Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Petter Jeppesen Giørtz

Fast Spatial Multi-level Models for Small Area Estimation

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglstad
June 2023

NTNU
Norwegian University of
Science and Technology

Petter Jeppesen Giørtz

Fast Spatial Multi-level Models for Small Area Estimation

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglstad
June 2023

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Abstract

An extension to the Besag-York-Mollié (BYM) is proposed, which can account for spatial variation across multiple scales. The multi-level model is presented as an alternative method for accurate small area estimation of demographic data, when direct methods from traditional survey statistics are unfeasible. In addition, the model is further developed and tested with the inclusion of covariates and informative prior distributions for the contribution to total variance from each level. Alongside the model itself, a range of computational techniques for fast inference are presented. These allow for extensive testing and validation of highly complex spatial models, and are implemented using the template model builder (TMB) package in `R`.

Validation of the multi-level model was done through a simulation study and a case study in India. This showed that the model is more robust than alternative single-level models when applied to a range of scenarios with variation on multiple spatial scales. It is especially useful when estimating on fine-scale levels on data where most of the variation happens on coarse scales. The use of informative priors for model parameters did not have a large impact on accuracy, but was useful during parameter estimation to counteract overestimation of the most dominant weight parameters. Estimates of the variables of interest on the finest scale were most accurate in terms of a selection of error metrics when using the multi-level model with covariates included. We were able to run this model on a range of data sets with an average run time of approximately 10 minutes, making it a viable model choice for accurate estimation large sets of variables on multiple levels.

Sammendrag

Det foreslås en utvidelse av Besag-York-Mollié (BYM)-modellen som kan ta hensyn til romlig variasjon på flere nivåer. Multi-level modellen presenteres som en alternativ metode for nøyaktig estimering av demografiske data på små områder, der direkte metoder fra tradisjonelle survey-statistics er utilstrekkelige. I tillegg blir modellen videreutviklet og testet med inkludering av kovariater og informativ priorfordeling for bidraget til totalvariansen fra hvert nivå. Sammen med selve modellen blir det presentert en rekke beregningsmetoder for rask inferens. Disse muliggjør omfattende testing og validering av svært komplekse romlige modeller og blir implementert ved hjelp av template model builder (TMB)-pakken i `R`.

Valideringen av flernivåmodellen ble gjennomført gjennom en simuleringsstudie og en casestudie i India. Dette viste at modellen er mer robust enn alternative single-level modeller når den brukes i ulike scenarier med variasjon på flere romlige skalaer. Den er spesielt nyttig når man estimerer på fin-skala nivåer for data der mesteparten av variasjonen skjer på grove skalaer. Bruken av informativ priorfordeling for modellparametere hadde ikke stor innvirkning på nøyaktigheten, men var nyttig under parameterestimeringen for å motvirke overestimering av de mest dominerende vektparametrene. Estimatene for variablene av interesse på fineste skala var mest nøyaktige i forhold til et utvalg feilmål når flernivåmodellen med inkluderte kovariater ble brukt. Vi klarte å kjøre denne modellen på en rekke datasett med en gjennomsnittlig kjøretid på omtrent 10 minutter, noe som gjør den til et levedyktig modellvalg for nøyaktig estimering av store mengder variabler på flere nivåer.

Preface

The work presented in this thesis was conducted during the spring of 2023 at the Department of Mathematical Sciences at the Norwegian University of Science and Technology (NTNU). This marks the completion of my Master's degree in Applied Physics and Mathematics. In this regard, I offer my thanks to NTNU and my supervisor Geir-Arne Fuglstad for exceptional guidance throughout the last year of my degree. I am very happy with the assignment I was given, and I hope the work on this field is continued by future Master's students. Finally, I would like to thank my family for their unwavering support during my studies, and my classmates for creating a motivating and social atmosphere in our study hall.

Petter Jeppesen Giørtz
Trondheim, 11th June 2023

Contents

Chapter 1

Introduction

Most countries in the world are split into different sets of non-overlapping regions such as states and counties. Every such set is called an administrative level, with the coarsest one being the admin 1 level, and on the finer scales come the admin 2 level, admin 3 level and so on. The fine-scale levels are typically nested within the coarser levels. Many low- and middle-income countries have deficient vital registration systems, meaning that there is no complete registration of births and deaths, which makes estimation of population dependent variables within the regions difficult. Therefore, data is collected through surveys to get useful estimates. A common problem with survey data from these countries is that the data is too sparse to estimate metrics of interest on the fine-scale administrative levels, when using traditional survey statistics methods such as direct estimates. Such regions are called 'small areas'. In small areas, making accurate estimates of for example the prevalence of demographic and health indicators has a high value, as this can aid local policymakers with making more informed decisions. Local policymakers exist on all administrative levels, making it desirable to produce estimates on each level. Ideally, estimates on different levels are both accurate and consistent with each other. Thus, in the field of small area estimation (SAE), model-based methods must be developed to achieve accurate estimates on every level, when traditional survey statistic methods are unfeasible.

A common approach is to apply spatial regression models that borrow strength in space by assuming a correlation between variables in neighbouring regions. These kinds of models are discussed in a range of recent research papers such as Utazi et al. (2021), Fuglstad et al. (2021) and L. D. Mercer et al. (2015). They can be used for estimates on a continuous scale, or discrete scale on different levels. This thesis considers discrete models that can be applied to different

administrative levels in a country. The Besag-York-Mollié (BYM) model is an example of a widely used discrete spatial regression model (Besag et al., 1991). It assumes a combination of a random effect and a spatial smoothing effect between regions on a single administrative level. The model can then be applied to make estimates on each level separately.

The problem with these models is that they only consider spatial variation on one administrative level at a time, when in reality there is variation across multiple spatial scales. Thus a natural extension of the BYM model is to enable it to combine effects across multiple administrative levels in order to obtain estimates at the finer levels, while also being able to produce improved estimates on the coarse levels. This motivates the goal of this thesis, which is to present a generalized spatial multi-level model for accurate small area estimation, that can make better estimates than alternative methods and provide useful insights through the model parameters. The model is validated through application to simulated and real data on key demographic indicators that the UN are monitoring in developing countries. If the new model performs well, it can for example be crucial for identifying small-scale regions in low-income countries where new resources and/or policies are necessary, in order to reach the sustainable developments goals (SDGs) as defined by the UN (United Nations General Assembly, 2015).

The new multi-level model is based on a weighted sum of single-level BYM models. Variables are treated as Gaussian Markov random fields (GMRFs) on each administrative level, where the distributions are modelled through a combination of the Besag model (Besag, 1974) and a random i.i.d. effect. These are simple model components that make the multi-level model easier to work with and interpret. The model can also include covariates as separate components, and informative prior distributions of the model parameters. Both of these additional aspects of the model are experimented with in the thesis.

In addition to the development of a new model, a large emphasis is put on different techniques that can be used to attain fast computations without losing accuracy. When working with multi-level spatial effects on fine scales, the computations become highly complex because of the number of latent variables involved. Therefore, we look into how the Laplace approximation can be used to calculate the maximum likelihood during model parameter optimization. This also includes the use of automatic differentiation (AD) for fast computation of exact derivatives, which are needed for the Laplace approximation. Finally, our model choice leads to GMRFs with sparse precision matrices. This means that we can drastically reduce the complexity of the matrix

operations that are needed, such as Cholesky factorization, to further speed up computations. The techniques are applied to construct and compare the spatial models through implementation in the TMB library in `R` (Kristensen, 2023).

To demonstrate the predictive accuracy of the multi-level model and compare it to alternative models, demographic data from India is used as an example in a simulation study. Here, real data provided by the Demographic and Health Surveys (DHS) Program is used to create realistic simulations. This is followed by a case study where the data is used to test the model. The data comes from a survey in India between 2019 and 2021. India is a country with considerable geographical inequality, almost 18% of the world's population, and three different administrative levels. It consists of 41 states, 676 districts, and 2347 subdistricts, of which approximately 10% were not visited in the DHS survey. This makes India a very interesting and appropriate example to test the new model on. The level of education and current employment status among adult females are used as the main example metrics in this thesis. The two metrics are treated as binary variables on the individual level, and spatial regression models are used to estimate the *prevalence* of the metrics on the three administrative levels.

