}}$ (b) $\mathrm{Var}(\hat{\boldsymbol{x}})$

Figure 7.6: Posterior means and variances of $\hat{\boldsymbol{x}}$ (base level)

respective variables over 30 simulations. The upper, middle and lower lines in each plot show maximum, average and minimum values, respectively. The Figures A.1b and A.1d reveal that the minimum and maximum values of $\boldsymbol{y}$ are scattered away from their mean (blue line), however, these values of $\boldsymbol{y}^*$ are relatively close to their mean proving that the proposed



(a) $\hat{\boldsymbol{x}}$ (Level-2) (b) $\hat{\boldsymbol{x}}$ (Level-3) (c) $\hat{\boldsymbol{x}}$ (Level-4) (d) $\hat{\boldsymbol{x}}$ (Level-5)

(e) Var($\hat{\boldsymbol{x}}$) (Level-2) (f) Var($\hat{\boldsymbol{x}}$) (Level-3) (g) Var($\hat{\boldsymbol{x}}$) (Level-4) (h) Var($\hat{\boldsymbol{x}}$) (Level-5)

Figure 7.7: Posterior means and variances of $\hat{\boldsymbol{x}}$ for various levels of smoothing (Scenario-I)

method of regional smoothing tunes values to their local means. The posterior variances are provided on log scale, their absolute values are pretty low, falling approximately between 0.0405 and 0.0998 for $\hat{\boldsymbol{x}}$ and between 0.0001 and 0.0025 for $\hat{\boldsymbol{y}}^*$.

7.1.3 Summary

We apply adaptive regional smoothing based on number of counties in a region (state) and consider three scenarios of regional and global smoothing. It is noticed that Scenario-I (both smoothing parameters change simultaneously) performs quite well. In case of other scenarios, either some areas are left un-smoothed or we need very high amount of smoothing in some areas to get regionally smoothed data, as evident from Figures A.2 and A.3.

Regarding the amount of smoothing, Level-3 and Level-4 of Scenario-I (Figures 7.7b and 7.7c) can be considered a better choice in the current set-up, as the posterior estimates do not look over smoothed.

7.2 Multivariate Regionally Smoothed Model

7.2.1 Simulation Parameters

Consider $J = 3$, three samples (diseases) are observed at a time, to illustrate regional smoothing in multivariate spatial model. Suppose incidences are observed on spatial structure of German administrative counties, which are 544. For each sample, the study variable has 544 observations $y_{ij}, y_{2j}, \ldots, y_{544j}$. The population size in each county is set to $N = 40,000$. The probability of risk π_{ij} is calculated from dividing the number of incidences in each county by its population size i.e., $y_{ij}/40,000$. The precision matrix $\mathbf{P}_{544\times544}$ is equivalent to neighborhood matrix $\mathbf{W}$, defined in Section 2.1.5 i.e., $P_{ii} = 0$, $P_{il} = 1$ if $i \sim l$ and $P_{il} = 0$, otherwise.

We suppose Federal states of Germany to be the local regions to apply regional smoothing. There are sixteen Federal states of Germany, therefore, $M = 16$. The neighborhood matrix $\mathbf{P}$ is partitioned state wise into sixteen block-diagonal sub-matrices as described in Section 5.2, $\mathbf{P}_1$ to $\mathbf{P}_{16}$, and $\boldsymbol{r}^{(j)}, = \{r_1^{(j)}, r_2^{(j)}, \ldots, r_{16}^{(j)}\}$.

To select numerical values of $r_k^{(j)}$; $k = 1, 2, \ldots, 16$, we classify states into three categories based on the number of counties (Figure 7.1). First category consists of states 2, 4, 10, 11, second category contains 1, 6, 7, 13 and third category contains 3, 5, 8, 9, 12, 14, 15, 16. Similar to univariate model, we use same value of $r_k^{(j)}$ for states within each category.

The simulation of smoothing parameters is done using various sets of $\boldsymbol{r}^{(j)}$ and $s^{(j)}$. Three possible scenarios are considered similar to univariate regionalized spatial model. Scenario-I: $\boldsymbol{r}^{(j)}$ and $s^{(j)}$ both change simultaneously; Scenario-II: $s^{(j)}$ changes with fixed $\boldsymbol{r}^{(j)}$; and Scenario-III: $\boldsymbol{r}^{(j)}$ changes with fixed $s^{(j)}$. Different values of $s^{(j)}$ are considered between 0.8 and 4.0, the minimum and maximum values of $\boldsymbol{r}^{(j)}$ per state are the same as shown in table in Section 7.1.1. The matrix of within interaction parameters C is a diagonal matrix of order 3×3 with values between 0.0005 to 0.1, depending on the choice of smoothing parameters ($\boldsymbol{r}^{(j)}$ and $s^{(j)}$).

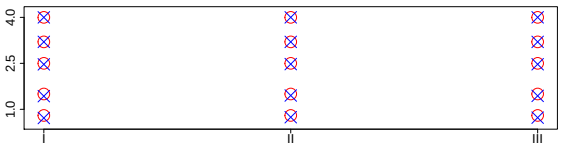
Since $N = 40,000$, the density of d in (5.7) reduces to a univariate Gaussian density instead of a mixture of Gaussians. The mean and variance of d are -10.59662 and $-2.5e^{-0.05}$, respectively (see appendix of Frühwirth-Schnatter et al. (2009)).

The model parameters are simulated 30 times and each time a Markov chain of 10,000 values is generated. Since y_i's are correlated (with neighbors) with strong positive spatial autocorrelation, therefore, to obtain an independent sample from simulations, every 10^{th} value is collected with the assumption that y_i^{th} and $(y_i + 10)^{\text{th}}$ are independent (Albert and Chib (1993)). Further, first 100 values are burn-in from Markov chain of 1,000 values to reduce any initial value effects, posterior estimates are then obtained from further reduced chain of 900 values.

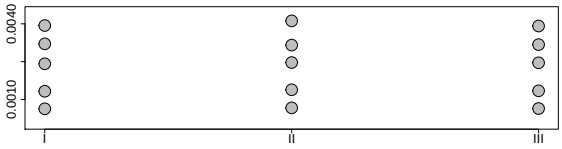
7.2.2 Results

We use different sets of $\boldsymbol{r}^{(j)}$ and $s^{(j)}$ to compare the amount of smoothing. Their true values with posterior means and variances are shown in Figures 7.8, A.4 and A.5. Figure 7.8a

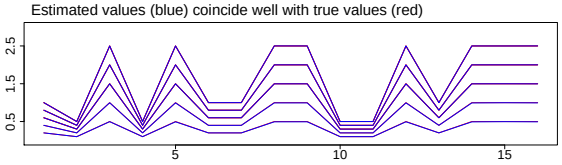
shows various levels of global smoothing parameter s. True values are represented by $\circ$ and their posterior means by $\times$. The posterior variances are shown by $\bullet$ in Figure 7.8b. The lowest value $s = 0.8$ against Level-1 is used as base level in all samples for comparison with other levels. Each line, for example, in Figure 7.8c connects sixteen regional smoothing parameters (true and estimated). The line closest to horizontal axis has lowest set of $\boldsymbol{r}^{(1)}$ values, we call it Level-1 of regional smoothing and use it as a base level to compare with other four levels in the same figure (moving from bottom to top). Figure 7.8 displays the results under first scenario when both $\boldsymbol{r}^{(j)}$ and $s^{(j)}$ change simultaneously from Level-1 or base level to Level-5 of smoothing. In the second scenario, Figure A.4 displays them when $\boldsymbol{r}^{(j)}$ is fixed at base level and Figure A.5 displays them under third scenario when $s^{(j)}$ is fixed at base level.



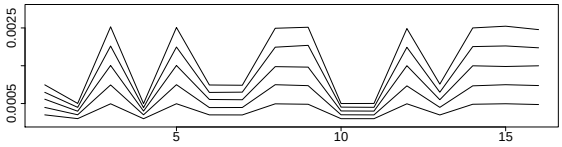
(a) True values $(s^{(1)}, s^{(2)}, s^{(3)})$ and posterior means $(\hat{s}^{(1)}, \hat{s}^{(2)}, \hat{s}^{(3)})$



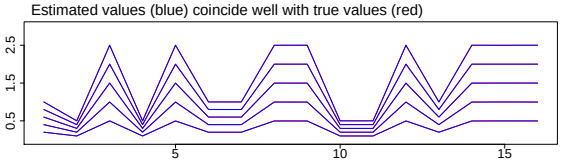
(b) Posterior variances of $\hat{s}^{(1)}, \hat{s}^{(2)}$ and $\hat{s}^{(3)}$



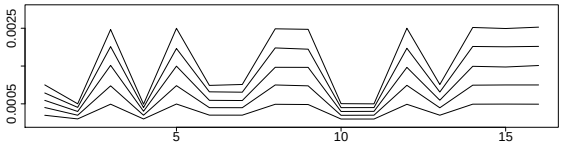
(c) True values $(\boldsymbol{r}^{(1)})$ and posterior means $(\hat{\boldsymbol{r}}^{(1)})$



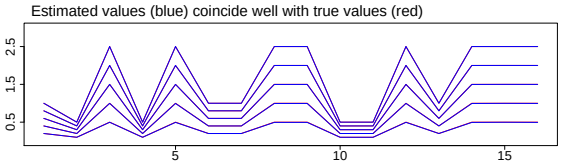
(d) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(1)}))$



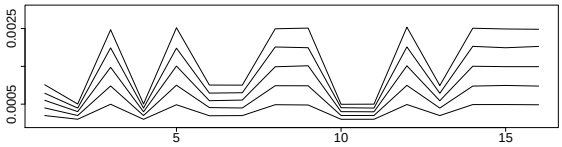
(e) True values $(\boldsymbol{r}^{(2)})$ and posterior means $(\hat{\boldsymbol{r}}^{(2)})$



(f) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(2)}))$



(g) True values $(\boldsymbol{r}^{(3)})$ and posterior means $(\hat{\boldsymbol{r}}^{(3)})$



(h) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(3)}))$

Figure 7.8: Smoothing parameters under Scenario-I

Base level indicates low amount of smoothing while Level-5 indicates strong amount. It is obvious from the figures that the parameters are estimated accurately no matter what amount of smoothing is used. Posterior variances of s vary from 0.0006 to 0.0041 in comparison to posterior variances or $\boldsymbol{r}$ which vary between 0.0001 and 0.0026. Posterior variances are low when true values are low (e.g. 0.0006 for $s = 0.8$ and 0.0001 for $\hat{r}_2, \hat{r}_4, \hat{r}_{10}$ and $\hat{r}_{11}$) and higher for high true values, it is expected as higher values mean that Markov chain proposes values from a wide range around true value.

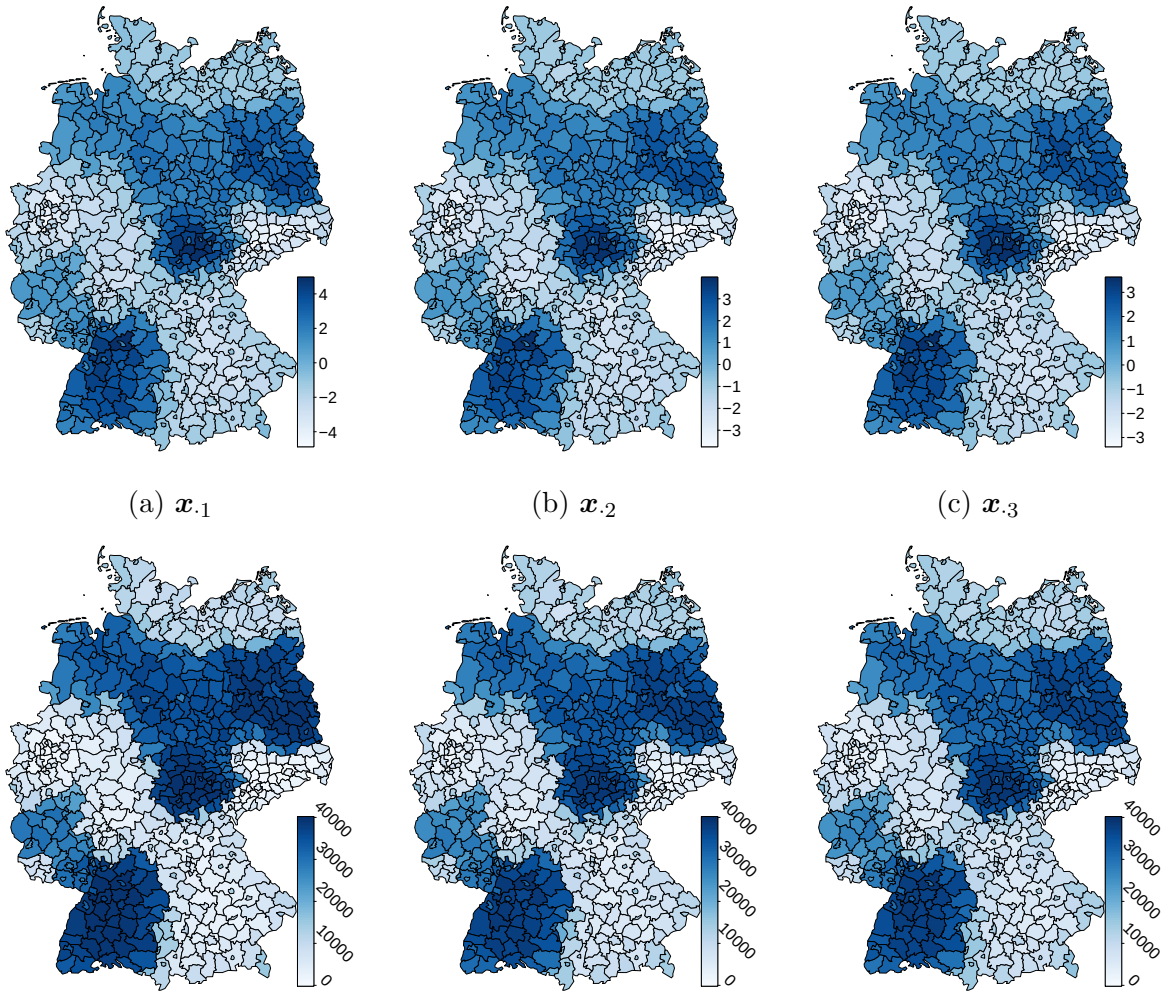
Figures 7.9a to 7.9c display simulated GMRFs from the neighborhood structure of 544 German counties. The conditional dependence of counties on their neighbors can be read from maps. The maps are produced with state wise mean values $(-1, -2, 1, 2, -1, -2, 1, 2, -1, -2, 1, 2)$ and precision matrix $\mathbf{P}_{\boldsymbol{rs}}^{(j)}$, for which base level values of both the smoothing parameters are used i.e., $r_2^{(j)} = r_4^{(j)} = r_{10}^{(j)} = r_{11}^{(j)} = 0.1, r_1^{(j)} = r_6^{(j)} = r_7^{(j)} = r_{13}^{(j)} = 0.2, r_3^{(j)} = r_5^{(j)} = r_8^{(j)} = r_9^{(j)} = r_{12}^{(j)} = r_{14}^{(j)} = r_{15}^{(j)} = r_{16}^{(j)} = 0.5$ and $s^{(j)} = 0.8$. The number of simulated incidences of samples 1, 2 and 3 are displayed in Figures 7.9d to 7.9f. They vary from 9 to 39,989 with extreme risk probabilities. The regions of Baden-Württemberg, Brandenburg and Thüringen have extremely high incidences while the regions Hesse, Saarland and Saxony have relatively low incidences.

The maps similar to Figure 7.9 are simulated for all three scenarios of smoothing parameters and the posterior estimates $\hat{\boldsymbol{x}}$ are obtained, they are given in Figure A.6 for base level, while for other levels and scenarios are presented in Figures A.7 to A.15. For each sample, compare posterior means and variances level wise, we see that the regions with extreme risk probabilities are smoothed toward their regional mean risk values and variances are large when the variation (heterogeneities on maps) in underlying risk is high but when the regionalized smoothing makes the risk relatively flat, these variances also become small.

The incidences are simulated in such a way that y_{ij} varies from as little as just 9 to as more as 39,989 out of $N = 40,000$. In 30 simulated data sets, the minimum and maximum values of y_{11}, y_{12} and y_{13} are respectively (432, 39,650), (1,242, 39,084) and (1,582, 38,777). Figure A.16 shows such extreme values from simulations in all the respective variables. The upper, middle and lower lines in each plot show maximum, mean and minimum values, respectively. Take a look at the values of y_{ij} and y_{ij}^* in Figures A.16b and A.16d, the minimum and maximum values of y_{ij}^* are relatively close to their mean proving that the proposed method of regional smoothing tunes local values to their mean values. The posterior variances are provided on log scale, their absolute values are pretty low, (0.0963, 0.1055) for $\hat{\boldsymbol{x}}_{\cdot 1}$, (0.0964, 0.1031) for $\hat{\boldsymbol{x}}_{\cdot 2}$, (0.0960, 0.1023) for $\hat{\boldsymbol{x}}_{\cdot 3}$, (0.0001, 0.0049) for $\hat{\boldsymbol{y}}_{\cdot 1}^*$, (0.0001, 0.0016) for $\hat{\boldsymbol{y}}_{\cdot 2}^*$ and (0.0001, 0.0012) for $\hat{\boldsymbol{y}}_{\cdot 3}^*$.