Prevalence refers to the *proportion or percentage* of individuals in a population who have a particular condition at a specific point in time or over a specified period. In areas with large populations prevalence can be used to approximate risk, which refers to the *probability* of an individual developing the condition over a specified period of time. By knowing the approximate risk of developing a specific condition (i.e. education or employment) one can determine whether measures should be taken to increase or decrease this risk, which is why it is useful to produce accurate estimates of prevalence and risk.

Prevalence mapping is a common use case for discrete spatial regression models. Therefore, the performance of the multi-level model is compared to that of common single-level models. Three alternative models are suggested. The first one considers a spatial effect on the admin 1 level, the second considers a spatial effect on the admin 2 level and the last one on the admin 3 level. Mean squared error (MSE) and continuous ranked probability score (CRPS) are used to rank the models. These error measurements are applied to the logit transform of the prevalence, which is used during modelling. This transformation leads to better statistical attributes, meaning that predictions become more accurate and error measurements are easier to compare. The main interest lies in whether the multi-level model outperforms the alternative models when applied to data from India with some degree of spatial variation on all three administrative levels. We also want to see if the inclusion of covariates and informative priors further

increases predictive accuracy.

This thesis has contents that are partly or completely based on sections from the project thesis by Giørtz (2022). This is due to the thesis being a continuation of the work that was performed during the project from 2022, with much of the necessary background knowledge being the same. Thus large parts of Chapter 3 and Chapter 4 are relatively similar to corresponding sections from the project, especially Section 3.7, Section 4.2 and Section 4.7 remain almost unchanged.

The different data sources and preprocessing needed to apply the spatial models is first presented in Chapter 2. In Chapter 3 of this thesis we go through the background knowledge that goes into the models and computational techniques that are used. The definitions of the new model and alternative models are described in Chapter 4, along with formulations of the metrics used for model evaluation. In this chapter we also present the details of the model implementation through the use of the TMB library. Chapter 5 presents a simulation study to evaluate the predictive accuracy of the models. There we investigate how the models' prevalence mappings differ from those of the alternative models through visualizations and numerical results. The models are applied to real data and compared through a case study on India in Chapter 6. The results and findings are then discussed in Chapter 7 along with some concluding remarks.

Chapter 2

India: Geography and data sources

2.1 Geography

India is chosen as the test case for the spatial models. It is a country that is divided into three distinct administrative levels, where the borders on each level are nested within the borders of the previous one. This is necessary to be able to test models that consider variation on multiple administrative levels in a simple and understandable way. India is also a developing country with a very large population, meaning that accurate estimates of important demographic data can have a big impact on future development in the country. The data on India's partition into subnational levels is provided by *GADM – Global Administrative Areas* (2023). The three levels consist of 41 states, 676 districts and 2347 subdistricts, and they are all shown Figure 2.1.

Spatial regression models capture correlation structures between neighbouring regions. This means that any region without neighbours cause issues, because they are independent of all other regions. All island territories without any neighbours are removed in order to avoid this issue when working with territories in India. This includes 2 states on the admin 1 level, 4 districts on the admin 2 level and 39 subdistricts on the admin 3 level.

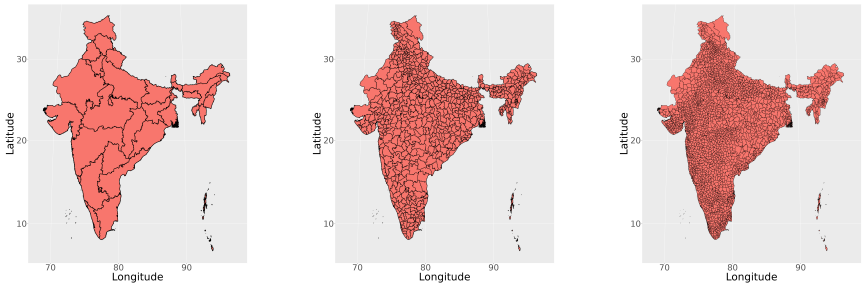


Figure 2.1: India's division into sets of regions on administrative levels 1, 2 and 3.

2.2 DHS survey data

Variables of interest

The data that motivates the development of the new models and inspires the simulation study comes from a DHS survey conducted in India between 2019 and 2021. In the survey, 724115 women between the ages 15 and 49, and 101839 men between the ages 15 and 54 were interviewed on a wide range of questions concerning, for example, health, economic situation and education. The responses of interest in this thesis are whether or not an individual has completed secondary education and if they currently are employed. These responses are chosen because they are good indicators of how far a country has come in the development phase, and there is a sufficient amount of available data that can be used for meaningful estimates. The main interest lies in estimating the *prevalence* of positive responses among females between the ages 20 and 39, within the administrative regions on all levels.

There are two reasons as to why females between the ages 20 and 39 are considered in this thesis. Firstly, they must be of an age where it is realistic to have completed secondary education and be employed. Secondly, this ensures that there is enough data to apply models on the admin 3 level. The DHS data can then be used to both create realistic simulations to test models, and to validate models through application to real data. Additionally, if we are able to make accurate estimates using the data from India, it is likely that the models can be used to make estimates of the variables of interest using similar datasets. This is a major advantage of using DHS data, as the DHS program has already conducted surveys in a wide range of countries of similar fashion as the survey in India.

Survey design

When working with analysis of survey data it is common to use design-based methods which take into account key aspects of the survey design. The main aspect that separates these methods from model-based methods is the assumption that the data is from a fixed finite population. This means that the responses from a specific household are deterministic, and any uncertainty is rooted in the survey sampling itself. Thus it is crucial to take into consideration which households were part of the survey and how they were selected.

According to the DHS report from India (Population Sciences - IIPS/India and ICF, 2022), there are three key ways that complex survey data such as DHS data differs from simple random samples. First, the population is divided into subgroups called strata. The division into strata is based on the population of different villages and the percentage of the population belonging to scheduled castes and scheduled tribes in the country. Then it is decided how many samples are drawn from each stratum. Although the population in the strata can be uneven, the number of samples from each of them is usually very similar. This is in order to ensure reliable estimates even in areas with small populations. Second, a number of primary sampling units (PSUs) are chosen within each stratum, and within each PSU there is a fixed number of households that are surveyed where the household selection is i.i.d within the strata. In the DHS survey in India the number was fixed at 22 households per PSU, and a total of 30456 PSUs were included in the survey. Lastly, the samples are made without replacement. The reasoning behind the survey design is that the precision per survey cost increases, compared to when using completely random sampling.

The samples in the DHS survey from India are scaled based on selection probability to account for the survey design. First the probability of choosing the PSU i in stratum h is calculated, followed by the probability of choosing a specific household in said PSU. Let a_h denote the number of surveyed PSUs in stratum h and let M_{hi} denote the number of households from PSU i in stratum h as reported by the sampling frame. The survey analysts used the latest census made in India which was in 2011 to decide M_{hi}. Sampling was then done based on probability proportional to number of households. The probability of choosing PSU i in stratum h is then

$$P_{hi}^{(1)} = a_h \frac{M_{hi}}{\sum_{j=1}^{N_h} M_{hj}}, \quad i = 1, 2 \dots N_h,$$

where N_h is the chosen number of PSUs in stratum h.

Further, in each PSU the number of surveyed households is denoted g_{hi} and the total number of households in the PSU at time of surveying is denoted L_{hi}. Then the probability of choosing a specific household is

$$P_{hi}^{(2)} = \frac{g_{hi}}{L_{hi}}.$$

The weight for an observation j among the surveyed households g_{hi} is the product of the inverses of these two probabilities

$$w_{hij} = (P_{hi}^{(1)} P_{hi}^{(2)})^{-1}.$$

With these weights the stratified sampling estimator is

$$Y_h = \frac{\sum_i \sum_j w_{hi} y_{hij}}{\sum_i \sum_j w_{hij}}.$$

This is a weighted mean of the observations y_{hij}, which denotes the jth observation from PSU i in stratum h. Note that we write Y_h as a stochastic variable, because the choice of households that are used to compute it is random, whereas y_{hij} is viewed as deterministic in this setting.

Direct estimates

The `survey` package (Lumley, 2004) allows for the computation of *direct estimates* based of a set of survey data using the aforementioned scaling. Producing direct estimates means that no spatial correlation is assumed, so that estimates within each administrative area only make use of the responses in that specific area. Such estimates are easily made when using binary data as inputs. Therefore, we produce estimates of prevalences of the chosen variables along with their respective standard deviations. This provides useful insights into whether or not more complex estimators are necessary to get useful results. Estimates of prevalences and standard deviations are made on logit scale using a quasibinomial model, described in L. Mercer et al. (2014).

After selecting only responses from females between the ages 20-39 from the DHS survey, a total of 436196 responses on education and 65777 responses on employment across India are used to produce the direct estimates. In Figure 2.2 the direct estimates of the prevalence of completed secondary education in India are plotted along with the coefficients of variance for each estimate on the admin 1, admin 2 and admin 3 levels. Similar plots for current employment are shown in Figure 2.3.

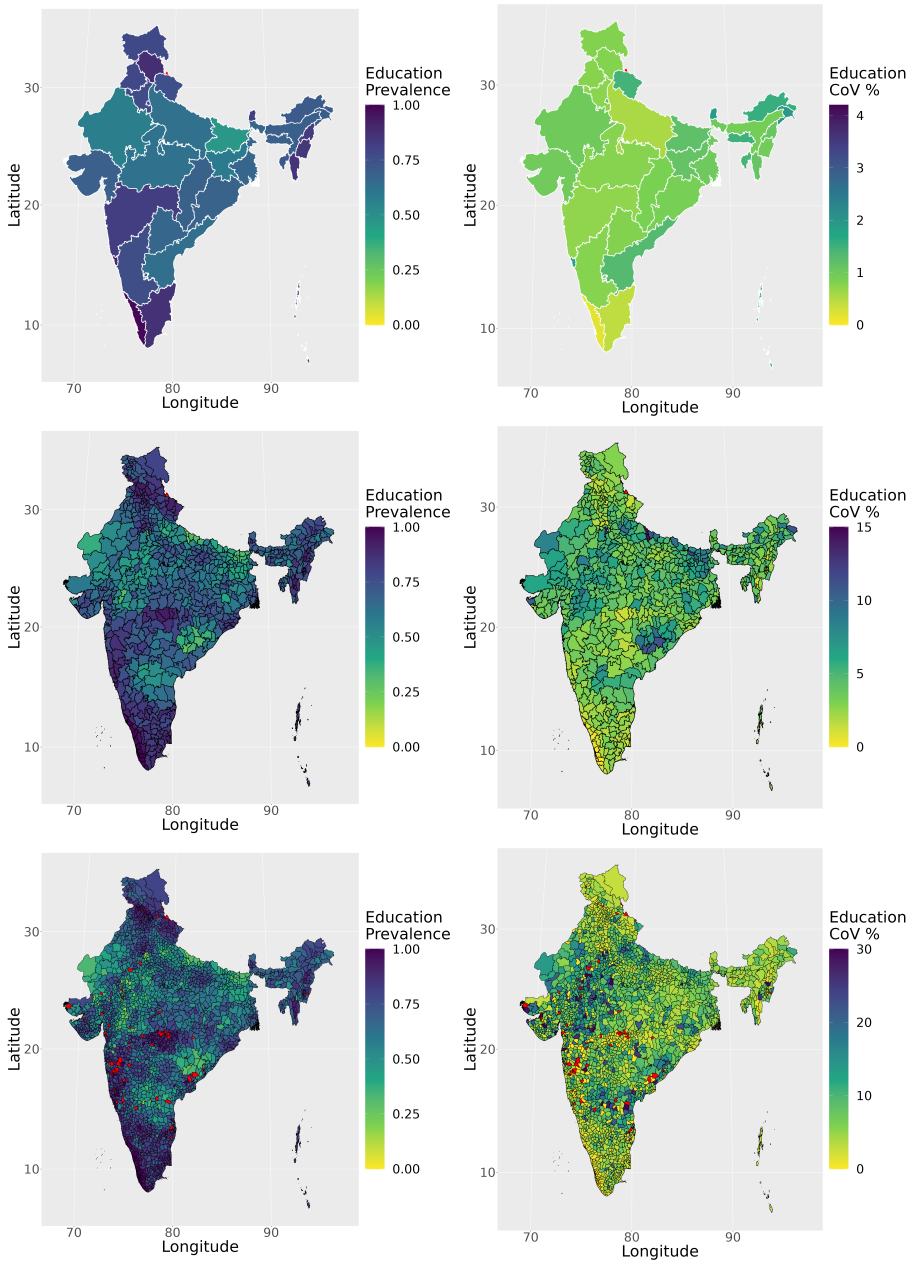


Figure 2.2: Direct estimates of prevalences (left side) and the associated coefficients of variance (right side) of completed secondary education in all admin 1 areas (top row), admin 2 areas (middle row) and admin 3 areas (bottom row). Areas that are colored red had no observations.

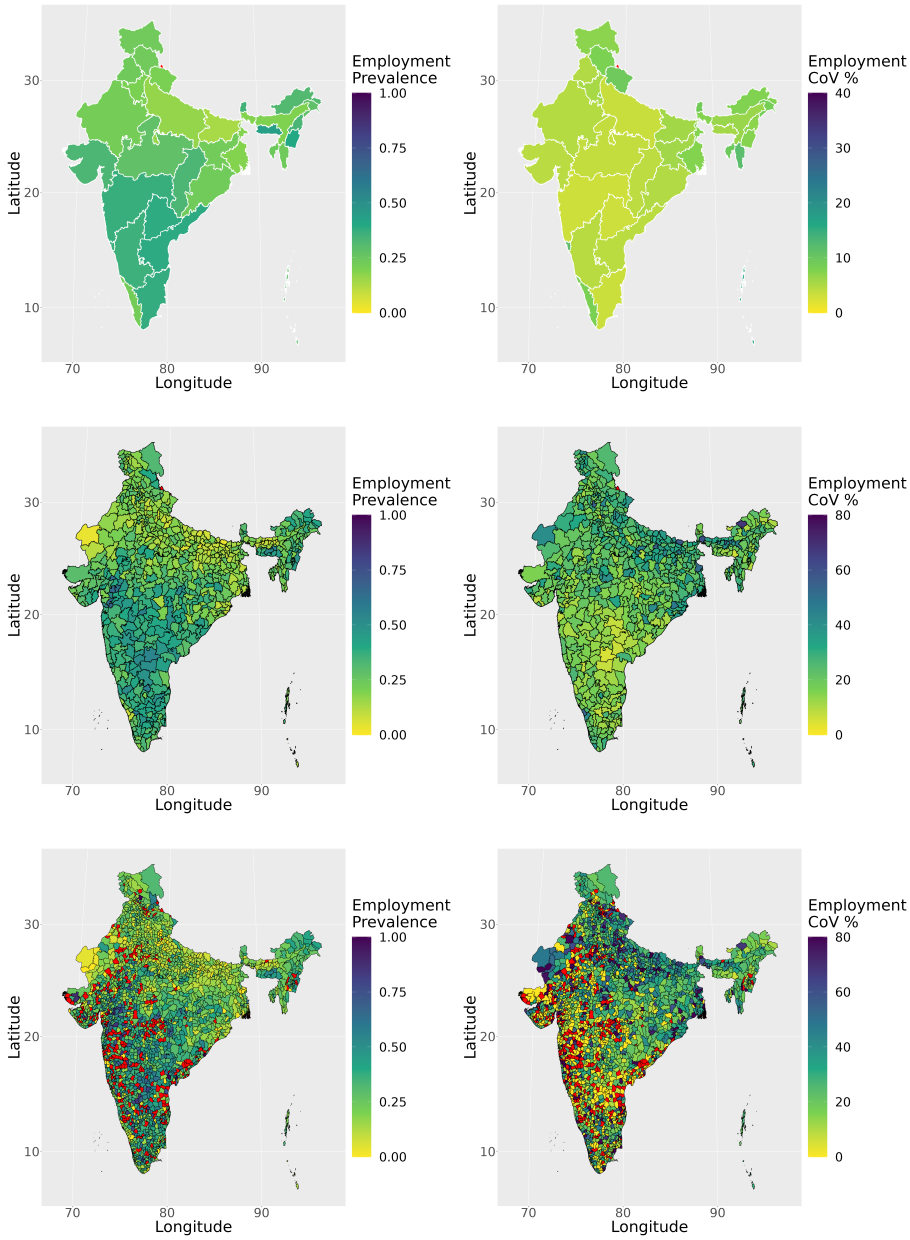


Figure 2.3: Direct estimates of prevalences (left side) and the associated coefficients of variance (right side) of current employment in all admin 1 areas (top row), admin 2 areas (middle row) and admin 3 areas (bottom row). Areas that are colored red had no observations.