7.2.3 Summary

We apply adaptive regional smoothing based on number of counties in a region (state) and consider three scenarios of regional and global smoothing. Scenario-I (both smoothing parameters change simultaneously) performs a little bit better but not as good as in the



(a) $\boldsymbol{x}_{\cdot 1}$ (b) $\boldsymbol{x}_{\cdot 2}$ (c) $\boldsymbol{x}_{\cdot 3}$

(d) $\boldsymbol{y}_{\cdot 1}$ (e) $\boldsymbol{y}_{\cdot 2}$ (f) $\boldsymbol{y}_{\cdot 3}$

Figure 7.9: Simulated GMRFs with regional smoothing and corresponding incidences (base level)

univariate regionalized spatial model.

Regarding the amount of smoothing, Level-4 and Level-5 of Scenario-I (or may be even higher values) can be preferred for the Federal states of Germany, as they produce reasonable well smoothed posterior estimates (Figures A.7 to A.9).

Chapter 8

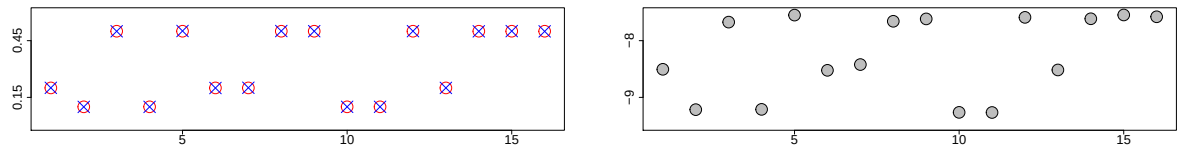
Real-life Application Study

8.1 Univariate Application

8.1.1 Regionalized Smoothing

Data of Oral cancer incidences are taken from Becker and Wahrendorf (2013), in which 15,466 individuals are affected in Germany from the period 1986–1990. These individuals are unequally spread over 544 German administrative counties or 16 Federal states. The number of incidences, y_i, vary from 1 to 501, and the population, N_i, ranges from 15,465 to 2,018,842. The incidence ratio (IR) for each county is computed from dividing y_i by N_i. The map of natural log of IRs (Figure 8.1) reveals that the incidences are not homogeneously distributed across Germany. The variation present in the map is due to different regional differences such as, different administrative set ups, heterogeneous population etc., this has already been discussed in Section 6.1.1.

To apply regional smoothing, we classify states based on the number of counties, the details are given in Section 7.1.1 with classification shown in Figure 7.1. We choose $r_1 = r_6 = r_7 = r_{13} = 0.2$, $r_2 = r_4 = r_{10} = r_{11} = 0.1$ and 0.5 value for the rest of states. The global smoothing parameter s is set to 0.8. The true values along side posterior means and log of variances are given in Figure 8.2. True values are represented by $\circ$ and their posterior means by $\times$ while log of variances are shown by $\bullet$. The variances of r lie between 0.0001 and 0.0005. The posterior mean of s is 0.7998 with 0.0008 posterior variances.



(a) True values ($\boldsymbol{r}$) and their posterior means ($\hat{\boldsymbol{r}}$) (b) Posterior variances of $\hat{\boldsymbol{r}}$ on log scale

Figure 8.2: Smoothing parameters for incidences of Oral cancer

The GMRF $\boldsymbol{x}$ from smoothed neighborhood matrix, their posterior means $\hat{\boldsymbol{x}}$ with posterior variances for each county are shown in (Figure 8.3). The global mean of all 544 observations of $\boldsymbol{x}$ is 0.1001, we can see on its map that some counties have mean value of as low as –4 and some of them have value of upto 4. The posterior means show that the inclusion of smoothing parameters in spatial binomial model has data free from extreme observations. The global mean of posterior means is –0.0117 and most of the counties have mean values close to this value, in fact, 12 out of 16 states have their regional mean values equal to global mean.

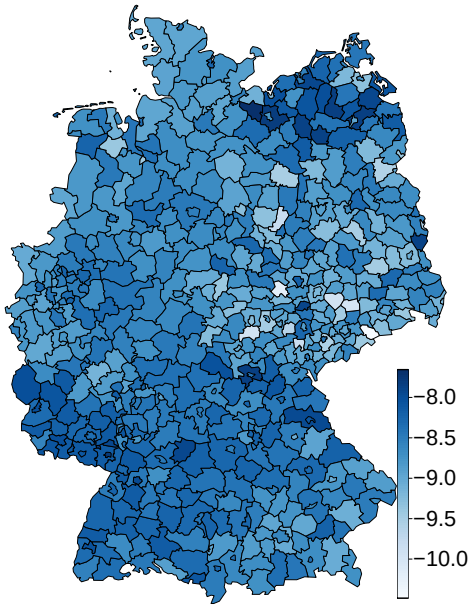
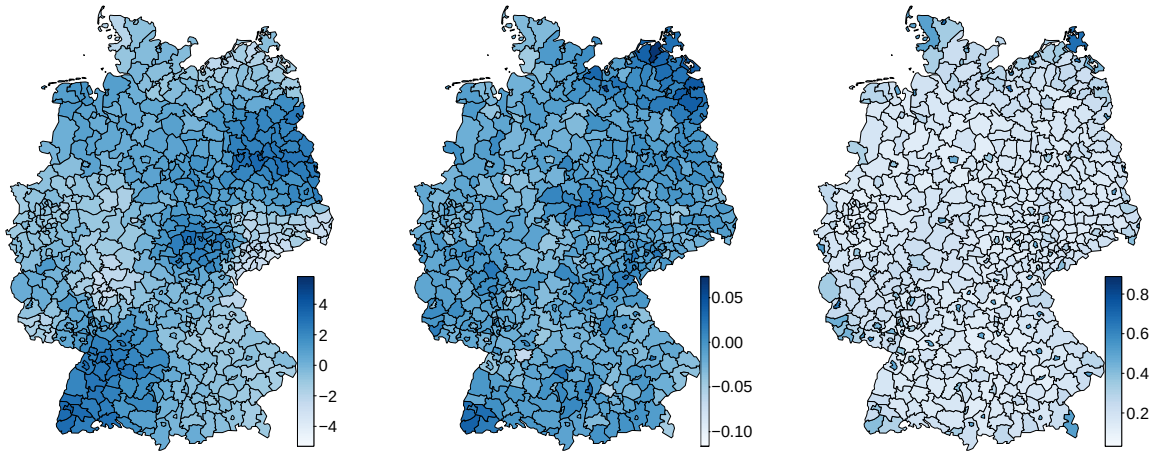


Figure 8.1: Log IRs

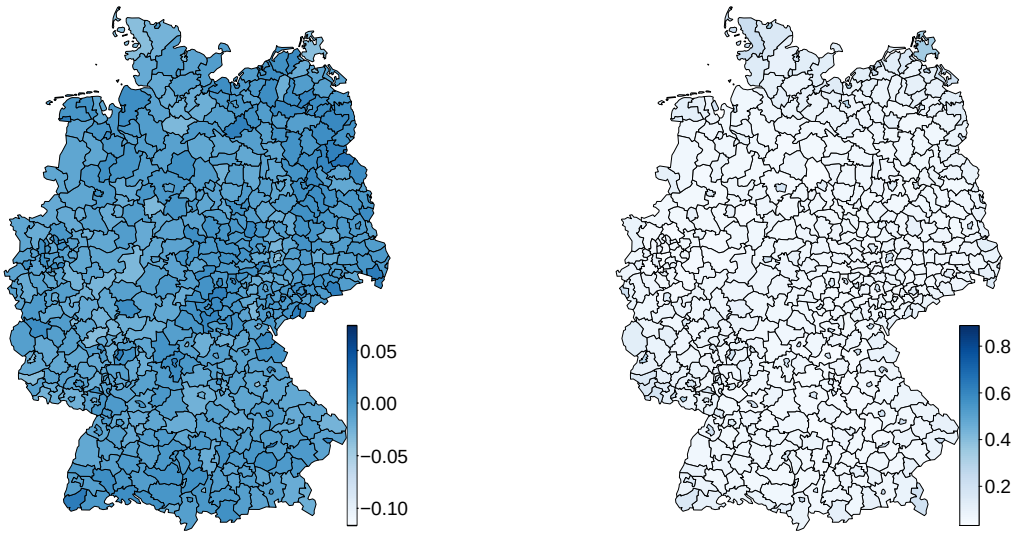
Coming to posterior variances in Figure 8.3c, three-fourth counties (408 out of 544) have a value of 0.25 or less. The counties which have higher variance than 0.25 are generally those which do not have many neighbors, they are either on the borders or they have so small area that their polygon is completely surrounded by a big county around them i.e., they are conditionally independent from all the counties except from the one surrounding them. Such counties do not borrow much information from other counties which explains their higher variation.



(a) $\boldsymbol{x}$ (b) $\hat{\boldsymbol{x}}$ (c) Var($\hat{\boldsymbol{x}}$)

Figure 8.3: Smoothed GMRF with their posterior means ($\hat{\boldsymbol{x}}$) and variances for incidences of Oral cancer

As a comparison, we have included another set of smoothing parameters from Level-3 of Scenario-I (Section 7.1.1) and then compare the results from our smoothed spatial model with the simple model of GMRF (commonly known as Besag-York-Mollie or BYM model of Besag et al. (1991)). It is to be noted that our model reduces to BYM model if all the regional smoothing parameters are set to either zero or to same constant. The results from increased smoothing are shown in Figure 8.4 and from BYM model are shown in Figure 8.5. The posterior means and variances of BYM model are given in Figures 8.5a and 8.5b. The global mean of true observations is 0.2584, while the global mean of posterior means is –0.0227. The Figures 8.3 to 8.5 are on the same scale, comparing them reveals that our proposed model is superior in smoothing variations at both regional and global levels. Posterior variances for smoothed spatial model are significantly low, only the counties on the borders have a little higher variation.

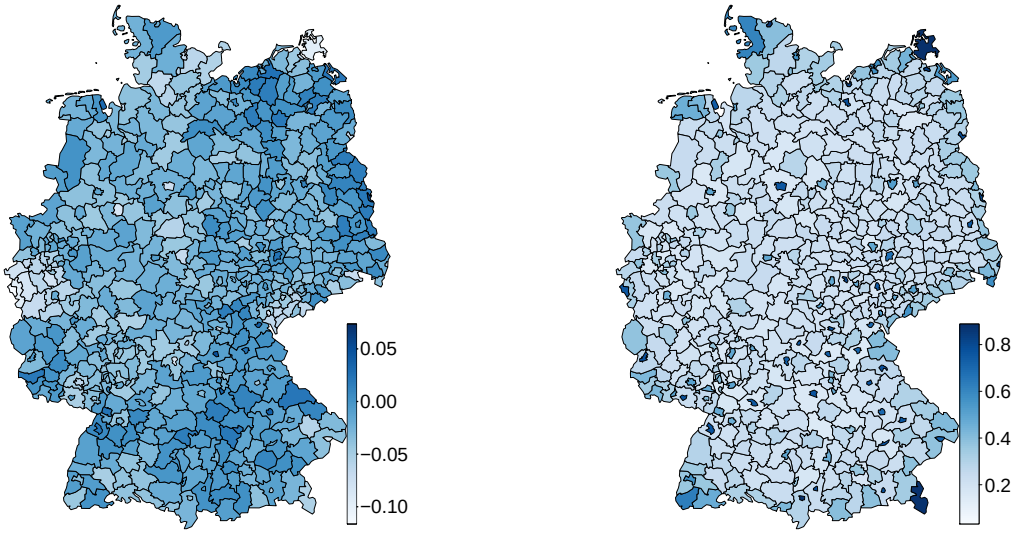


(a) $\hat{\boldsymbol{x}}$ (b) Var($\hat{\boldsymbol{x}}$)

Figure 8.4: Smoothed GMRF (higher levels of parameters) with their posterior means ($\hat{\boldsymbol{x}}$) and variances for incidences of Oral cancer

8.1.2 Clustering of Alike Spatial Units

The real data of Oral cancer incidences is used to find clusters exhibiting alike risks. The number of incidences y_i are given for each county and expected incidences e_i are obtained using $N_i \times \pi_i$, where π_i is the probability of risk in binomial density (4.1). Following the procedure described in Section 6.3, we first elicit different configurations of clusters from data on three previous time points. Since the data from previous time points is not available, therefore, some random noise is added to current data to depict incidences from previous time points. For the first previous time point, uniform random noise $U(-0.3, 0.3)$ is added to



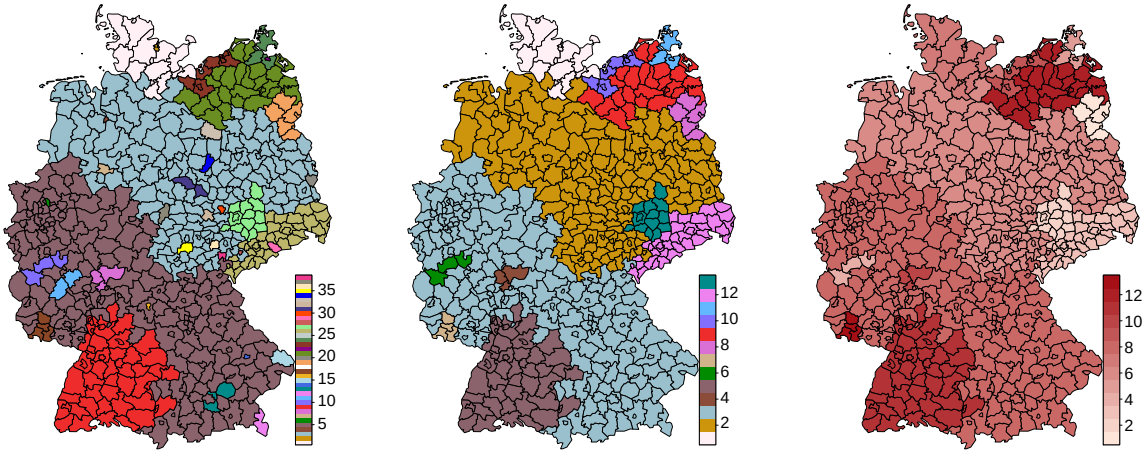
(a) $\hat{\boldsymbol{x}}$ (b) Var($\hat{\boldsymbol{x}}$)

Figure 8.5: Posterior means of simple GMRF (BYM model) ($\hat{\boldsymbol{x}}$) along with variances for incidences of Oral cancer

y_i and $U(-50, 50)$ is added to e_i. For the second previous time point, $U(-0.6, 0.6)$ is added to y_i and $U(-100, 100)$ is added to e_i. Finally, $U(-0.9, 0.9)$ is added to y_i and $U(-150, 150)$ is added to e_i for the third previous time point. The ratios (y_i/e_i) are obtained from data sets of previous time points to create different configurations of clusters.

Figure 8.6a displays the 38 identified clusters based on risk levels of Oral cancer. It can be seen that there are some clusters which contain only one or two counties and this is much likely that these tiny clusters could be a part of some big cluster around them. Therefore, it is appropriate to merge these single and/or small clusters with their big neighboring clusters by restricting the cluster size to at least three counties, which seems a somewhat reasonable minimum size for current data. There are a total of 25 such small clusters which are merged with their big neighboring clusters, making a total of 13 final clusters, which are displayed in Figure 8.6b.

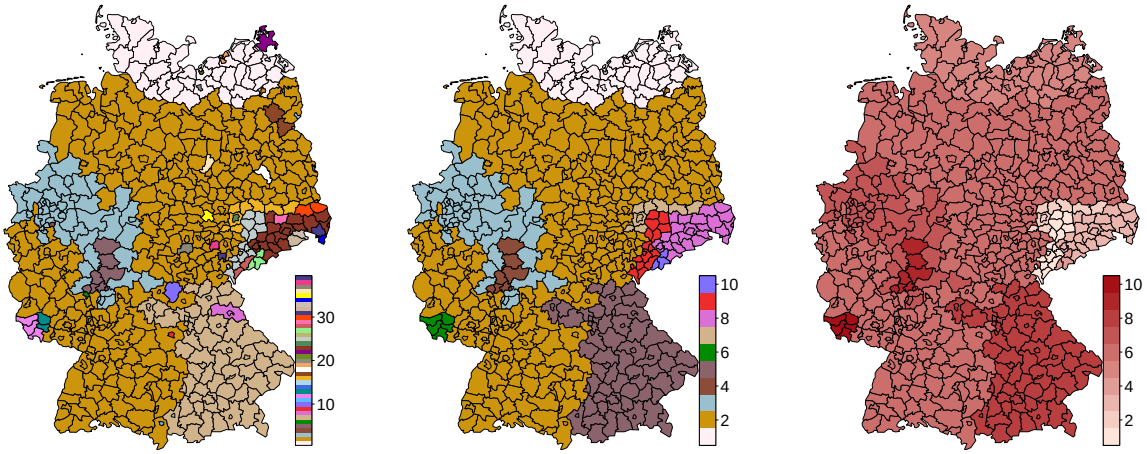
The clusters identified based on risk levels of Oral cancer only tell us that all the spatial units within each cluster have alike risks but they do not show whether a particular cluster is least/most affected than the other clusters. In order to compare them based on their risk levels, we obtain cluster based IRs to see how the clusters are ranked, as discussed in Section 6.4. The ranking of identified clusters in Figure 8.6b is given in Figure 8.6c, a value of 1 indicates that the cluster is least affected of all and a value of 13 means that the cluster is most affected. It has been found that cluster 7 is the most affected, followed by clusters 9 and 5 and so on until the least affected cluster 8. Cluster 7 contains the counties from Saarland on the south-west border with France, cluster 9 contains counties from Mecklenburg-Vorpommern on the north-east border with Poland and cluster 5 has a



(a) Clusters (original) (b) Final clusters (c) Ranking of clusters

Figure 8.6: Risk clusters from smoothed model for Oral cancer

large portion of Baden-Württemberg again with the border of France. The assumptions made in Section 6.2 are therefore true that most affected clusters are either along the border of France or are in the north east Germany.