An obvious weakness of the direct estimate approach is that it is not possible to produce direct estimates in many areas on the admin 3 level. This is due to the lack of a sufficient amount of data on such a fine scale, where some admin 3 areas have zero responses. The coefficients of variance are also high, especially for the estimates on the admin 3 level when looking at employment. These results motivate the need for model-based approaches to obtain estimates on fine scales, and to improve accuracy on the admin 1 and 2 levels.

2.3 Covariate raster

In addition to the application of spatial regression in the model-based estimation, covariates can be included as an extra model component. Covariate values are assigned to each administrative area based on mapping available data from WorldPop (2018) to the areas. The covariates that are used in this thesis are chosen based of analysis from the DHS India report (Population Sciences - IIPS/India and ICF, 2022).

The report states for example that "Among both females and males, the median number of years of schooling is higher in urban areas than in rural areas (7.5 years versus 4.0 years among females and 8.8 years versus 6.5 years among males)" and "The employment level is much higher among less educated persons, highest among persons with less than 5 years of schooling (89% among men and 34% among women)". With the reported analytics in mind, the following covariates are chosen to be included in the simulation study, with all the necessary data being available online.

Population density

It is clear from the DHS report that there are large differences between populations from urban and rural areas. A simple way to capture some of these differences is to include population density as a covariate. Data from WorldPop (2018) is openly available and used to compute population density. WorldPop provides estimated population numbers on a 100m×100m grid across the whole country, separated by genders and age groups. Thus the estimates can be mapped to areas on the admin 1, admin 2 and admin 3 levels, and added together to obtain population counts in each area. Population density is then calculated by dividing by the area of the regions. The population counts also serve another purpose, as they can be used to compute weights for estimation on admin 1 and admin 2 level through weighted sums of admin 3 estimates. Details on how this is calculated are described in Section 4.5.

Nighttime lights

Data on nighttime lights is another interesting metric provided by WorldPop (2018). Looking at the amount of nighttime lights relative to the population within a region can give an indication of the amount of wealth, level of development and degree of urbanisation in the region. All of these factors are expected to affect both education and employment, making nighttime lights a relevant covariate to consider during estimation.

Chapter 3

Background

The direct estimate approach presented in Chapter 2 is clearly not usable for small area estimation when working with data similar to the DHS survey data from India. An alternative method that can provide estimates even in areas without any available data, is to instead use a model-based approach. In this chapter we explain the theory behind a commonly used such approach, along with how it can be turned into a computationally efficient method for accurate estimation.

3.1 Areal spatial modelling

To obtain model-based estimates in a set of areas, a useful method is to apply areal spatial modelling. The term areal implies that estimates are made for an entire administrative area, and not for individuals within the area. The DHS survey data is provided with geographical points for each data point on a continuous scale. In order to apply areal spatial modelling these data points are all mapped to their respective admin 1, admin 2 and admin 3 areas, which consequently converts the data to a discrete format.

A common approach for modelling discrete spatial data is to create models based on the relationships between neighbouring regions. One way this can be done is through the usage of a covariance function based on the distances between centroids of the administrative areas, inspired by the centroid method introduced in Fisher (1936). However, such models typically require far too much heavy computation when working on fine scales. Another, more easily applicable method is the use of Spatial Autoregressive (SAR) models. In these models, the correlation structure between neighbouring regions is specified independently of their size and shape, leading to considerably less complexity compared to for

example the centroid method.

When choosing an areal spatial model for fine-scale estimation, it is crucial to consider the computational challenges the model entails. Fitting a model to data across a set of more than one thousand regions leads to complex operations that can be time consuming. Thus the model choice should be based on reasonable assumptions that make the model usable within an acceptable time frame while producing sufficiently accurate estimates. One widely used assumption that entails both computational advantages and increased interpretability is to assume that the variable of interest follows a multivariate Gaussian distribution.

A variable $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)^T$ that follows a multivariate Gaussian distribution can be defined through a mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, so that $\mathrm{E}[\eta_i] = \mu_i$ and $\mathrm{Cov}(\eta_i, \eta_j) = \Sigma_{ij}$. The probability density function is then

$$\pi(\boldsymbol{\eta}) = (2\pi)^{-n/2} \|\boldsymbol{\Sigma}\|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})\right), \quad \boldsymbol{\eta} \in \mathbb{R}^n.$$

Working with such Gaussian variables has multiple advantages when fitting complex spatial models. The Gaussian distribution is one of the most commonly used probability distributions, making it easy to recognize and interpret through only looking at the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$. It is also a well-behaved distribution that has many desirable statistical properties. For example, the maximum likelihood estimates of the mean and variance of a Gaussian distribution are known to be efficient, meaning that they are consistent and have the smallest variance among all such estimators.

However, prevalence is limited to the interval between 0 and 1, and clearly does not follow a Gaussian distribution. To avoid this issue models are applied to the logit transform of the prevalence. Thus for a prevalence p in an administrative area i it is assumed that

$$\eta_i = \mathrm{logit}(p_i) = \log\left(\frac{p_i}{1 - p_i}\right)$$

follows a Gaussian distribution.

3.2 Gaussian Markov Random Fields

The assumption of a multivariate Gaussian distribution still involves certain computational challenges. Specifically, when $\boldsymbol{\Sigma}$ is a dense covariance matrix there are several mathematical operations that become highly complex during

model fitting, such as computing the inverse and determinant of $\mathbf{\Sigma}$. Thus it is computationally beneficial to introduce conditions that reduce the complexity. To this end, the neighbourhood structure between a set of non-overlapping regions in a country can be exploited by assuming that the variable of interest follows the Markov property.

The Markov property states that the distribution of a variable within a region is *conditionally independent* of all other regions, given the values in all neighbouring regions. It is a useful property when the neighbourhood structure between administrative areas can be defined as an undirected graph on the form $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes corresponding to the administrative areas and $\mathcal{E}$ is the set of edges corresponding to all pairs of neighbouring areas. To visualize this, displays of the graph structures on admin 1 level and admin 2 level in India are shown in Figure 3.1.

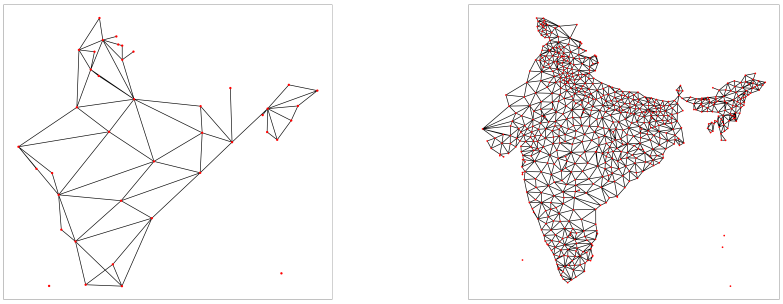


Figure 3.1: Graph structure of India on admin 1 level (left) and on admin 2 level (right). The nodes represent administrative areas and the edges represent links between neighbouring areas.

When a pair of nodes i, j are neighbours we denote it as $i \sim j$ and $\{i, j\} \in \mathcal{E}$. For a node $i \in \mathcal{V}$ we define the set of all neighbours of i as $N(i) = \{j \in \mathcal{V} : i \sim j\}$. The global Markov property then says that

$$\pi(\eta_i|\boldsymbol{\eta}_{-i}) = \pi(\eta_i|\boldsymbol{\eta}_{N(i)}), \quad \boldsymbol{\eta} \in \mathbb{R}^n.$$

Another way to phrase this is that for two non-neighbouring region i, j we have that

$$\eta_i|\boldsymbol{\eta}_{-ij} \perp\!\!\!\perp \eta_j|\boldsymbol{\eta}_{-ij}, \quad \{i, j\} \notin \mathcal{E}. \tag{3.1}$$

The graph structures of administrative levels as the ones in Figure 3.1 are clearly very sparse, meaning that most nodes are conditionally independent of each

other. The Markov property can be visualized on the admin 1 graph in India as in Figure 3.2. Here, the distribution of η_i on the green node in the figure is independent of all other nodes if the values on the three blue nodes are known. Similar behaviour happens on all other nodes as well, making the conditional distributions significantly less complex, thanks to the Markov property.

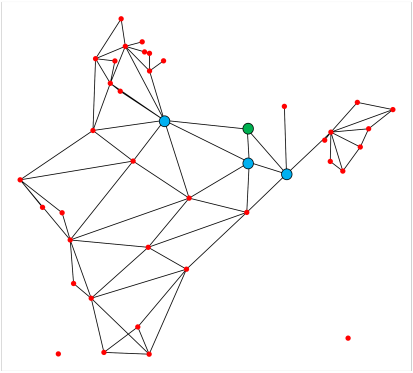


Figure 3.2: Illustration of the Markov property in action. When the nodes are assumed to possess the Markov property, the distribution of a variable on the green node is independent of all the red nodes, if the values are known on the three blue nodes.

A widely used approach when modelling discrete spatial data that makes use of both the assumption of normally distributed variables and the Markov property is called Gaussian Markov random fields (GMRF). It is defined in Rue and Held (2005) as

Definition 3.2.1 (Gaussian Markov Random Field)**.** For $n \in \{1, 2 \ldots\}$ a random vector $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)^T \in \mathbb{R}^n$ is called a GMRF with respect to a labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q} > 0$ iff its density has the form

$$\pi(\boldsymbol{\eta}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu})^T \mathbf{Q} (\boldsymbol{\eta} - \boldsymbol{\mu})\right), \quad \boldsymbol{\eta} \in \mathbb{R}^n, \tag{3.2}$$

and

$$\mathrm{Q}_{i,j} \neq 0 \quad \Longleftrightarrow \quad \{i, j\} \in \mathcal{E} \; \forall \; i \neq j.$$

The covariance matrix $\boldsymbol{\Sigma}$ is defined so that the elements are $\Sigma_{i,j} = \mathrm{Cov}(\eta_i, \eta_j)$. This means that the complexity of the matrix is unaffected by conditional independence. Therefore, the precision matrix $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ is used instead in the

definition. Choosing to use the sparse precision matrix during modelling proves to entail several computational advantages, which are explained further in Section 3.7.

It is possible to assume that GMRFs have precision matrices that are not of full rank. Such cases are called *improper GMRFs*, and are useful when trying to capture long-range dependencies between nodes in a graph where the mean vector is not given. In order to apply an improper GMRF in a meaningful way, one needs to introduce constraints to account for the rank deficiency of the precision matrix. If the $n \times n$ precision matrix $\mathbf{Q}$ is of rank $n - k$, there must be k independent constraints to get an identifiable model. Rue and Held (2005) define an improper GMRF as

Definition 3.2.2 (Improper GMRF)**.** Let $\mathbf{Q}$ be an $n \times n$ symmetric positive semi-definite (SPSD) matrix with rank $n - k > 0$. Then $\boldsymbol{\eta} = (\eta_1, \eta_2, \ldots, \eta_n)^T$ is an improper GMRF of rank $n - k$ with parameters $(\boldsymbol{\mu}, \mathbf{Q})$, if its density is

$$\pi(\boldsymbol{\eta}) = (2\pi)^{-\frac{n-k}{2}} (|\mathbf{Q}|^*)^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{\eta} - \boldsymbol{\mu})^T \mathbf{Q}(\boldsymbol{\eta} - \boldsymbol{\mu})\right), \quad \boldsymbol{\eta} \in \mathbb{R}^n.$$

where $|\mathbf{Q}|^*$ is the product of all non-zero eigenvalues of $\mathbf{Q}$. Further, $\boldsymbol{\eta}$ is an improper GMRF with respect to the labelled graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where

$$\mathrm{Q}_{i,j} \neq 0 \quad \Longleftrightarrow \quad \{i, j\} \in \mathcal{E} \; \forall \; i \neq j.$$

Note that although an improper GMRF does not have a mean and precision matrix $(\boldsymbol{\mu}, \mathbf{Q})$ formally, we still denote the two parameters as the mean and precision matrix for convenience.

A special case of improper GMRFs is an *intrinsic GMRF* of first order, as defined by Rue and Held (2005)

Definition 3.2.3 (Intrinsic GMRF)**.** An intrinsic GMRF of first order is an improper GMRF of rank $n - 1$ where $\mathbf{Q}\boldsymbol{1} = \boldsymbol{0}$.

From the definition it follows that the IGMRF is invariant to the addition of a constant $c \cdot \boldsymbol{1}$. This means that it is able to capture the deviation from any global mean level across the graph, without the mean level having to be given. As the mean level is often unknown and/or not our main interest, this is a desirable trait of IGMRFs that is commonly exploited in spatial modelling, for example when modelling prevalence through $\boldsymbol{\eta}$.

3.3 The Besag model

Among the most widely used IGMRFs are the intrinsic conditional auto regressive models (ICAR), first discussed in Besag (1974). The ICAR models have a general

density function defined as

$$\pi(\boldsymbol{x}) \propto \kappa^{(n-1)/2} \exp\left(-\frac{\kappa}{2}\sum_{i\sim j} w_{ij}(x_i - x_j)^2\right), \quad \boldsymbol{x} \in \mathbb{R}^n, \tag{3.3}$$

where the precision parameter $\kappa > 0$ and symmetric weights w_{ij} are chosen to capture the impact that the squared differences between direct neighbours have on the likelihood. The simplest example of such a model is the Besag model, where all weights $w_{ij} = 1 \quad \forall \quad i \sim j$, making it a generalization of the classical random-walk in a two-dimensional space. Hence the conditional distribution of the GMRF is

$$x_i|\boldsymbol{x}_{-i} \sim \mathcal{N}\left(\frac{1}{nb(i)}\sum_{j\in N(i)} x_j, \frac{1}{\kappa \cdot nb(i)}\right),$$

with $nb(i)$ being the number of neighbours node i has. The conditional mean is simply the mean of all neighbouring values, and the precision parameter controls the conditional variance. An important advantage of the simple form of the Besag model is that the structure matrix corresponding to the model is easy to define and implement, and has the form

$$\mathrm{R}_{i,j} = \begin{cases} nb(i), & i = j, \\ -1, & i \sim j, \\ 0, & \text{otherwise.} \end{cases}$$

The precision matrix is then $\mathbf{Q} = \kappa\mathbf{R}$, which can be shown to be a matrix of rank $n - 1$. The rank deficiency of the IGMRF is resolved by introducing a sum-to-zero constraint.

However, it can be difficult to interpret the precision parameter κ, as it only controls conditional variance, and not marginal variance. As a result, the total variation across graphs of different complexities, such as in Figure 3.1, can vary a lot despite using the same precision parameter. As the two graphs in fact cover the same country, it is unrealistic that there is much more total variation from the mean on admin 2 level than on admin 1 level. Therefore, it is desirable to scale the precision matrices $\mathbf{Q}$ so that the precision parameter primarily controls the marginal variance across the graphs rather that the conditional.