(a) Clusters (original) (b) Final clusters (c) Ranking of clusters

Figure 8.7: Risk clusters without smoothing for Oral cancer

For comparison with the BYM model, we are able to identify 39 clusters (Figure 8.7a) using their model. After merging small clusters (one or two-county clusters) with their big neighboring clusters, a final list of 10 clusters is shown in Figure 8.7b with their ranking in Figure 8.7c. Cluster 6 has been found the most affected, followed by clusters 4 and 5 and so on until the least affected cluster 9. Cluster 6 contains the counties from Saarland on the

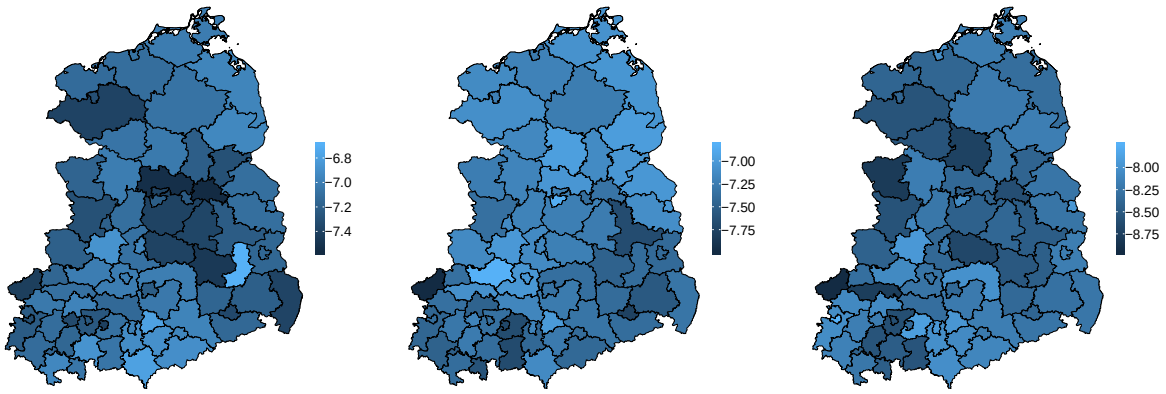
south-west border with France, cluster 5 contains counties from Bayern on the south-east border with Austria and Czech Republic and cluster 4 has counties from Hessen. The top four least affected clusters all lie in Sachsen.

The clusters identified with our regionalized spatial model highlight areas which are linked to higher risk of Oral cancer in the literature (Knorr-Held and Raßer (2000)), whereas the BYM model identifies only a part of those areas. Further, we rank the identified clusters in a systematic way according to severity of risk.

8.2 Multivariate Application

8.2.1 Regionalized Smoothing

The real data is taken from German Joint Cancer Register (GKR-Krebsatlas) for the year 2014 against the Colon, Lung and Pancreatic cancers. The affected individuals belong to 77 counties of eastern Germany in the states of Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachne-Anhalt and Thüringen. A quick look of the data tells us that per county at least 28 and at most 1,793 individuals are affected by Colon cancer ($y_{\cdot 1}$), between 20 and 2,285 are affected by Lung cancer ($y_{\cdot 2}$), and between 8 and 634 are affected by Pancreatic cancer ($y_{\cdot 3}$). A total of 11,885 people are affected by Colon cancer, 11,004 are affected by Lung cancer and 3,841 are affected by Pancreatic cancer. The population size per county, N_i, ranges from 36,208 to 3,469,849. The log of IRs (y_{ij}/N_i) for three cancers are given in Figure 8.8. The map reveals that the incidences are not homogeneously distributed across counties. The variation present in the maps is due to different factors such as different administrative set ups, heterogeneous population etc., please refer to discussion in Section 6.1.1.



(a) Colon (b) Lung (c) Pancreatic

Figure 8.8: Log IRs of Colon, Lung and Pancreatic cancers for the year 2014

To apply regional smoothing, states are classified into categories based on their counties. Figure 8.9 displays the six categories, states 4 and 5 have almost same number of counties, 13

counties in state 4 and 14 counties in state 5, therefore, we apply same amount of smoothing to them i.e., $r_4 = r_5 = 0.3$. The values for other states are $r_1 = 0.1, r_2 = 0.4, r_3 = 0.2$ and $r_6 = 0.5$. The global smoothing parameter s is set to 0.6.

The true values of multivariate GMRFs from regionalized spatial smoothing along side posterior means and posterior variances are given in Figure 8.10. True values are represented by $\circ$, their posterior means by $\times$ and log of posterior variances are shown by $\bullet$. The posterior mean of s for Colon cancer is 0.5997 with 0.01 posterior variance, for Lung cancer $\hat{s} = 0.6004$ with variance 0.02 and $\hat{s} = 0.6002$ with 0.03 variance for Pancreatic cancer.

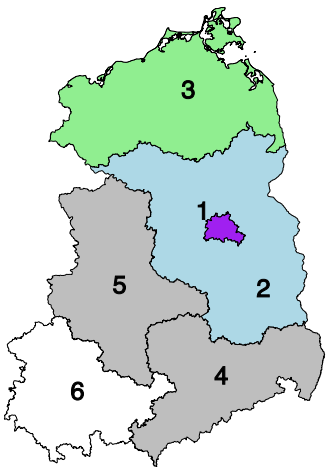


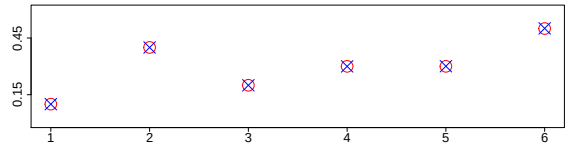
Figure 8.9: Classification of states for regional smoothing

The maps of $\hat{\boldsymbol{x}}$ and their posterior variances are shown in Figure 8.11. The global mean of all the 77 observations for Colon cancer is 1.0798, for Lung cancer it is -1.6880 and for Pancreatic cancer it is 0.7006. We can see that some counties have means as low as -4 and some have as more as 4. The posterior means in Figures 8.11d to 8.11f show that the proposed model has reduced such local variations by applying regional smoothing on each state. The global mean of posterior means for Colon cancer is 0.0028, for Lung cancer is -0.0170 and 0.0052 for Pancreatic cancer. Most of the counties after smoothing in Figures 8.11d to 8.11f have their mean values close to global mean.

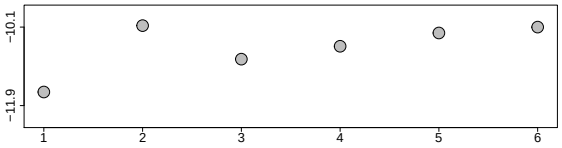
The posterior variances in Figures 8.11g to 8.11i for all the cancer fall below 0.0500. Only a few counties show relatively higher variance, they are generally those counties which do not have many neighbors.

As a comparison, we have included another set of smoothing parameters ($r_1 = 0.3, r_2 = 1.2, r_3 = 0.6, r_4 = r_5 = 0.9, r_6 = 1.5$ and $s = 1.8$) and then compare the results from our smoothed spatial model with the simple model of GMRF (BYM model). The results from increased smoothing are shown in Figure 8.12 and from BYM model are shown in Figure 8.13. The posterior means and variances of BYM model are given in Figures 8.13d to 8.13i.

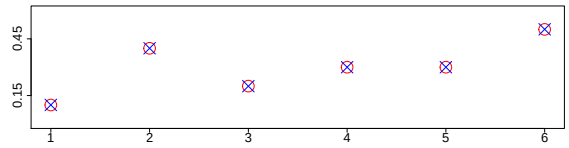
The posterior estimates look equally smooth from both the regionalized and simple GMRF model with almost same amount of variation in posterior means. However the difference becomes more clear when we perform identification of clusters on these models, we will see in the next section that the simple GMRF model does not properly identify clusters, atleast not for Colon and Lung cancers using current data.



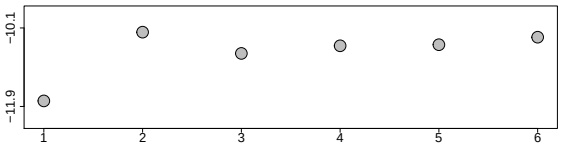
(a) True values ($\boldsymbol{r}^{(1)}$) and their posterior means ($\hat{\boldsymbol{r}}^{(1)}$)



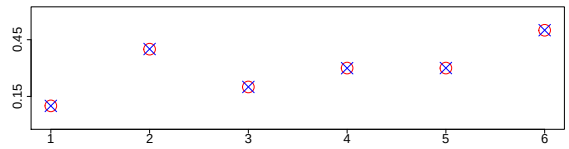
(b) Posterior variances of $\hat{\boldsymbol{r}}^{(1)}$ on log scale



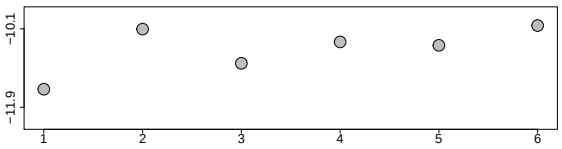
(c) True values ($\boldsymbol{r}^{(2)}$) and their posterior means ($\hat{\boldsymbol{r}}^{(2)}$)



(d) Posterior variances of $\hat{\boldsymbol{r}}^{(2)}$ on log scale



(e) True values ($\boldsymbol{r}^{(3)}$) and their posterior means ($\hat{\boldsymbol{r}}^{(3)}$)



(f) Posterior variances of $\hat{\boldsymbol{r}}^{(3)}$ on log scale

Figure 8.10: Smoothing parameters for Colon, Lung and Pancreatic cancers

8.2.2 Clustering of Alike Spatial Units

Following the procedure explained in Section 6.3, the risk clusters are identified for real data sets of Colon, Lung and Pancreatic cancers. The observed incidences are given for each county and expected incidences are computed using $N_i \times \pi_{ij} : j = 1, 2, 3$, where π_{ij} is the probability of risk in binomial density (5.1). The algorithm is applied separately on each cancer data. The data from previous time points is not available, therefore, a uniform random noise is added to current data to depict incidences from previous time points. For the Colon cancer, uniform random noises $U(-2, 2), U(-4, 4)$ and $U(-6, 6)$ are added to y_{i1} and $U(-200, 200), U(-400, 400)$ and $U(-600, 600)$ are added to e_{i1} for three previous time point data sets. Similarly, for the Lung cancer, $U(-2, 2), U(-4, 4)$ and $U(-6, 6)$ are added to y_{i2} and $U(-10, 10), U(-15, 15)$ and $U(-20, 20)$ are added to e_{i2}, For the Pancreatic cancer, $U(-1, 1), U(-2, 2)$ and $U(-3, 3)$ are added to y_{i3}, and $U(-100, 100), U(-200, 200)$ and $U(-300, 300)$ are added to e_{i3}. Incidence risks $(y_{ij}/e_{ij} : j = 1, 2, 3)$ are obtained from data sets of previous time points to create different configurations of clusters.

A total of 12 clusters are identified for Colon cancer, 13 are identified for Lung cancer and 15 are identified for Pancreatic cancer, they are shown in (Figures 8.14a to 8.14c). However, there are some singleton clusters with only one county, we merge them with their

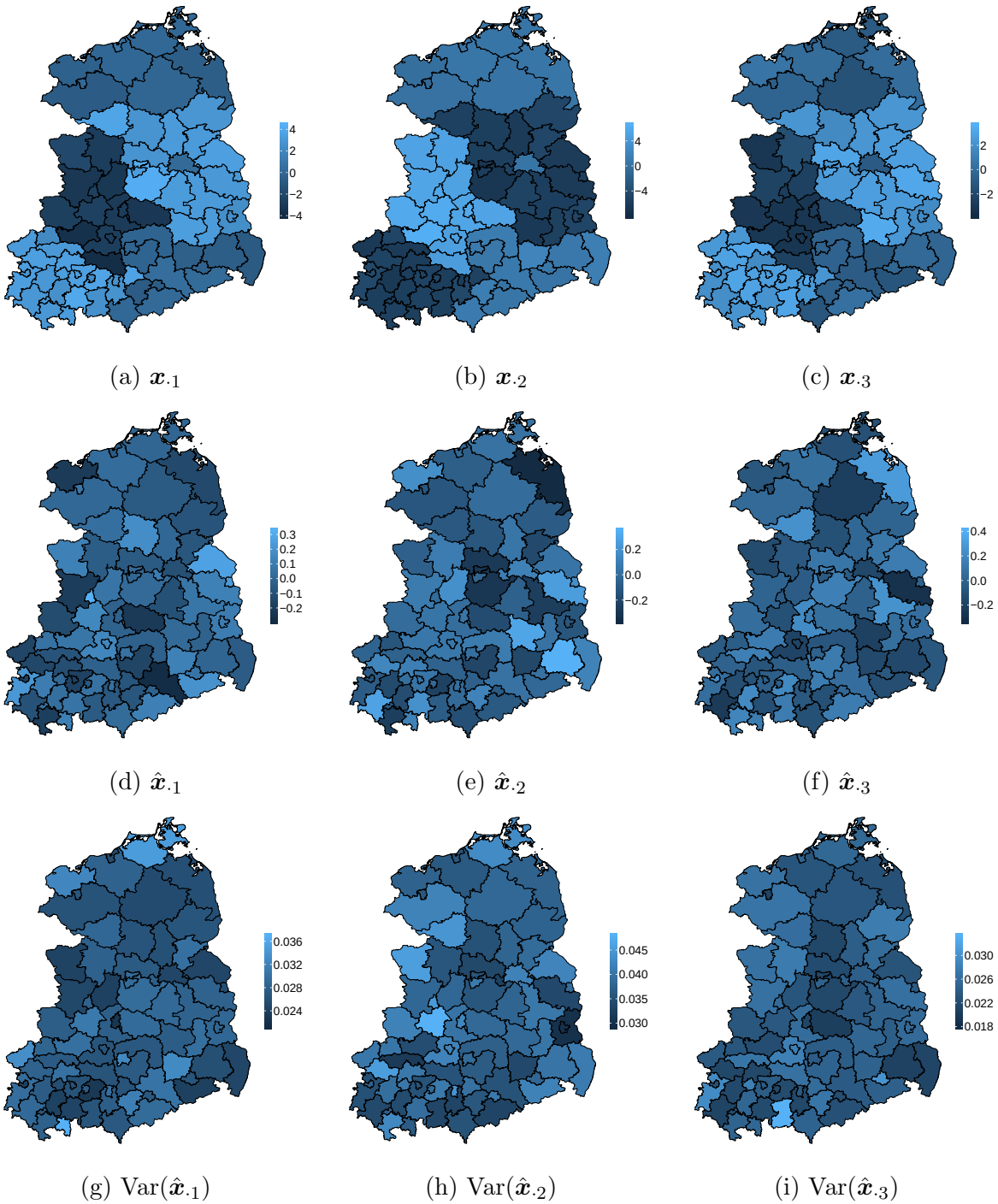


Figure 8.11: Smoothed GMRFs with their posterior means and variances for Colon ($\boldsymbol{x}_{\cdot 1}$), Lung ($\boldsymbol{x}_{\cdot 2}$) and Pancreatic ($\boldsymbol{x}_{\cdot 3}$) cancers

big neighboring clusters. After merging, we find 9 clusters for Colon, 6 for Lung and 8 for Pancreatic cancer, they are shown in (Figures 8.14d to 8.14f).

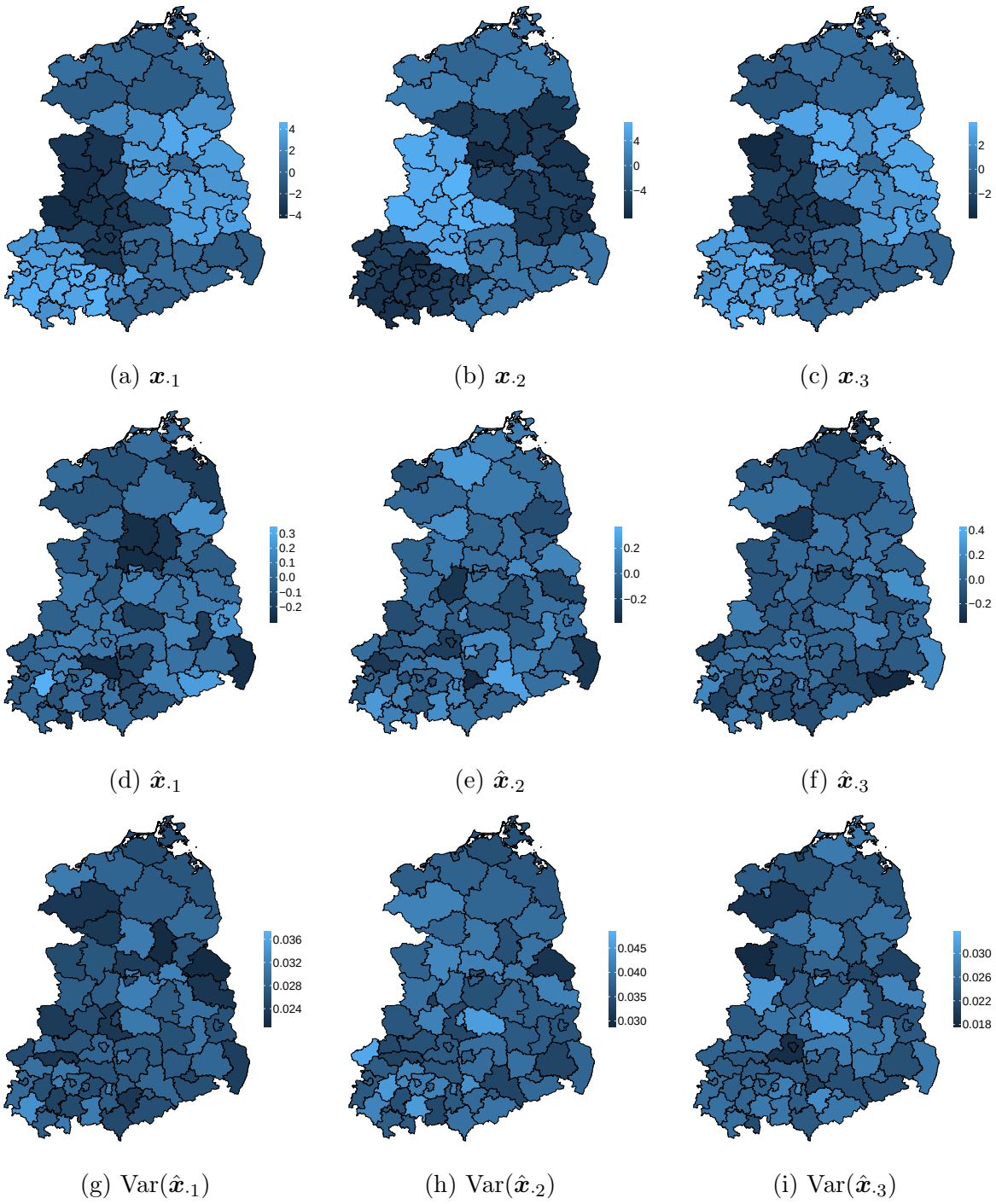


Figure 8.12: Smoothed GMRFs (higher levels of parameters) with their posterior means and variances for Colon ($\boldsymbol{x}_{\cdot 1}$), Lung ($\boldsymbol{x}_{\cdot 2}$) and Pancreatic ($\boldsymbol{x}_{\cdot 3}$) cancers

The clusters identified based on risk levels only tell us that all the counties within each cluster have alike risks but they do not show which cluster is affected badly by any of the

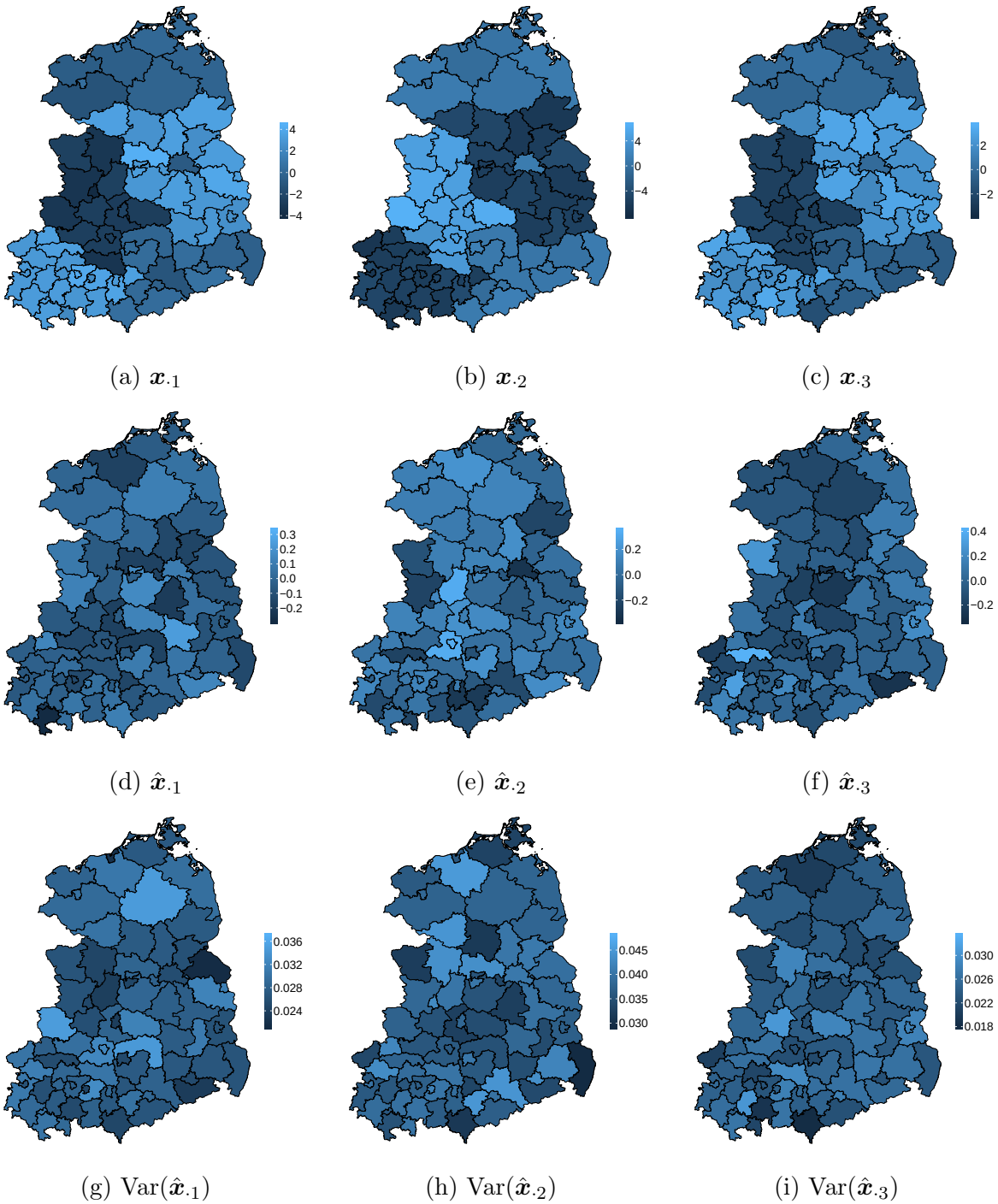
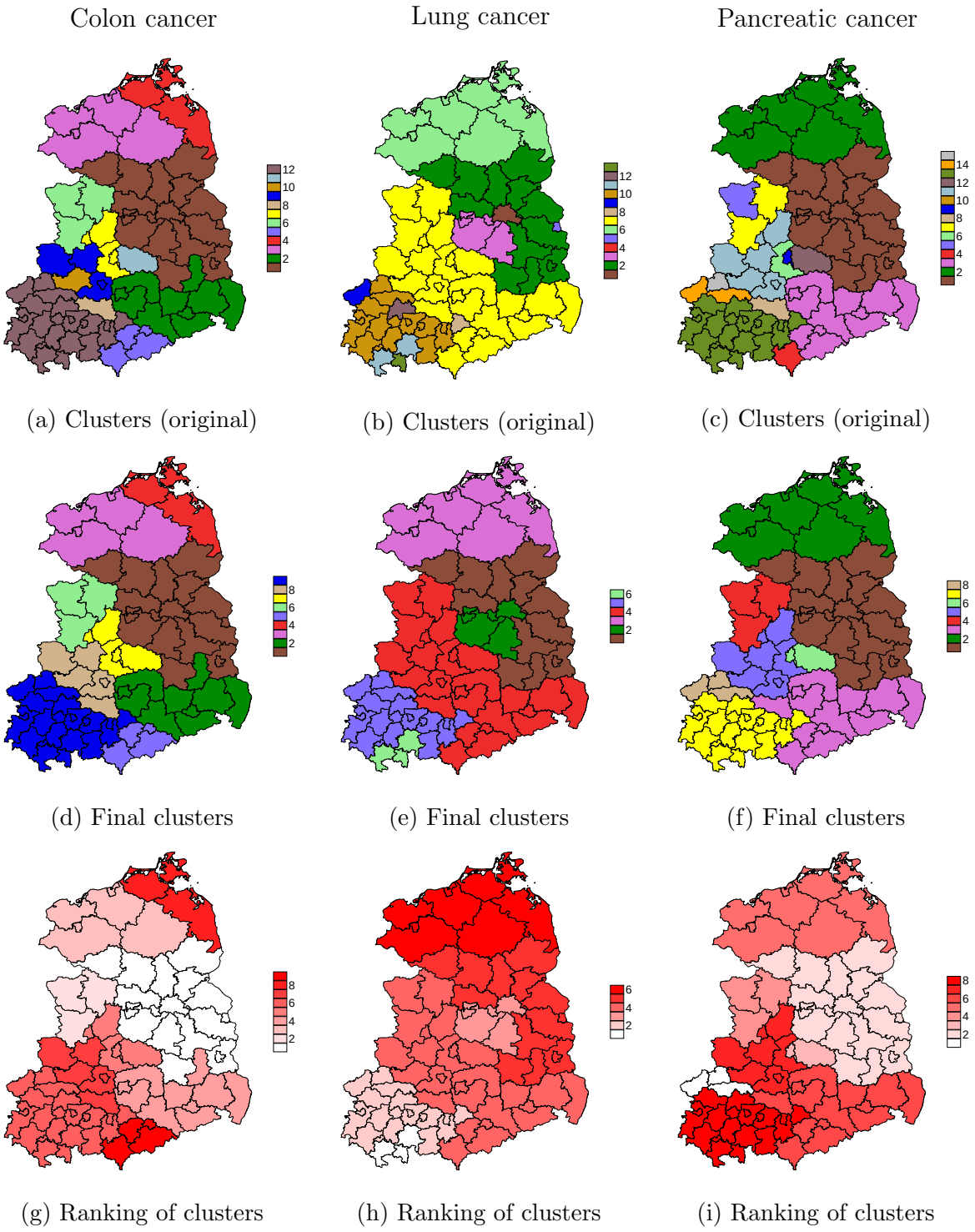


Figure 8.13: Simple GMRFs (BYM Model) without smoothing, their posterior means and variances for incidences of Colon ($\boldsymbol{x}_{\cdot 1}$), Lung ($\boldsymbol{x}_{\cdot 2}$) and Pancreatic ($\boldsymbol{x}_{\cdot 3}$) cancers

disease. We compare them based on their risk levels, we obtain cluster based IRs to see how the clusters are ranked (Section 6.4 for details). The ranking of final clusters is given



(a) Clusters (original) (b) Clusters (original) (c) Clusters (original)

(d) Final clusters (e) Final clusters (f) Final clusters

(g) Ranking of clusters (h) Ranking of clusters (i) Ranking of clusters

Figure 8.14: Risk clusters from smoothed model for Colon, Lung and Pancreatic cancers

in Figures 8.14g to 8.14i, a value of 1 indicates that the cluster is least affected of all and a

higher value indicates that the cluster is more affected.

It has been found that parts of Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt and Thüringia are mostly affected with Colon and Pancreatic cancers, while the Lung cancer has affected a large part of north-east Germany leaving only Thüringia the least affected.

For comparison with the BYM model, we are able to identify 3 clusters of Colon cancer and 8 clusters of Pancreatic cancer (Figures A.17a and A.17c) using their model. However their approach does not identify any cluster in case of Lung cancer (Figure A.17b). After merging small clusters (one or two-county clusters) for Colon and Pancreatic cancers, 2 clusters of Colon cancer are finalized and 8 clusters of Pancreatic cancer are finalized (Figures A.17d and A.17f). Following Section 6.4, the ranking of these clusters is given in Figures A.17g and A.17i. The success of regionalized spatial model is evident from the identification of clusters. The Anderson et al. (2014) approach successfully identifies risk clusters when applied on the regionally smoothed data in contrast to BYM model.

Chapter 9

Discussion

Spatial heterogeneity in disease maps requires specific treatment as standard CAR models fail to capture localized spatial correlation formed by adjacent regions. The local regions must be smoothed on local levels before applying any estimation techniques on such data. For this purpose, we consider spatial structure of German counties to analyze local spatial patterns of disease incidences. We propose a novel approach of local smoothing at various regional levels (Federal states) by incorporating smoothing parameters in spatial correlation matrix. We assume strong positive spatial correlation between adjacent regions on spatial structure. We divide the structure into 16 non-overlapping regions corresponding to Federal states and then a smoothing parameter is specified for each region. The degree of smoothing is decided based on the number of counties in each region.

In Chapters 4 and 5, we deal with the univariate and multivariate spatial models. We suppose that number of incidences in each county follows a binomial density with risk probability depending on some coefficient β_i through a logit function. The spatial structure of German counties is utilized in form of prior density of this coefficient. The novel method of incorporating regional smoothing parameters ($\boldsymbol{r}$) into correlation matrix is discussed. We explain the use of auxiliary variables to ease MCMC sampling for drawing from posterior density.

The simulation study is conducted using various levels of regional and global smoothing parameters. The results and discussions from Chapter 7 show that the proposed method successfully smoothes local heterogeneities and helps in flattening the risk across regions.

The proposed method is applied to two data sets; on Oral cancer incidences in Germany for univariate regionalized spatial model and on Colon, Lung and Pancreatic cancers in north eastern part of Germany for multivariate regionalized spatial model. We see in univariate regionalized spatial model that 12 out of 16 states have mean values equal to global mean value of all the observations. Posterior variances of only a quarter of counties are noticeable. The analysis of real data reveals that the proposed novel approach smoothes local level means to global mean. Posterior variances per county are high only for counties which either lie on the borders or are completely surrounded by a big county around them making them significantly less correlated with other counties.

We compare our method of regionalized smoothing with simple model of GMRF (Besag-

York-Mollie or BYM model of Besag et al. (1991)). Figures 8.3 to 8.5 present estimates from the univariate models. Comparing them reveals that our proposed model is superior in smoothing variations at both regional and global levels. Posterior variances for smoothed spatial model are significantly low. Figures 8.11 to 8.13 present estimates from multivariate regionalized and multivariate BYM models. The posterior estimates look equally smooth from both the model with almost same amount of variation in posterior means. However, when we perform identification of clusters on these models, the simple GMRF model does not properly identify clusters, atleast not for Colon and Lung cancers using current data.

A number of factors contribute to spatial heterogeneities in risk such as heterogeneous population, different administrative structures, etc. therefore, we propose a regionalized spatial model which smoothes extreme observations locally to flatten the heterogeneities. The benefit of smoothing can additionally be utilized in the identification of risk clusters. The areas with similar risk can be identified easily after regionalized smoothing.

We utilize the method proposed by Anderson et al. (2014) and identify a total of 13 clusters in case of Oral cancer, 9 clusters of Colon, 6 clusters of Lung and 8 clusters of Pancreatic cancers. They are shown in Figures 8.6b and 8.14d to 8.14f. All the counties within these clusters presumably have similar risk i.e. they demonstrate that all the counties within a cluster are approximately equally affected by the respective cancer.

The most affected clusters from Oral cancer contains the counties from Saarland, Mecklenburg-Vorpommern and Baden-Württemberg. On the other hand, the BYM model (Besag et al. (1991)) identifies 10 clusters which covers the counties from Saarland, Bayern and Hessen. Both the approaches identify Sachsen as least effected. The clusters identified with our regionalized spatial model highlights areas which are linked to higher risk of Oral cancer in the literature (Knorr-Held and Raßer (2000)), whereas the BYM model identifies only a part of those areas.

The identified clusters are further ranked from least to most affected based on IRs. Figures 8.6c and 8.14g to 8.14i show the ranking of identified clusters. The success of regionalized smoothing is evident from the identification of clusters in multivariate model. When the Anderson et al. (2014) approach is applied to smoothed data and the BYM model for comparison, only two clusters of Colon cancer are identified with the BYM model but 9 clusters are identified on smoothed data. Further, no cluster of Lung cancer is identified on BYM model, in contrast to 6 clusters on smoothed data (Figures 8.14 and A.17).

Finally, there are some areas to which the proposed method of smoothing can be extended. For example, we use a uniform smoothing within a region, asymmetric smoothing based on geographic background and/or disease history can be used to regionally smooth spatial correlation matrix. Other areas include the challenge of introducing regional smoothing parameters to a similarity based non spatial precision matrix. Such precision matrices exist when no spatial positive auto-correlation is present in data, a similarity based non spatial precision matrix introduced in Baptista et al. (2016) may be useful in this direction.

Bibliography

Agarwal, A. and Daumé, H. (2010). A geometric view of conjugate priors. *Machine Learning*, 81(1):99–113.

Albert, J. H. and Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 88(422):669–679.

Anderson, C., Lee, D., and Dean, N. (2014). Identifying clusters in Bayesian disease mapping. *Biostatistics*, 15(3):457–469.

Anselin, L. (2001). Spatial econometrics. In Baltagi, B. H., editor, *A Companion to Theoretical Econometrics*, chapter 14. Blackwell Publishing Ltd.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in Regional Science*, 89(1):3–25.

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press.

Baptista, H., Mendes, J. M., MacNab, Y. C., Xavier, M., and Caldas-de Almeida, J. (2016). A Gaussian random field model for similarity-based smoothing in Bayesian disease mapping. *Statistical Methods in Medical Research*, 25(4):1166–1184.

Bayes, T. (1763). An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, FRS communicated by Mr. Price, in a letter to John Canton, AMFR S. *Philosophical Transactions of the Royal Society of London*, (53):370–418.

Becker, N. and Wahrendorf, J. (2013). *Krebsatlas der Bundesrepublik Deutschland/Atlas of cancer mortality in the Federal Republic of Germany 1981–1990*. Springer-Verlag.

Berger, J. O. (2013). *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media.

Bernardo, J. M. and Smith, A. F. (2009). *Bayesian Theory*, volume 405. John Wiley & Sons.

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*, 43(1):1–20.