Sørbye and Rue (2014) showed a simple solution to appropriately scale the structure matrix $\mathbf{R}$. The method is to scale it by the geometric mean of the marginal variances of a GMRF that has $\mathbf{Q} = 1 \cdot \mathbf{R}$ as its precision matrix. Thus we get

$$\mathbf{Q}^* = \kappa(c\mathbf{R})$$

$$c = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log(\mathbf{R}_{ii}^{-})\right),$$

with $\mathbf{R}^{-}$ being the generalized inverse of the structure matrix which satisfies $\mathbf{R}\mathbf{R}^{-}\mathbf{R} = \mathbf{R}$. This scaling makes it easier to interpret and compare precision parameters of ICAR models that are applied to graphs of varying complexity. Thus all use of the Besag model throughout this thesis will include scaled versions of the precision matrices $\mathbf{Q}$.

3.4 Hierarchical models of binomial data

The Besag model and similar spatial models are not suited for modelling of binomially distributed data directly, such as the DHS data on education and employment. Instead, these models are applied to the logit transformation of the prevalence parameter from the binomial distribution, as explained in Section 3.1. Thus the spatial models are in fact a part of what is called hierarchical spatial models. These consist of three main components which are an observation model, a latent model and parameters that are to be estimated.

The observation model is based on the observable variables, which are assumed to be binary responses valued 0 or 1 such as the variables introduced in Chapter 2. In an area i there are m_i observations, and the jth observation is denoted $Y_{ij}, \quad j = 1, 2, \ldots, m_i$. The set of observations from that area then follows a binomial distribution

$$Y_i | p_i \sim \text{Bin}(m_i, p_i),$$

where $Y_i = \sum_{j=1}^{m_i} Y_{ij}$.

Further, the latent model is used for the parameter p_i. To achieve this the set of parameters $\boldsymbol{p}$ from all $i = 1, 2, \ldots, N$ areas are considered together. The latent model then assumes that the logit transform of these parameters follows a multivariate Gaussian distribution

$$\text{logit}(\boldsymbol{p}) = \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1}).$$

For the parameters that are to be estimated in this distribution we introduce an expansion upon the classical Besag model, which was proposed in Besag et al. (1991), called the Besag-York-Mollié (BYM) model. Here, an alternative to the Besag model is presented, where a normal i.i.d. component is paired with the ICAR component in each region. This model is used because it is reasonable to assume that there is some degree of variation in the prevalences that is not

explained by spatial effects, but is instead considered to be completely random. The model uses a weighted sum of these two components

$$\boldsymbol{\eta} = w_B\boldsymbol{u} + w_R\boldsymbol{v}, \quad \boldsymbol{u} \sim \mathcal{N}(0, \mathbf{Q_B}^{-1}), \quad \boldsymbol{v} \sim \mathcal{N}(0, \mathbf{Q_R}^{-1}), \tag{3.4}$$

where $\boldsymbol{u}$ is the ICAR component and $\boldsymbol{v}$ is the i.i.d. component, and they can have either separate or a shared precision parameter κ. Riebler et al. (2016) suggest a reparametrization of this model that reduces the complexity of parameter estimation. This is done by scaling the Besag component according to Section 3.3 and using a single precision parameter to control the total marginal variance, resulting in the model

$$\boldsymbol{\eta} = \frac{1}{\kappa}(\sqrt{\phi}\boldsymbol{u} + \sqrt{1-\phi}\boldsymbol{v}), \quad \boldsymbol{u} \sim \mathcal{N}(0, \mathbf{Q_B}^{-1}), \quad \boldsymbol{v} \sim \mathcal{N}(0, \mathbf{Q_R}^{-1}). \tag{3.5}$$

An intercept μ can also be included when the expected mean value is not equal to 0. Then the objective is to determine the weight parameter ϕ, the precision parameter κ and the intercept μ, in order make accurate estimations of prevalences based on a set of binomially distributed spatial data.

3.5 Inclusion of covariates in spatial models

To further expand upon the BYM model in Equation 3.5, covariates can be included to increase the accuracy of estimations. Naturally, the values of variables across a country is not only determined by spatial correlation effects, but can also just as well be affected by local circumstances. Examples of covariates that can have a significant effect on the variables of interest are presented in Section 2.3. Building upon the model in (3.5), an intercept and covariates are included by setting

$$\boldsymbol{\eta} = \mathbf{G}\boldsymbol{\beta} + \frac{1}{\kappa}(\sqrt{\phi}\boldsymbol{u} + \sqrt{1-\phi}\boldsymbol{v}), \quad \boldsymbol{u} \sim \mathcal{N}(0, \mathbf{Q_B}^{-1}), \quad \boldsymbol{v} \sim \mathcal{N}(0, \mathbf{Q_R}^{-1}), \tag{3.6}$$

where the $3 \times N$ matrix $\mathbf{G} = [\mathbf{1}, \boldsymbol{G_1}, \boldsymbol{G_2}]^T$ contains the observed covariate values on each node, and $\boldsymbol{\beta}$ is a parameter vector with β_0 being the intercept. In this case $\mathbf{1}$, $\boldsymbol{G_1}$ and $\boldsymbol{G_2}$ represent the intercept, values for population density and values for nighttime lights, respectively. The log transform is used on the covariate values, as this is deemed more likely to be linearly correlated with the responses of interest. Through this simple way of modelling with covariates, a larger amount of available data can be used for estimations, which is likely to improve predictive accuracy while still using interpretable parameters $\boldsymbol{\beta}$.

3.6 Selecting prior distributions for model parameters

Spatial models such as the BYM model contain specific parameters that are estimated based on training on sets of real or simulated data. Definition of these parameters involves defining their initial values and prior distributions. There are mainly two options when defining the priors. The first option is to use vague priors. A vague prior distribution is one that contains little information or is intentionally uninformative. This is useful for situations where there is limited knowledge about the parameter, or to avoid introducing unwanted bias into the modelling process. A common vague prior for weight parameters is the uniform distribution on the unit interval

$$w \sim \text{Unif}(0, 1),$$

and for unrestricted parameters such as the intercept and covariate parameters a normally distributed prior with a large variance is commonly used

$$\beta \sim \mathcal{N}(0, V^2),$$

where V is chosen to be a large number.

The alternative to vague priors is informative priors. These can be used when there is information available that makes it possible to make reasonable assumption about the parameters. When these assumptions are correct, models are more likely to estimate the correct parameters and make more accurate estimations. However, there is a risk involved as the estimations can be worse if the wrong assumptions are made about the parameters. Therefore, it is interesting to investigate how much accuracy in prediction and parameter estimation is to gain from setting good, informative priors, and the consequences of setting bad priors. Then it can be decided whether or not it is worth it to try to introduce informative priors, or if vague priors lead to sufficiently accurate results and is a safer choice.

Weight parameters are usually set on the unit interval $w \in [0, 1]$. When there is reason to believe that a weight should tend more towards a certain value on the interval, it is common to set the prior to a beta distribution

$$w \sim \text{Beta}(a, b).$$

This way the mean and variance can be controlled by the parameters a and b, where

$$\text{E}(w) = \frac{a}{a + b},$$

$$\mathrm{Var}(w) = \frac{ab}{(a+b)^2(a+b+1)}.$$

In Figure 3.3 a few different plots are shown to illustrate how parameters can be chosen to adjust the prior mean and variance.

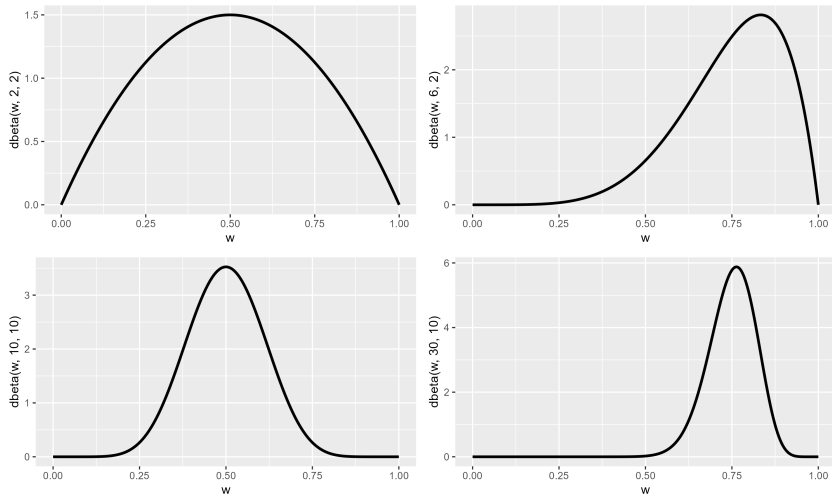


Figure 3.3: Prior beta distributions with different choices of parameters. Here with $a = 2$, $b = 2$ (top left), $a = 6$, $b = 2$ (top right) $a = 10$, $b = 10$ (bottom left), and $a = 30$, $b = 10$ (bottom right).