Best, N. G., Arnold, R. A., Thomas, A., Waller, L. A., and Conlon, E. M. (1999). Bayesian models for spatially correlated disease and exposure data. In *Bayesian Statistics 6: Proceedings of the Sixth Valencia International Meeting*, volume 6, pages 131–156. Oxford University Press.

Botella-Rocamora, P., Martinez-Beneito, M. A., and Banerjee, S. (2015). A unifying modeling framework for highly multivariate disease mapping. *Statistics in Medicine*, 34(9):1548–1559.

Brooks, S., Gelman, A., Jones, G., and Meng, X.-L. (2011). *Handbook of Markov chain Monte Carlo*. CRC press.

Charras-Garrido, M., Abrial, D., Goër, J. D., Dachian, S., and Peyrard, N. (2012). Classification method for disease risk mapping based on discrete hidden Markov random fields. *Biostatistics*, 13(2):241–255.

Chen, Y. (2012). On the four types of weight functions for spatial contiguity matrix. *Letters in Spatial and Resource Sciences*, 5(2):65–72.

Congdon, P. D. (2020). *Bayesian Hierarchical Models: With Applications Using R*. CRC Press.

Cowles, M. K. (2004). Review of WinBUGS 1.4. *The American Statistician*, 58(4):330–336.

Cressie, N. (1993). *Statistics for spatial data: Wiley series in probability and Statistics*. Wiley–Interscience, New York.

Dean, C., Ugarte, M., and Militino, A. (2001). Detecting interaction between random region and fixed age effects in disease mapping. *Biometrics*, 57(1):197–202.

Frühwirth-Schnatter, S., Frühwirth, R., Held, L., and Rue, H. (2009). Improved auxiliary mixture sampling for hierarchical models of non-Gaussian data. *Statistics and Computing*, 19(4):479–492.

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press.

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Getis, A. (2009). Spatial weights matrices. *Geographical Analysis*, 41(4):404–410.

Hodges, J. S. (2013). *Richly parameterized linear models: additive, time series, and spatial models using random effects*. CRC Press.

Jeffreys, H. (1961). *Theory of probability*. Oxford University Press.

Jin, X., Banerjee, S., and Carlin, B. P. (2007). Order-free co-regionalized areal data models with application to multiple-disease mapping. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69(5):817–838.

Knorr-Held, L. and Raßer, G. (2000). Bayesian detection of clusters and discontinuities in disease maps. *Biometrics*, 56(1):13–21.

Kulldorff, M. (1997). A spatial scan statistic. *Communications in Statistics-Theory and Methods*, 26(6):1481–1496.

Laplace, P. S. (1812). Marquis de. *A Philosophical Essay on Probabilities*.

Lawson, A. B. (2018). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*. Chapman and Hall/CRC.

Lee, D. and Mitchell, R. (2013). Locally adaptive spatial smoothing using conditional auto-regressive models. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(4):593–608.

Leroux, B. G., Lei, X., and Breslow, N. (2000). Estimation of disease rates in small areas: a new mixed model for spatial dependence. In *Statistical models in epidemiology, the environment, and clinical trials*, pages 179–191. Springer.

LeSage, J. and Pace, R. K. (2009). *Introduction to spatial econometrics*. Chapman and Hall/CRC.

Lu, H., Reilly, C. S., Banerjee, S., and Carlin, B. P. (2007). Bayesian areal wombling via adjacency modeling. *Environmental and Ecological Statistics*, 14(4):433–452.

MacNab, Y. C. (2016). Linear models of coregionalization for multivariate lattice data: A general framework for coregionalized multivariate CAR models. *Statistics in Medicine*, 35(21):3827–3850.

Martínez-Beneito, M. A. (2013). A general modelling framework for multivariate disease mapping. *Biometrika*, 100(3):539–553.

Martínez-Beneito, M. A. and Botella-Rocamora, P. (2019). *Disease Mapping: From Foundations to Multidimensional Modeling*. CRC Press.

Matúš, F. (1992). On equivalence of Markov properties over undirected graphs. *Journal of Applied Probability*, 29(3):745–749.

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Richardson, S., Thomson, A., Best, N., and Elliott, P. (2004). Interpreting posterior relative risk estimates in disease-mapping studies. *Environmental Health Perspectives*, 112(9):1016–1025.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: Theory and applications*. CRC press.

Schaible, W. L. (2013). *Indirect estimators in US Federal programs*, volume 108. Springer Science & Business Media.

Simões, P. and Natário, I. (2016). Spatial econometric approaches for count data: An overview and new directions. *International Journal of Economics and Management Engineering*, 10(1):348–357.

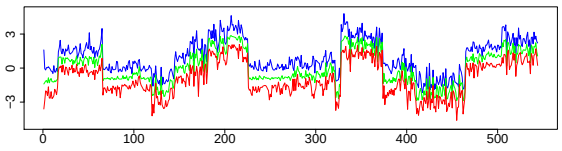
Stern, H. S. and Cressie, N. (2000). Posterior predictive model checks for disease mapping models. *Statistics in Medicine*, 19(17-18):2377–2397.

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46(sup1):234–240.

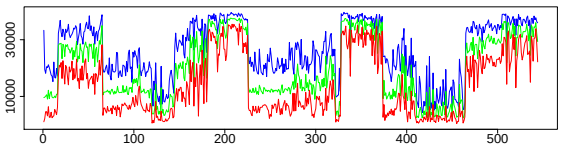
Tong, Y. L. (2012). *The multivariate normal distribution*. Springer Science & Business Media.

Wakefield, J. (2007). Disease mapping and spatial regression with count data. *Biostatistics*, 8(2):158–183.

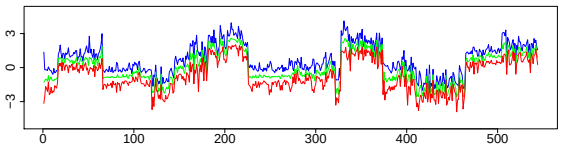
Appendix



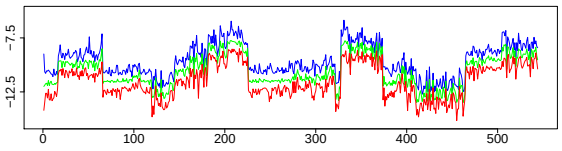
(a) $\boldsymbol{x}$



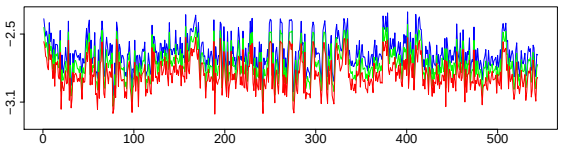
(b) $\boldsymbol{y}$



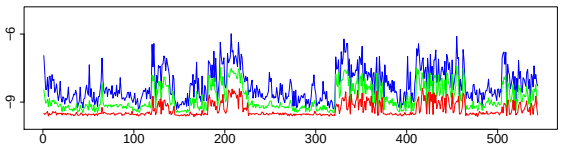
(c) $\hat{\boldsymbol{x}}$



(d) $\hat{\boldsymbol{y}}^*$

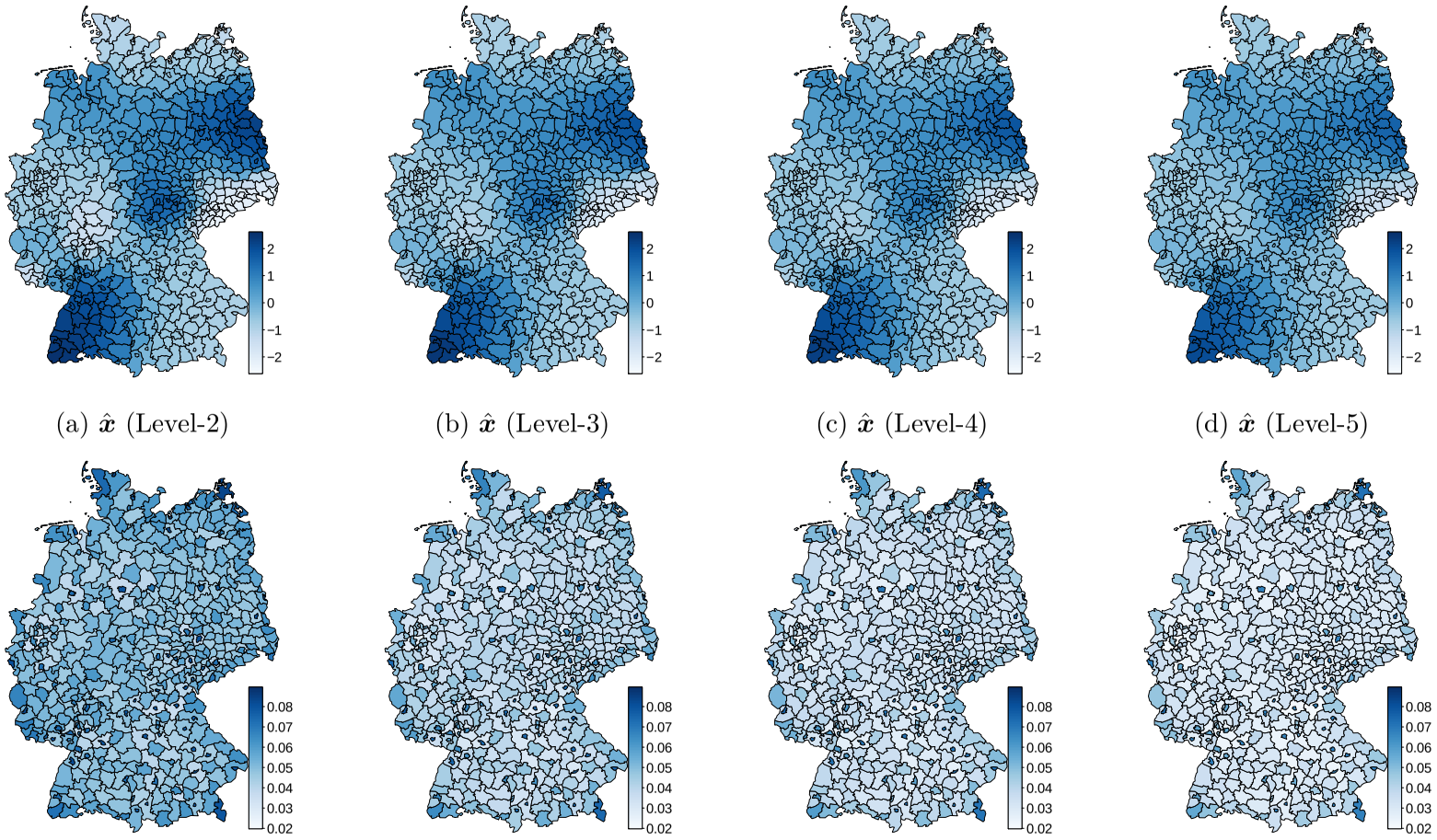


(e) Log of posterior variances of $\hat{\boldsymbol{x}}$



(f) Log of posterior variances of $\hat{\boldsymbol{y}}^*$

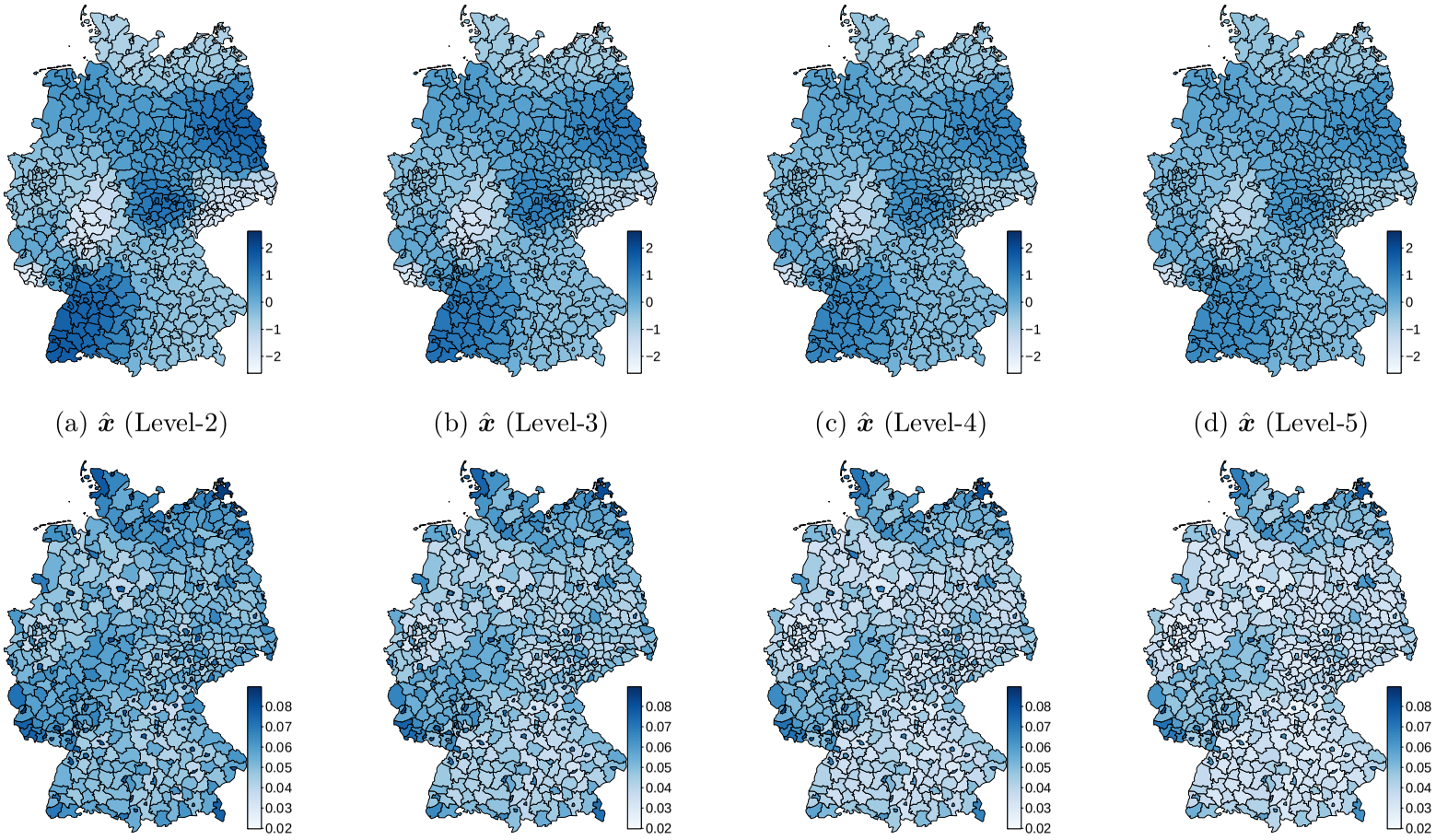
Figure A.1: Variation in $\boldsymbol{x}$, $\boldsymbol{y}$, $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}^*$ with their posterior variances on log scale



(a) $\hat{\boldsymbol{x}}$ (Level-2) (b) $\hat{\boldsymbol{x}}$ (Level-3) (c) $\hat{\boldsymbol{x}}$ (Level-4) (d) $\hat{\boldsymbol{x}}$ (Level-5)

(e) Var($\hat{\boldsymbol{x}}$) (Level-2) (f) Var($\hat{\boldsymbol{x}}$) (Level-3) (g) Var($\hat{\boldsymbol{x}}$) (Level-4) (h) Var($\hat{\boldsymbol{x}}$) (Level-5)

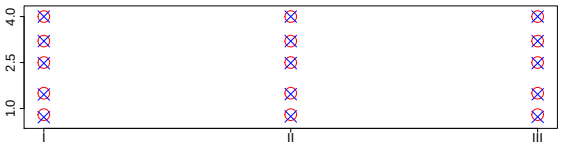
Figure A.2: Posterior means and variances of $\hat{\boldsymbol{x}}$ for various levels of smoothing (Scenario-II)



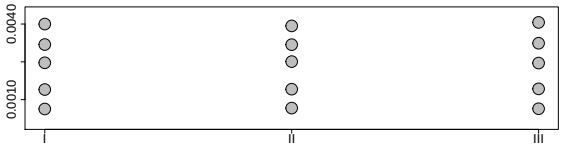
(a) $\hat{\boldsymbol{x}}$ (Level-2) (b) $\hat{\boldsymbol{x}}$ (Level-3) (c) $\hat{\boldsymbol{x}}$ (Level-4) (d) $\hat{\boldsymbol{x}}$ (Level-5)

(e) Var($\hat{\boldsymbol{x}}$) (Level-2) (f) Var($\hat{\boldsymbol{x}}$) (Level-3) (g) Var($\hat{\boldsymbol{x}}$) (Level-4) (h) Var($\hat{\boldsymbol{x}}$) (Level-5)

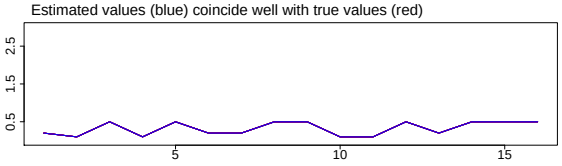
Figure A.3: Posterior means and variances of $\hat{\boldsymbol{x}}$ for various levels of smoothing (Scenario-III)



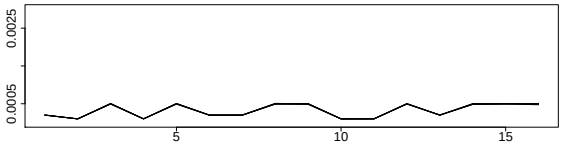
(a) True values $(s^{(1)}, s^{(2)}, s^{(3)})$ and posterior means $(\hat{s}^{(1)}, \hat{s}^{(2)}, \hat{s}^{(3)})$