In many models there are sets of weights for different model components where the weights in a set should sum up to 1. A set of beta priors are combined with this constraint for a multivariate informative prior in the Dirichlet distribution (Dirichlet, 1850), which can be used as a prior for such a set of weights. For K weights the prior is defined as

$$\boldsymbol{w} \sim \mathrm{Dir}(\boldsymbol{\alpha}),$$

where the probability density function is

$$f(\boldsymbol{w}, \boldsymbol{\alpha}) = \frac{1}{\mathrm{B}(\boldsymbol{\alpha})} \prod_{i=1}^{K} w_i^{\alpha_i - 1}, \quad \mathrm{B}(\boldsymbol{\alpha}) = \frac{\prod_{i=1}^{K} \Gamma(\alpha_i)}{\Gamma(\alpha_0)}, \quad \alpha_0 = \sum_{i=1}^{K} \alpha_i,$$

and for $0 < w_1, \ldots, w_K < 1$ we have $\sum_{i=1}^{K} w_i = 1$. Here, $\boldsymbol{\alpha}$ is a parameter vector that controls the shape of the prior distribution, with larger values indicating

stronger prior beliefs that the associated weight is large. From the mean and covariance of the variables following a Dirichlet distribution, it is clear that it can be interpreted as a generalization of the beta distribution to higher dimensions as

$$\mathrm{E}(w_i) = \frac{\alpha_i}{\alpha_0},$$

$$\mathrm{Var}(w_i) = \frac{\alpha_i(\alpha_0 - \alpha_i)}{\alpha_0^2(\alpha_0 + 1)}.$$

The prior distributions of the weights $\boldsymbol{w}$ can be adjusted similarly to in Figure 3.3, using that

$$w_i \sim \mathrm{Beta}(\alpha_i, \alpha_0 - \alpha_i).$$

Parameters that are not restricted like weights can still use informative priors that follow normal distributions, but adjust the mean and variance to match available information. For example, if there is strong evidence that the intercept in a model should be approximately 1, the prior distribution

$$\beta_0 \sim \mathcal{N}(1, 1)$$

can be a reasonable choice. When parameters are restricted to the positive half-plane, such as the precision parameter $\kappa \in (0, \infty)$, the common choice for the prior is a gamma distribution. This can also be either vague or informative so that

$$\kappa \sim \mathrm{Gamma}(\alpha, \beta), \quad \mathrm{E}(\kappa) = \frac{\alpha}{\beta}, \quad \mathrm{Var}(\kappa) = \frac{\alpha}{\beta^2},$$

where the two parameters control whether the prior is vague or informative.

3.7 Techniques for fast computations

The use of a hierarchical spatial model as presented in Section 3.4 together with a set of prior distributions for the model parameters as presented in Section 3.6 leads to highly complex likelihood functions. For a set of areas the likelihood becomes a combination of multiple binomial distributions, multivariate Gaussian distributions and the prior distributions. This entails that heavy computations are needed to fit the models to large sets of data, such as the DHS data across the fine-scale administrative levels in India. Thus we need to introduce efficient computational techniques so that models can be applicable in practice.

3.7.1 Estimation of the likelihood through the Laplace approximation

Estimation of model parameters, denoted as $\boldsymbol{\theta}$, involves computing the maximum likelihood of said parameters together with the unknown random effects, denoted

as $\boldsymbol{z}$. However, computing exact likelihoods can be very computationally expensive. For example, when working with the fine-scale administrative levels in India, $\boldsymbol{z}$ contains thousands of latent variables. In addition, we are often working with non-Gaussian likelihoods such as a binomial or Poisson likelihood, or combinations of multiple likelihoods. As a result, it in unfeasible to compute exact likelihoods when applying spatial regression models. In this thesis we are using the negative joint log-likelihood, denoted as $f(\boldsymbol{z}, \boldsymbol{\theta})$, as it has more useful attributes for optimization purposes. The objective is to maximize

$$L(\boldsymbol{\theta}) = \int_{\mathbb{R}^n} \exp(-f(\boldsymbol{z}, \boldsymbol{\theta}))dz, \quad \boldsymbol{\theta} > 0.$$

As we cannot efficiently compute this exactly, we instead use an accurate and efficient approximation called the Laplace approximation, as described in Skaug and Fournier (2006). It is computed through three main steps. First, the minimizer of $f(\boldsymbol{z}, \boldsymbol{\theta})$ with respect to $\boldsymbol{z}$ is computed and defined as

$$\hat{\boldsymbol{z}}(\boldsymbol{\theta}) = \arg\min_{\boldsymbol{z}} f(\boldsymbol{z}, \boldsymbol{\theta}), \quad \boldsymbol{\theta} > 0.$$

Second, $\mathbf{H}(\boldsymbol{\theta})$ is computes, which denotes the Hessian of $f(\boldsymbol{z}, \boldsymbol{\theta})$ with respect to $\boldsymbol{z}$ evaluated at $\hat{\boldsymbol{z}}(\boldsymbol{\theta})$, written as

$$\mathbf{H}(\boldsymbol{\theta}) = \frac{\partial^2}{\partial \boldsymbol{z}^2} f(\hat{\boldsymbol{z}}(\boldsymbol{\theta}), \boldsymbol{\theta}), \quad \boldsymbol{\theta} > 0.$$

Finally, the Laplace approximation for the marginal likelihood is computed as

$$L^*(\boldsymbol{\theta}) = (2\pi)^{n/2} |\mathbf{H}(\boldsymbol{\theta})|^{-1/2} \exp(-f(\hat{\boldsymbol{z}}, \boldsymbol{\theta})), \quad \boldsymbol{\theta} > 0.$$

The estimates for the parameters $\boldsymbol{\theta}$ are henceforth obtained by minimizing the negative log of the Laplace approximation. The objective function becomes

$$-\log(L^*(\boldsymbol{\theta})) = -\frac{n}{2}\log(2\pi) + \frac{1}{2}\log(|\mathbf{H}(\boldsymbol{\theta})|) + f(\hat{\boldsymbol{z}}, \boldsymbol{\theta}), \quad \boldsymbol{\theta} > 0,$$

which can be minimized by means of standard nonlinear optimization algorithms such as BFGS. In addition to obtaining an estimate $\hat{\boldsymbol{\theta}}$ after optimization, the variance of the estimator, or any differentiable function of the estimate $\phi(\hat{\boldsymbol{\theta}})$, can be estimated as such

$$\widehat{\mathrm{Var}(\phi(\hat{\boldsymbol{\theta}}))} = -\frac{\partial}{\partial \boldsymbol{\theta}}\phi(\hat{\boldsymbol{\theta}}) \left(\frac{\partial^2}{\partial \boldsymbol{\theta}^2} \log\Big(L^*(\hat{\boldsymbol{\theta}})\Big) \right)^{-1} \frac{\partial}{\partial \boldsymbol{\theta}}\phi(\hat{\boldsymbol{\theta}})^T,$$

which is reduced to the inverse Hessian of the log likelihood with respect to $\boldsymbol{\theta}$ when $\phi(\boldsymbol{\theta})$ is the identity function.

Part of these calculations is finding the first- and second derivatives of the objective function. These computations require special techniques in order to be fast, which is necessary when estimating parameters in complex models. One such technique that can efficiently compute exact derivatives is called automatic differentiation.

3.7.2 Computation of derivatives using automatic differentiation

Due to Fournier et al. (2012), a good method for fast computation of exact derivatives when working with complex spatial models is through the use of automatic differentiation (AD). This method has many advantages over alternative, inexact methods such as finite differences. Firstly, derivatives are needed for the Laplace approximation to the marginal likelihood. As this is already an approximation, using approximates to the first- and second order derivatives would lead to exponentially growing errors, which can be detrimental when optimizing model parameters. Secondly, methods such as finite differences require many function evaluations. We are working with comprehensive models that have complex likelihoods. Thus many function evaluations are computationally expensive, making finite differences inefficient.