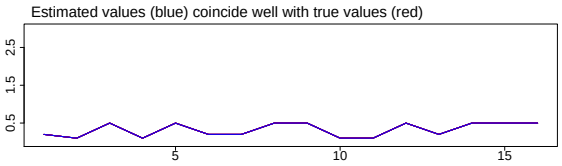
(b) Posterior variances of $\hat{s}^{(1)}, \hat{s}^{(2)}$ and $\hat{s}^{(3)}$



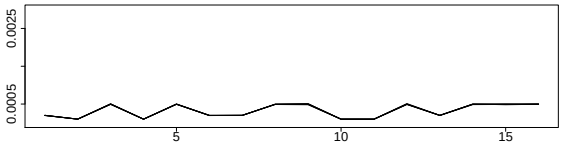
(c) True values $(\boldsymbol{r}^{(1)})$ and posterior means $(\hat{\boldsymbol{r}}^{(1)})$



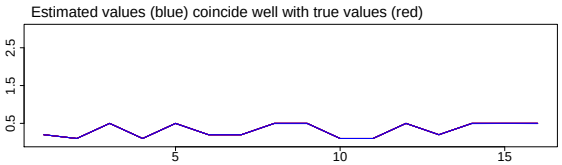
(d) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(1)}))$



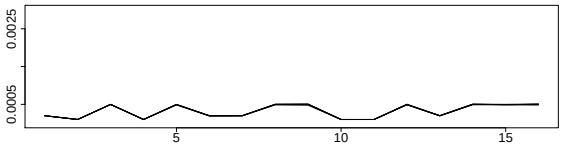
(e) True values $(\boldsymbol{r}^{(2)})$ and posterior means $(\hat{\boldsymbol{r}}^{(2)})$



(f) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(2)}))$

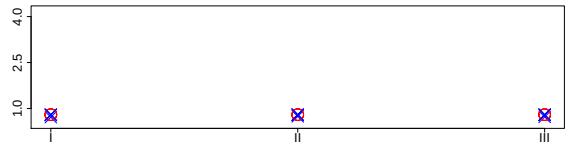


(g) True values $(\boldsymbol{r}^{(3)})$ and posterior means $(\hat{\boldsymbol{r}}^{(3)})$

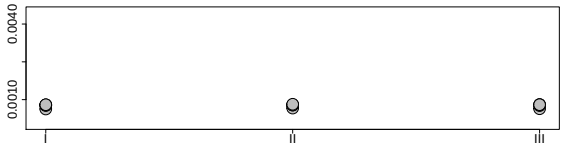


(h) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(3)}))$

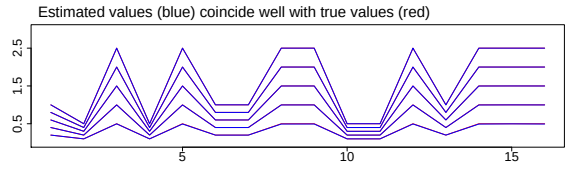
Figure A.4: Smoothing parameters under Scenario-II



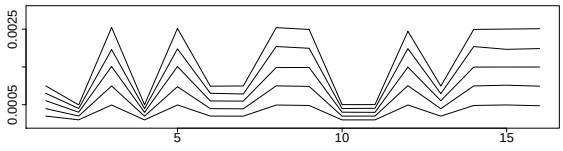
(a) True values $(s^{(1)}, s^{(2)}, s^{(3)})$ and posterior means $(\hat{s}^{(1)}, \hat{s}^{(2)}, \hat{s}^{(3)})$



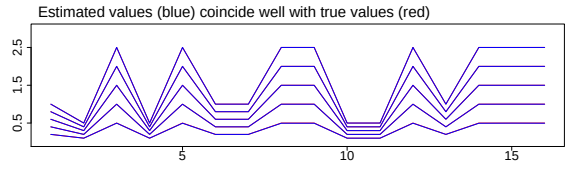
(b) Posterior variances of $\hat{s}^{(1)}, \hat{s}^{(2)}$ and $\hat{s}^{(3)}$



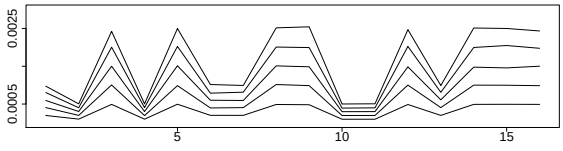
(c) True values $(\boldsymbol{r}^{(1)})$ and posterior means $(\hat{\boldsymbol{r}}^{(1)})$



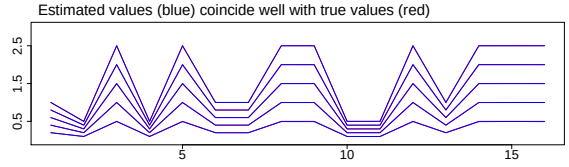
(d) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(1)}))$



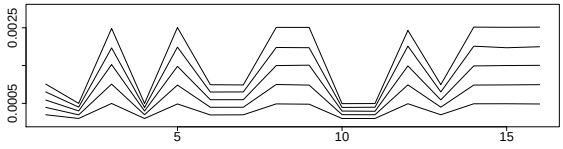
(e) True values $(\boldsymbol{r}^{(2)})$ and posterior means $(\hat{\boldsymbol{r}}^{(2)})$



(f) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(2)}))$

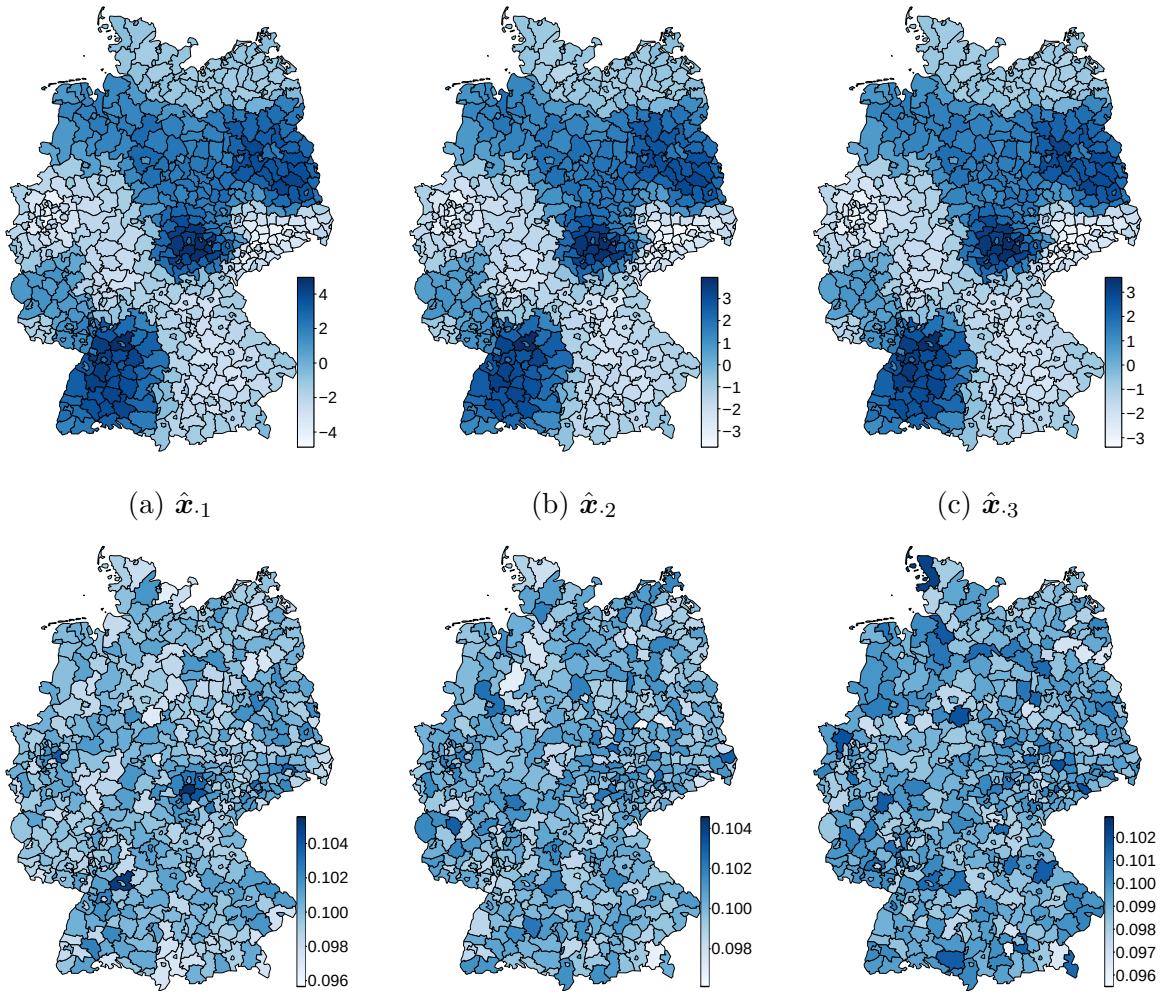


(g) True values $(\boldsymbol{r}^{(3)})$ and posterior means $(\hat{\boldsymbol{r}}^{(3)})$



(h) Posterior variances $(Var(\hat{\boldsymbol{r}}^{(3)}))$

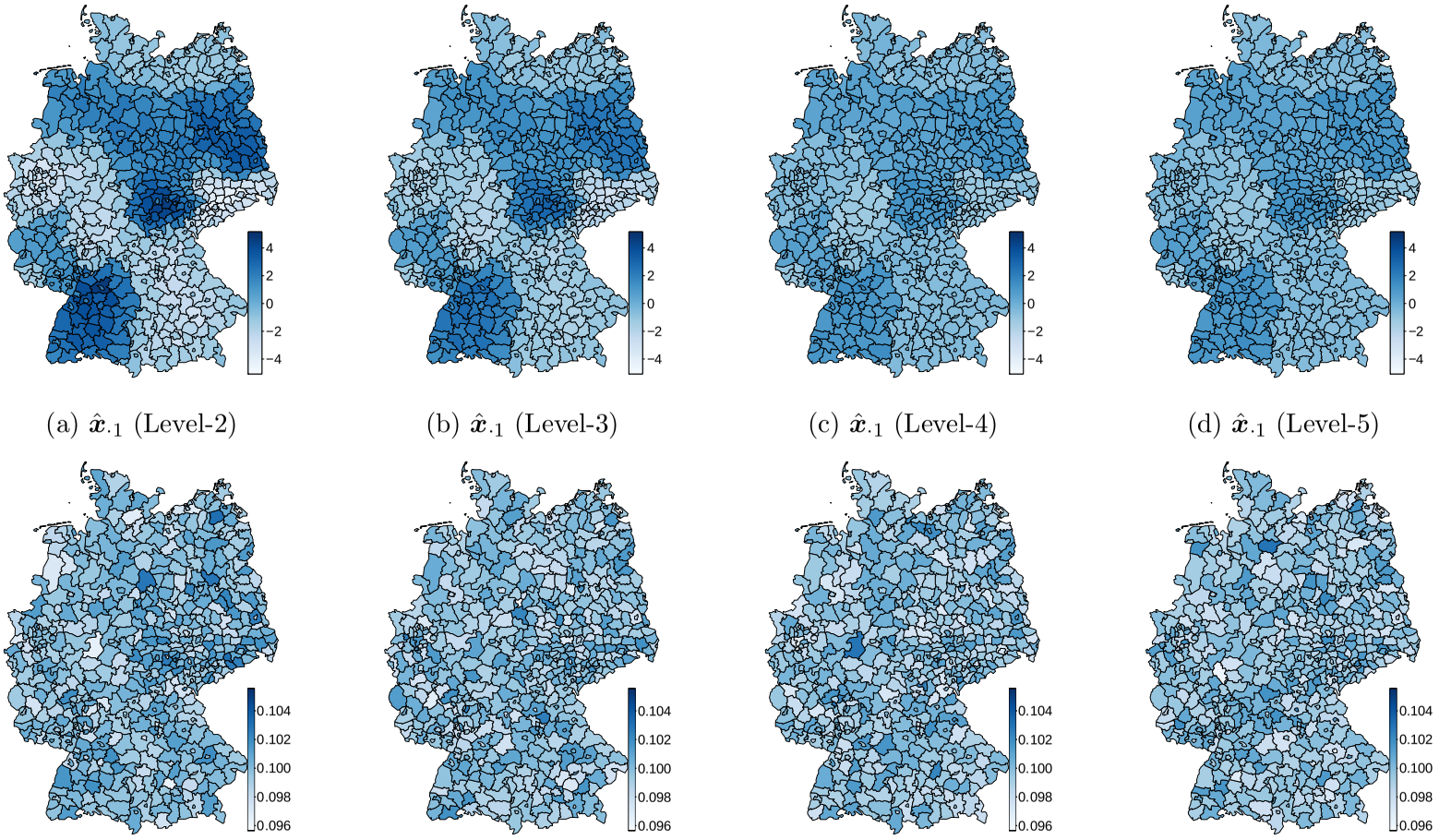
Figure A.5: Smoothing parameters under Scenario-III



(a) $\hat{\boldsymbol{x}}_{\cdot 1}$ (b) $\hat{\boldsymbol{x}}_{\cdot 2}$ (c) $\hat{\boldsymbol{x}}_{\cdot 3}$

(d) $\mathrm{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (e) $\mathrm{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (f) $\mathrm{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$

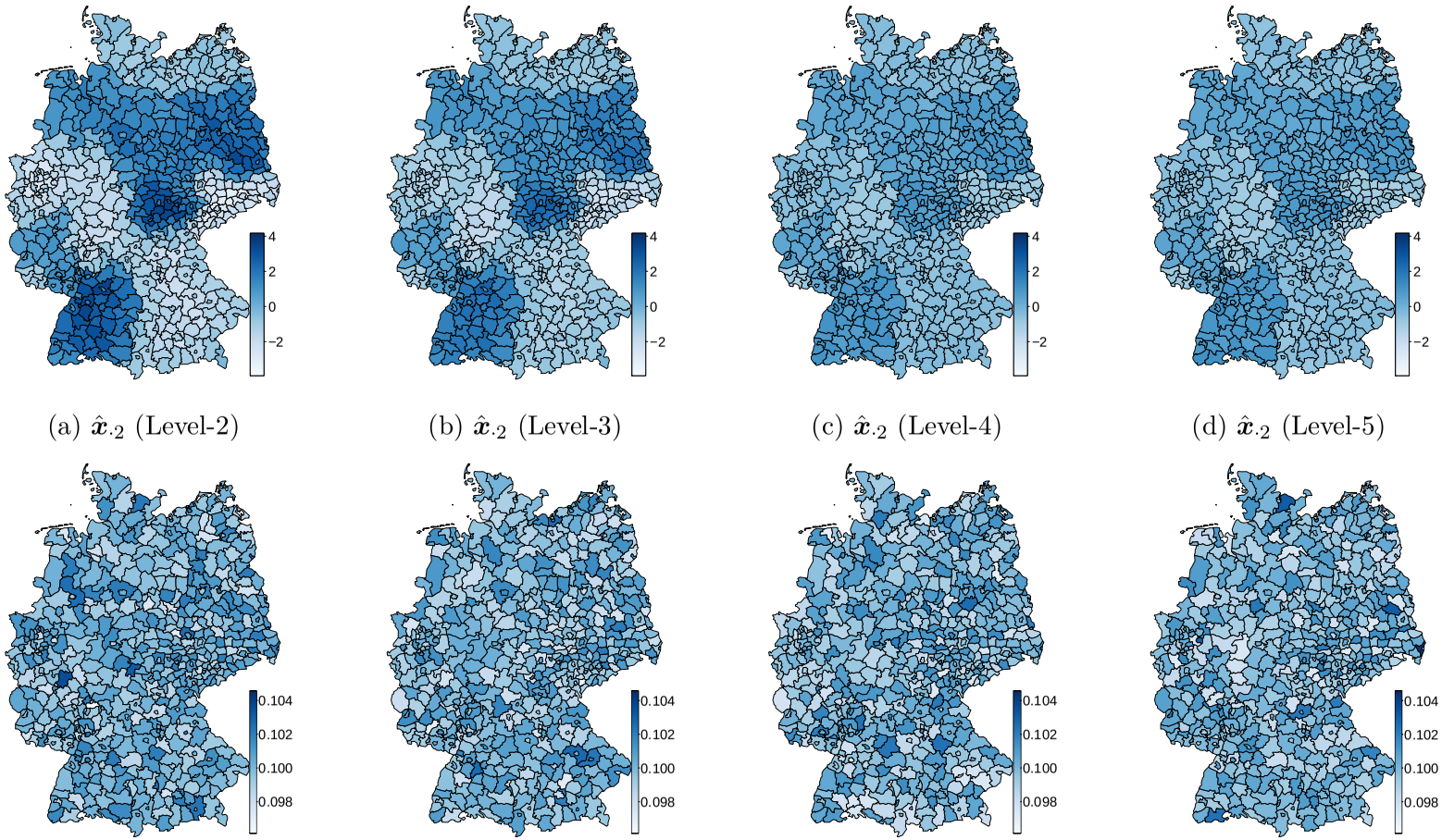
Figure A.6: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 1}, \hat{\boldsymbol{x}}_{\cdot 2}$ and $\hat{\boldsymbol{x}}_{\cdot 3}$ (base level)



(a) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-5)

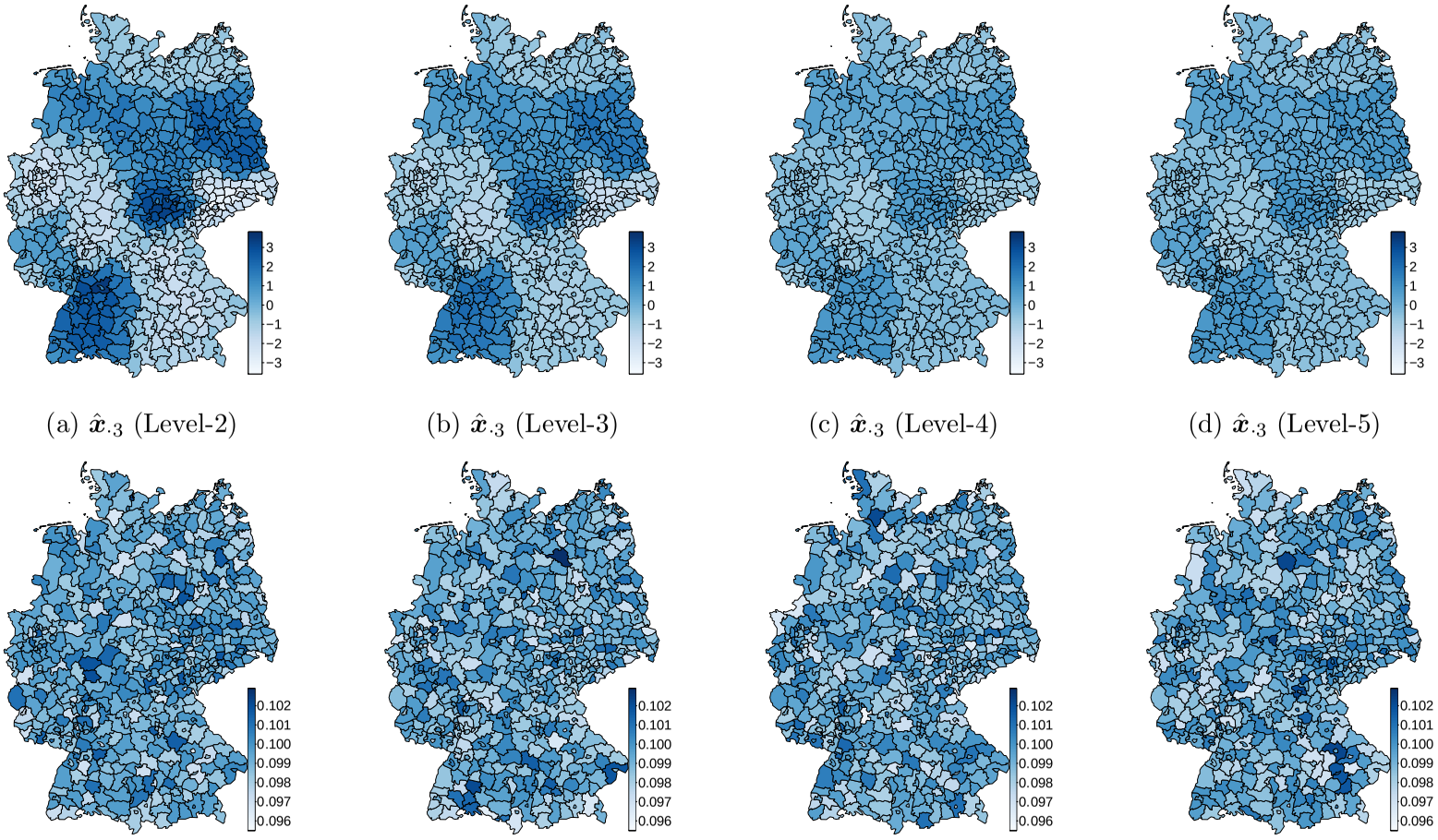
Figure A.7: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 1}$ for various levels of smoothing under Scenario-I



(a) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-5)

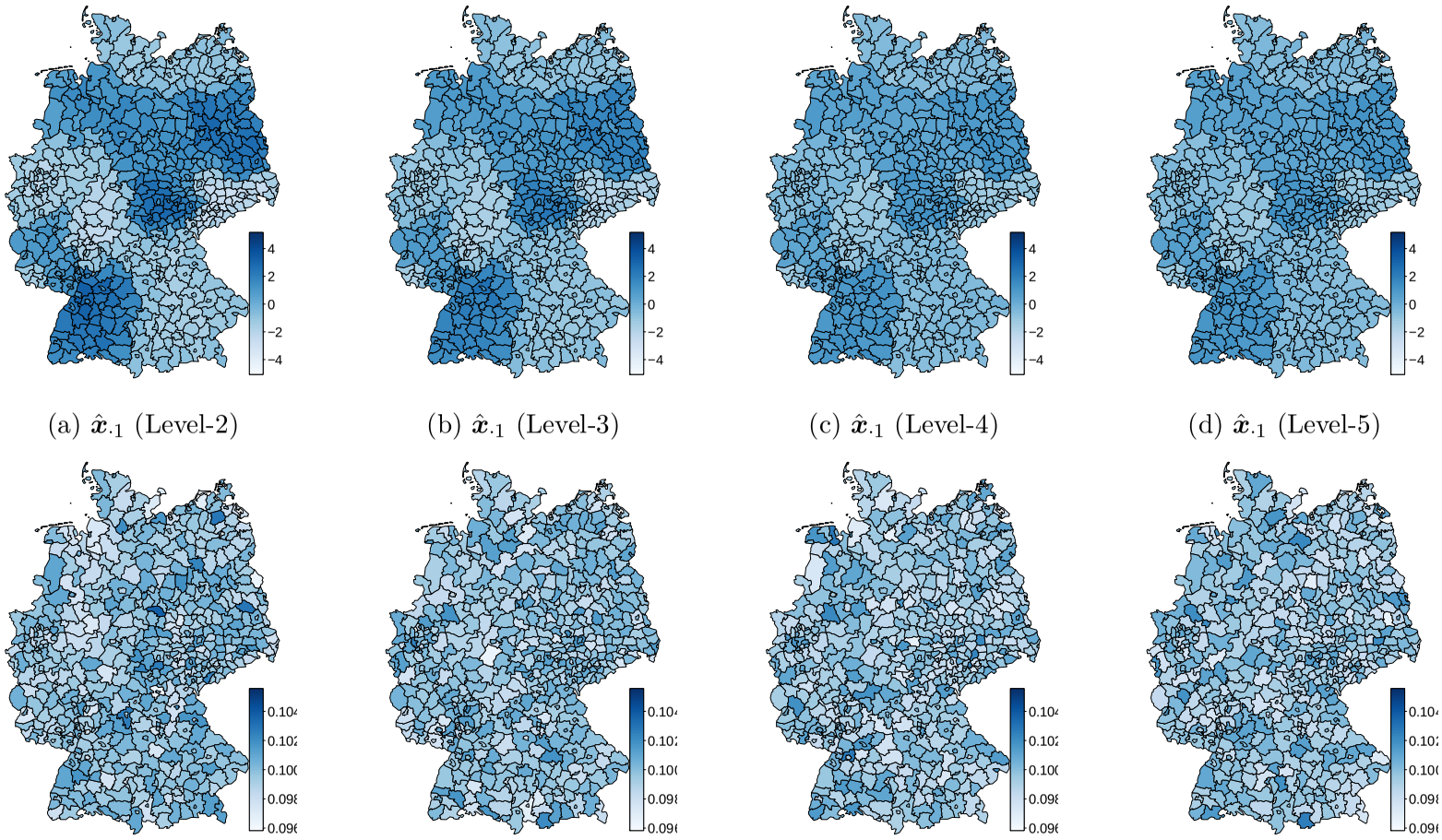
Figure A.8: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 2}$ for various levels of smoothing under Scenario-I



(a) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-5)

(e) Var($\hat{\boldsymbol{x}}_{\cdot 3}$) (Level-2) (f) Var($\hat{\boldsymbol{x}}_{\cdot 3}$) (Level-3) (g) Var($\hat{\boldsymbol{x}}_{\cdot 3}$) (Level-4) (h) Var($\hat{\boldsymbol{x}}_{\cdot 3}$) (Level-5)

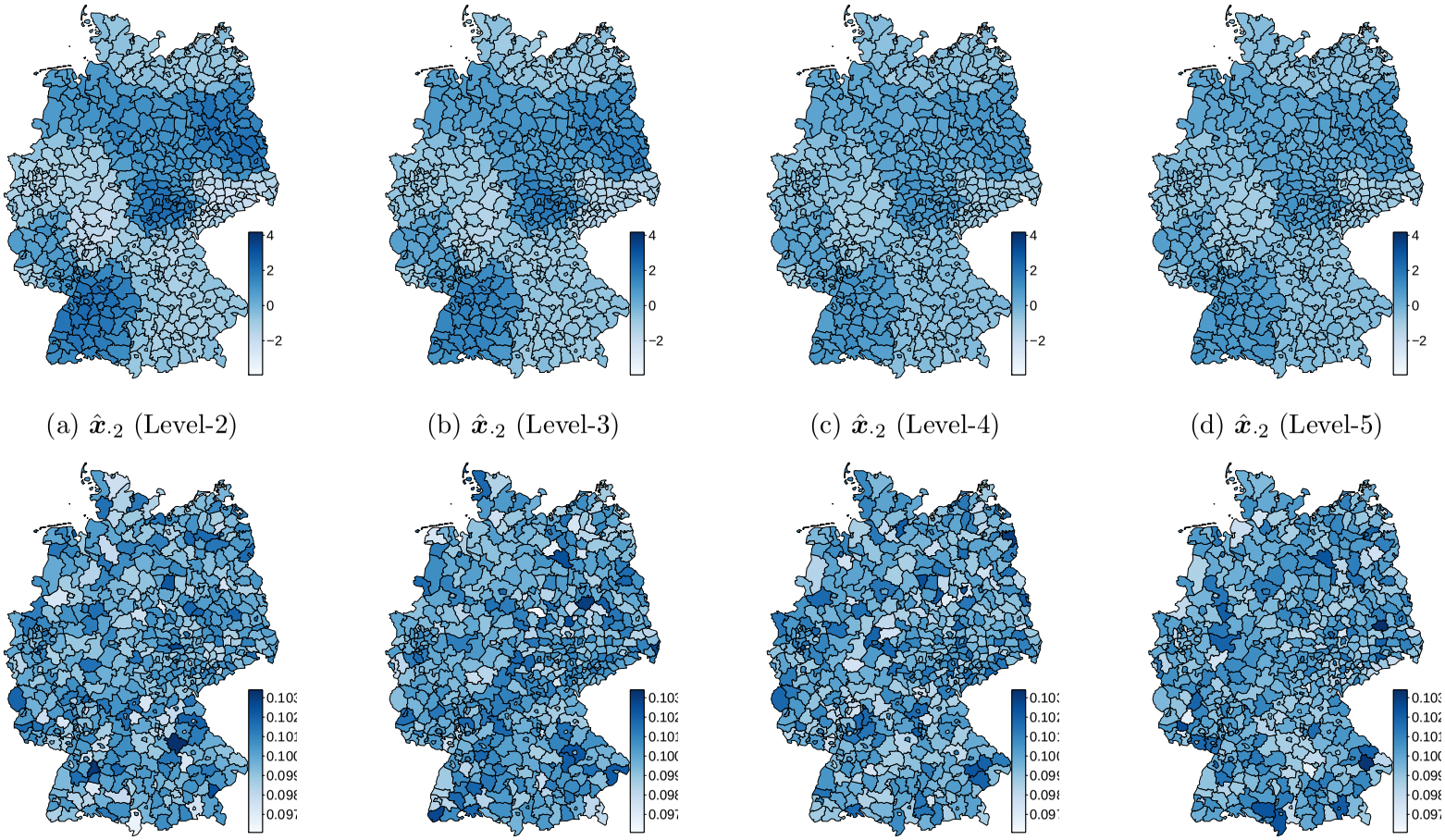
Figure A.9: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 3}$ for various levels of smoothing under Scenario-I



(a) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-5)

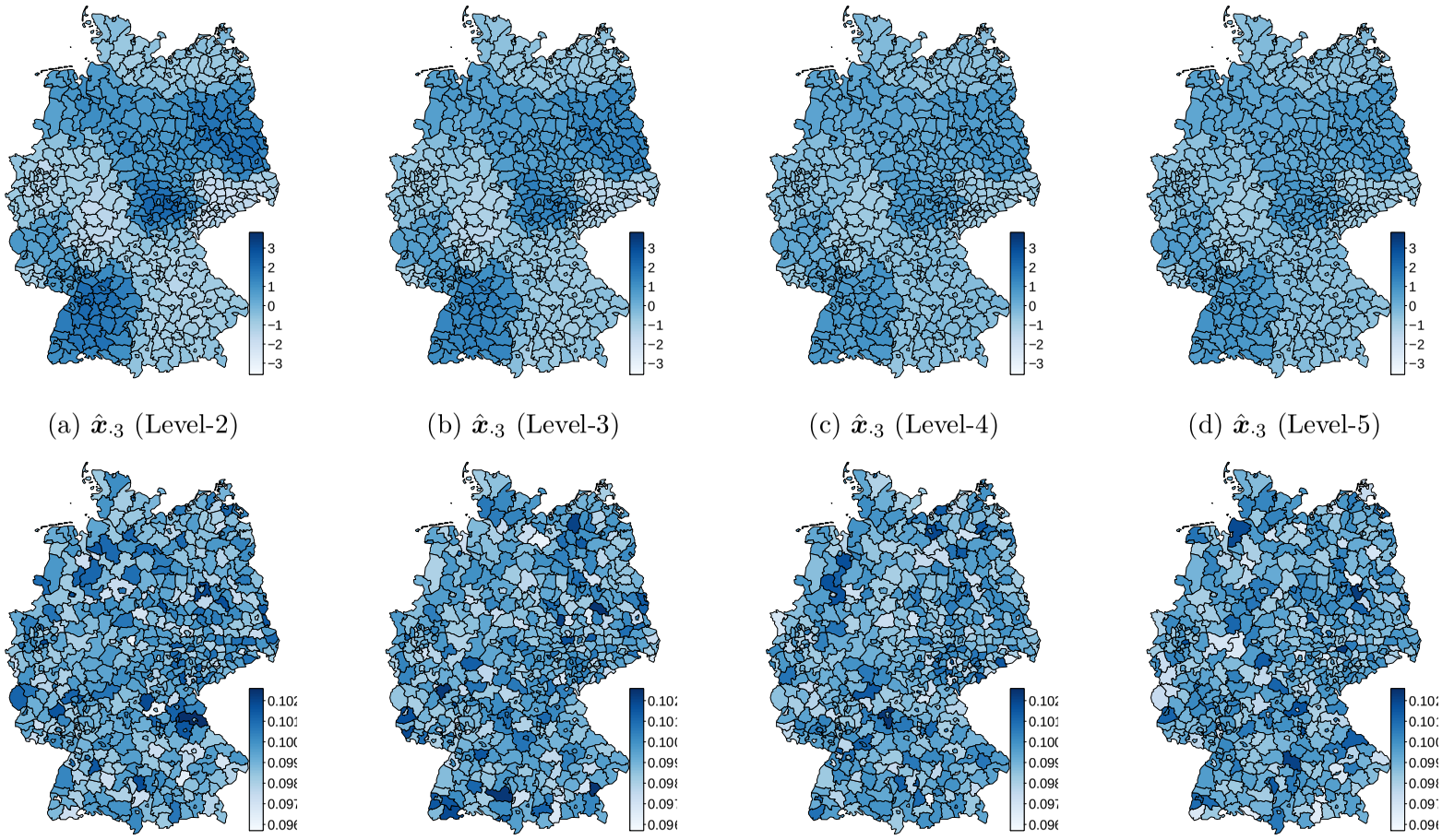
Figure A.10: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 1}$ for various levels of smoothing under Scenario-II



(a) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-5)

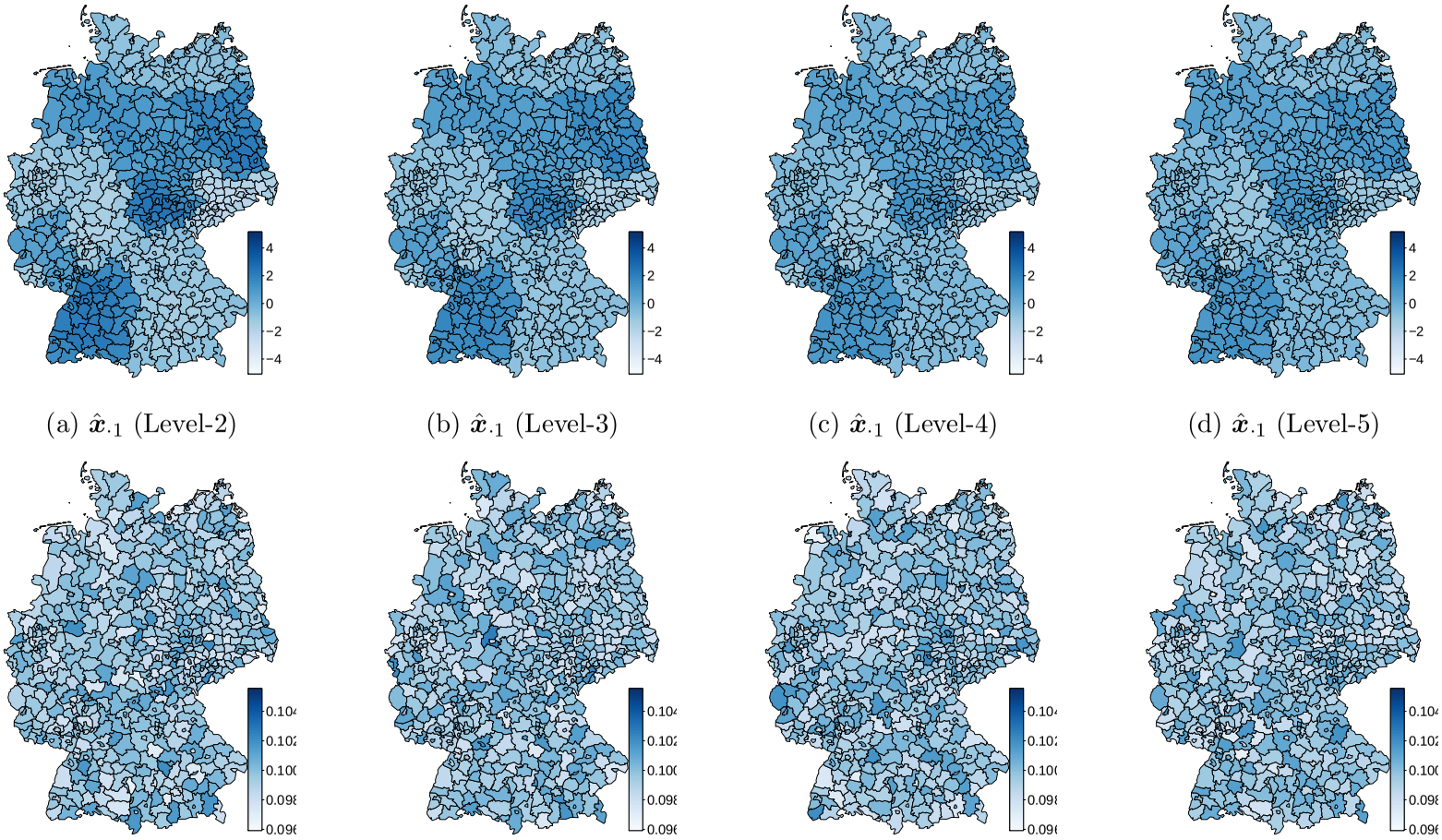
Figure A.11: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 2}$ for various levels of smoothing under Scenario-II



(a) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-5)

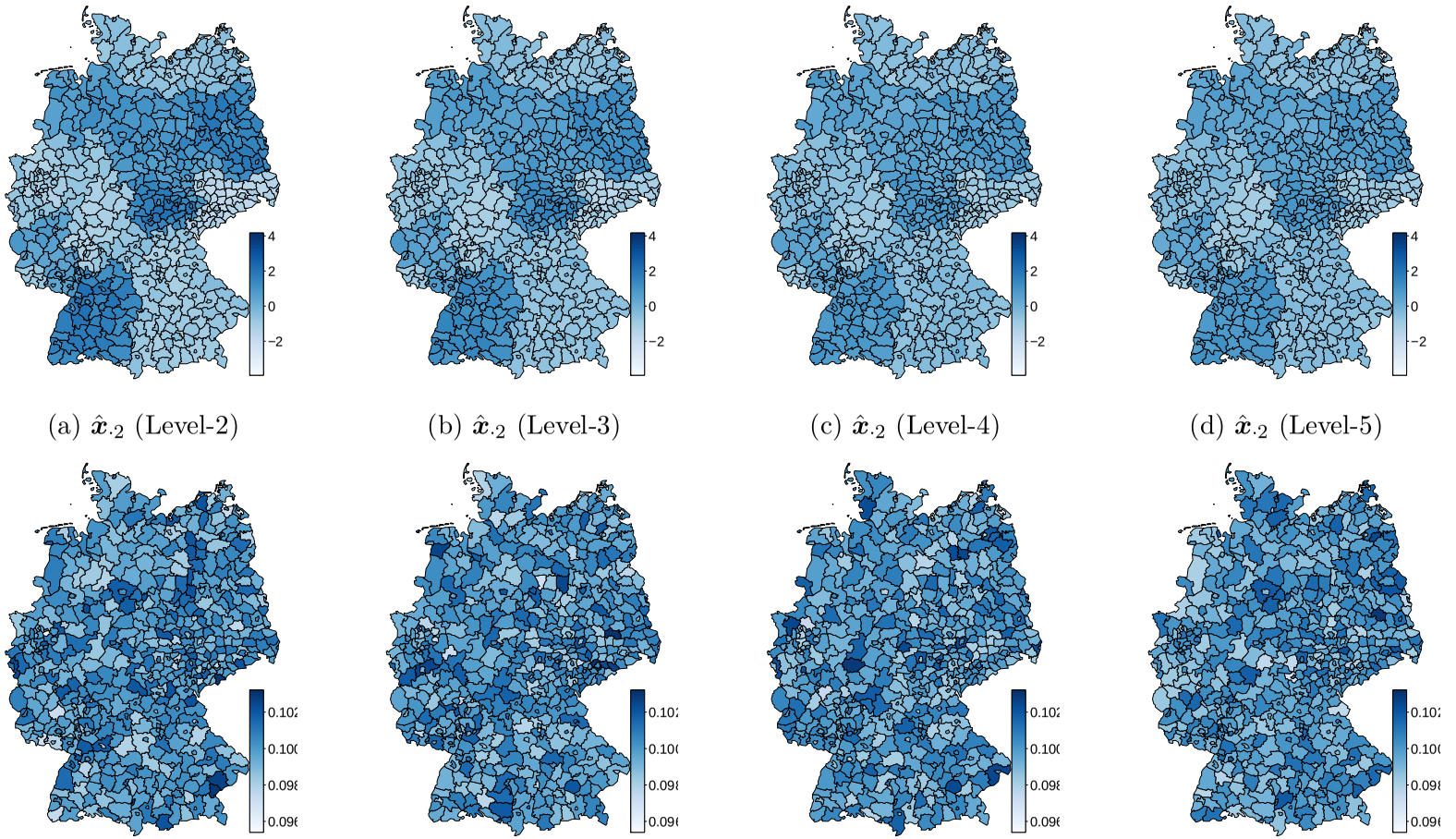
Figure A.12: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 3}$ for various levels of smoothing under Scenario-II



(a) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 1}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 1})$ (Level-5)

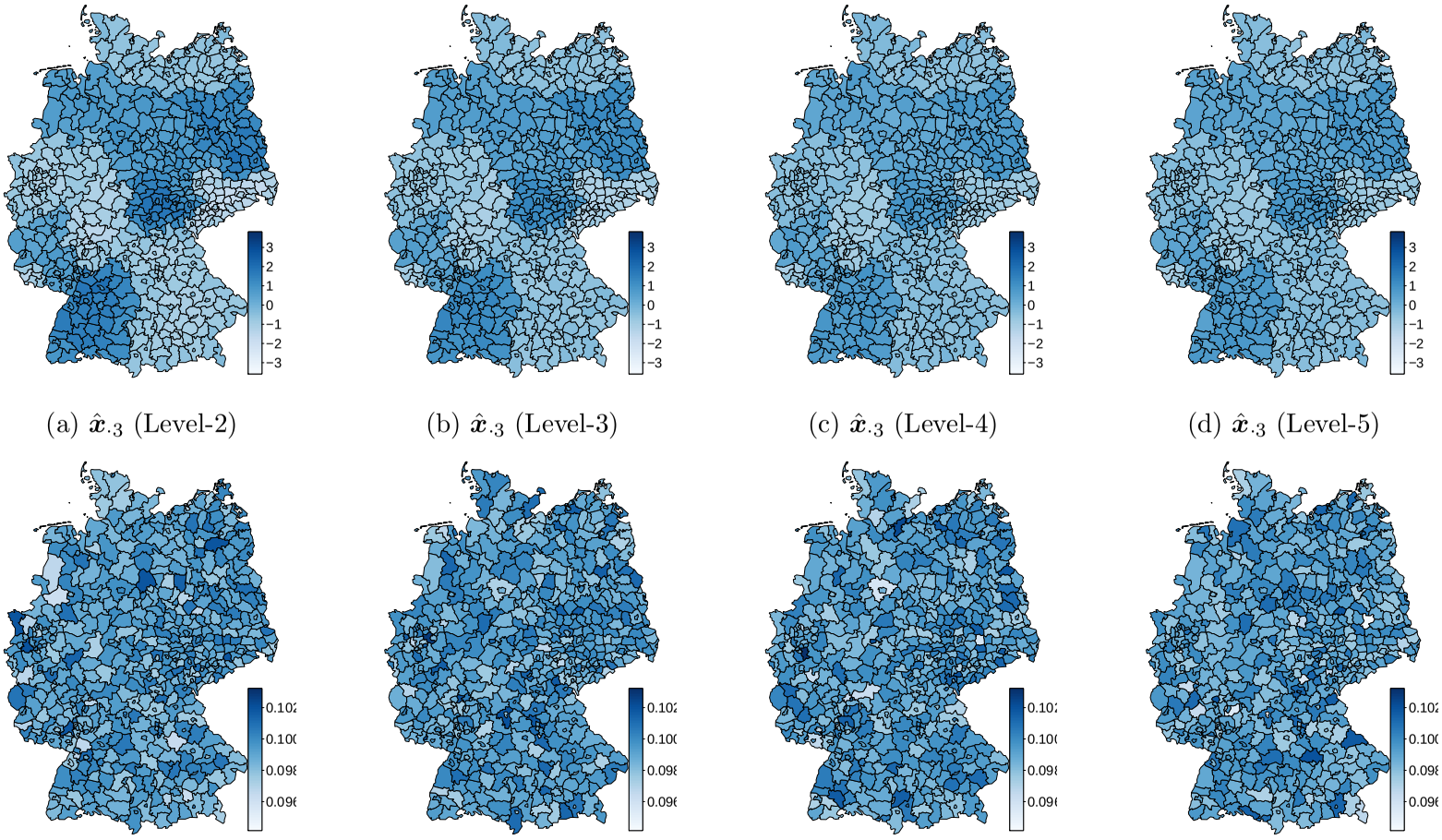
Figure A.13: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 1}$ for various levels of smoothing under Scenario-III



(a) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 2}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 2})$ (Level-5)

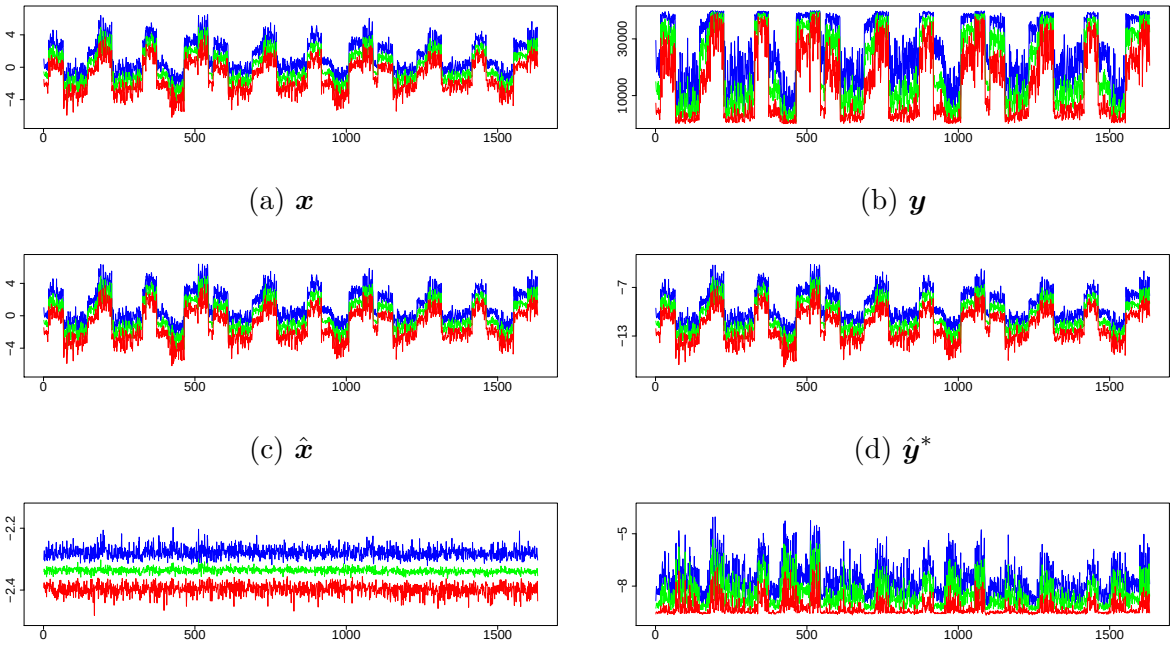
Figure A.14: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 2}$ for various levels of smoothing under Scenario-III



(a) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-2) (b) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-3) (c) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-4) (d) $\hat{\boldsymbol{x}}_{\cdot 3}$ (Level-5)

(e) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-2) (f) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-3) (g) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-4) (h) $\text{Var}(\hat{\boldsymbol{x}}_{\cdot 3})$ (Level-5)

Figure A.15: Posterior means and variances of $\hat{\boldsymbol{x}}_{\cdot 3}$ for various levels of smoothing under Scenario-III



(a) $\boldsymbol{x}$

(b) $\boldsymbol{y}$

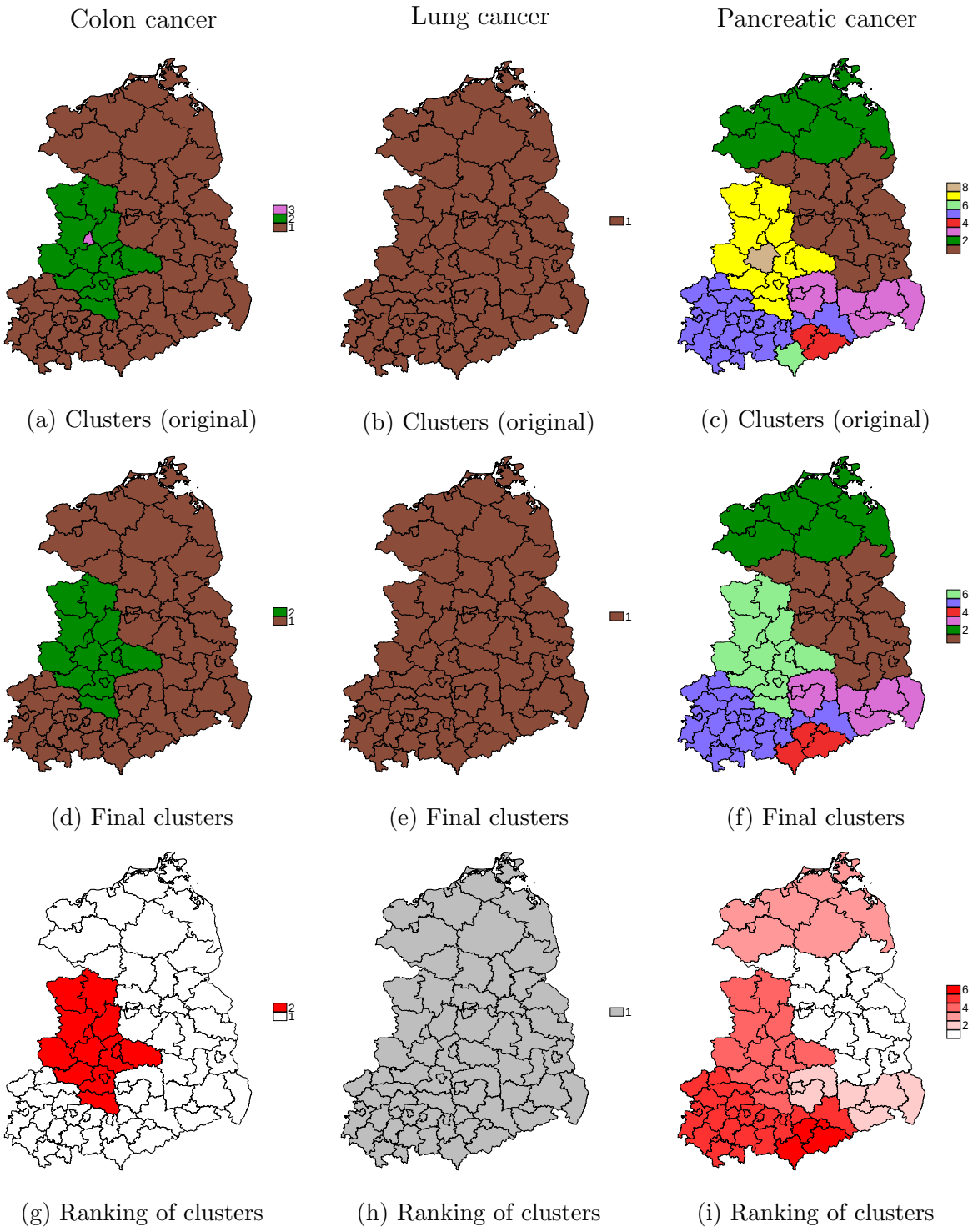
(c) $\hat{\boldsymbol{x}}$

(d) $\hat{\boldsymbol{y}}^*$

(e) Log of posterior variances of $\hat{\boldsymbol{x}}$

(f) Log of posterior variances of $\hat{\boldsymbol{y}}^*$

Figure A.16: Variation in $\boldsymbol{x}$, $\boldsymbol{y}$, $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{y}}^*$ with their posterior variances on log scale



(a) Clusters (original) (b) Clusters (original) (c) Clusters (original)

(d) Final clusters (e) Final clusters (f) Final clusters

(g) Ranking of clusters (h) Ranking of clusters (i) Ranking of clusters

Figure A.17: Risk clusters without smoothing for Colon, Lung and Pancreatic cancers