Automatic differentiation breaks down any complex function to a sequence of elemental unary or binary operations on floating point representations of real numbers. Derivation of the complex statement is hence the product of multiple simple partial derivatives following the chain rule. As an example, we can look at the expression for computing the squared error of a single point after applying linear regression $S_i = (y_i - (a + bx_i))^2$. This can be split into five steps

$$
\begin{aligned}
t_1 &= bx_i & \frac{\partial t_1}{\partial b} &= x_i, \ \frac{\partial t_1}{\partial x_i} = b \\
t_2 &= a + t_1 & \frac{\partial t_2}{\partial a} &= 1, \ \frac{\partial t_2}{\partial t_1} = 1 \\
t_3 &= y_i - t_2 & \frac{\partial t_3}{\partial y_i} &= 1, \ \frac{\partial t_3}{\partial t_2} = -1 \\
t_4 &= t_3^2 & \frac{\partial t_4}{\partial t_3} &= 2t_3 \\
S_i &= t_4 & \frac{\partial S_i}{\partial t_4} &= 1.
\end{aligned}
$$

Using combinations of these partial derivatives leads to simple computations of

partial derivatives through use of the chain rule

$$\frac{\partial S_i}{\partial a} = \frac{\partial t_2}{\partial a}\frac{\partial t_3}{\partial t_2}\frac{\partial t_4}{\partial t_3}\frac{\partial S_i}{\partial t_4} = -2(y_i - (a + bx_i)),$$

$$\frac{\partial S_i}{\partial b} = \frac{\partial t_1}{\partial b}\frac{\partial t_2}{\partial t_1}\frac{\partial t_3}{\partial t_2}\frac{\partial t_4}{\partial t_3}\frac{\partial S_i}{\partial t_4} = -2x_i(y_i - (a + bx_i)).$$

In AD there are two main strategies that are used to compute these partial derivatives, namely the forward mode and the reverse mode. For models with many estimated parameters the latter is the preferred choice. Reverse mode AD consists of three main steps that lead to the partial derivatives of the objective function.

- Compute and store all intermediate quantities $t_1, t_2, t_3 \ldots$.
- Define the partial derivatives and use them to compute the objective function's sensitivity to each intermediate variable $\frac{\partial S}{\partial t_1}, \frac{\partial S}{\partial t_2}, \frac{\partial S}{\partial t_3} \ldots$
- Extract the gradient of the objective function by finding the shortest path through a sensitivity and the chain rule $\frac{\partial S}{\partial a} = \frac{\partial S}{\partial t_2}\frac{\partial t_2}{\partial a}$, $\frac{\partial S}{\partial b} = \frac{\partial S}{\partial t_1}\frac{\partial t_1}{\partial b}$.

This technique generalizes well to any model that has an objective function that can be decomposed in a similar manner, which is the case for the models presented in Chapter 4.

3.7.3 Exploitation of sparsity for fast matrix operations

The neighbourhood structures between regions in India (or in any other country) lead to highly sparse precision matrices. Using the sparse $n \times n$ precision matrices $\mathbf{Q}$ in computations rather than the dense covariance matrices $\mathbf{\Sigma}$ allows us to considerably reduce the complexity of matrix operations that are needed during optimization. For example, part of computing the Laplace approximation, is evaluating the likelihood of GMRFs, given by (3.2). Thus the determinant of the precision matrices have to be computed. The determinant of a dense matrix has a computational complexity of $\mathrm{O}(n^3)$. However, when working with a sparse symmetric positive definite matrix, such as the precision matrix $\mathbf{Q}$, this can be reduced to $O(n^{3/2})$. This is done by first finding the Cholesky factorization of $\mathbf{Q}$

$$\mathbf{L}\mathbf{L}^T = \mathbf{Q},$$

which has the complexity $\mathrm{O}(n^{3/2})$. The determinant is then computed as

$$|\mathbf{Q}| = |\mathbf{L}|^2 = \prod_{i=1}^{n} L_{ii}^2,$$

with complexity $O(n)$. Another part of the density in (3.2) is computing $(\boldsymbol{x}-\boldsymbol{\mu})^T\mathbf{Q}(\boldsymbol{x}-\boldsymbol{\mu})$. In regular multivariate Gaussian distributions it is common to instead denote this as $(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})$. This involves computing the inverse of a dense covariance matrix, which has a complexity of $O(n^3)$. By instead computing the precision matrix in advance, we completely avoid having to compute this inverse. In addition, the sparsity of $\mathbf{Q}$ allows for considerably reduced complexity of multiplying it with the vector $(\boldsymbol{x}-\boldsymbol{\mu})$.

When sparse matrix operations are combined with automatic differentiation to compute the Laplace approximation to the marginal likelihood, we end up with a very efficient way of fitting certain spatial regression models. This allows us to develop new, highly complex models and apply them to large sets of data. In the following chapter a new model for binomial multi-level spatial regression is presented, along with a description of how validation of the model is done and how it is implemented through the use of techniques for fast computations.

Chapter 4

Binomial spatial regression models

4.1 Binomial observation model

The first part of the models that are developed in this thesis is the observation model. Specific notation is used for observations on different administrative levels. Let $n_j^{(1)}$, $Y_j^{(1)}$ and $p_j^{(1)}$ denote the number of surveyed individuals, the number of positive responses and the associated prevalence in admin 1 area j, respectively. The number of admin 1 areas is denoted as N_1. On admin 2 level, the corresponding variables are denoted $n_k^{(2)}$, $Y_k^{(2)}$ and $p_k^{(2)}$ in admin 2 area k, with N_2 being the number of admin 2 areas. Similarly, we use the notation $n_l^{(3)}$, $Y_l^{(3)}$ and $p_l^{(3)}$ for the admin 3 area l, and N_3 for the number of admin 3 areas. In the beginning of this chapter the focus is on modelling prevalence on the admin 3 level, where it is assumed that

$$Y_l^{(3)}|p_l^{(3)} \sim \text{Binomial}(n_l^{(3)}, p_l^{(3)}), \quad l = 1, 2, \dots N_3.$$

The spatial models are then applied to the prevalence $p_l^{(3)}$ through a logit link function, as explained in Section 3.4,

$$\text{logit}(p_l^{(3)}) = \eta_l^{(3)}.$$

Having defined the observation model, the next step is to define the second part of the hierarchical models, which is the latent spatial models that are applied to $\boldsymbol{\eta}^{(3)}$.

4.2 Latent models

4.2.1 Single-level models

Single level models only consider variation on one level. The first model considers variation on the admin 1 level, and is denoted 'Admin 1'. We let $j[l]$ denote the index of the admin 1 area that admin 3 area l lies in. The model for estimation on admin 3 level is then defined as

$$\eta_l^{(3)} = \mu + \frac{1}{\sqrt{\kappa}}\left(\sqrt{\phi_1}u_{j[l]}^{(1)} + \sqrt{1-\phi_1}v_{j[l]}^{(1)}\right), \quad l = 1, 2, \dots N_3, \tag{4.1}$$

Where μ is the intercept, κ is the precision parameter and $\phi_1 \in [0, 1]$ is the weight between the Besag component $\boldsymbol{u}^{(1)}$ and i.i.d. component $\boldsymbol{v}^{(1)}$. As in Equation (3.4), these model components follow multivariate Gaussian distributions

$$\begin{aligned}\boldsymbol{u}^{(1)} &= (u_1^{(1)}, u_2^{(1)}, \dots u_{N_1}^{(1)}) \sim \mathcal{N}(0, (\mathbf{Q_B}^{(1)})^{-1}),\\ \boldsymbol{v}^{(1)} &= (v_1^{(1)}, v_2^{(1)}, \dots v_{N_1}^{(1)}) \sim \mathcal{N}(0, (\mathbf{Q_R}^{(1)})^{-1}),\end{aligned}$$

where $\mathbf{Q_B}^{(1)}$ and $\mathbf{Q_R}^{(1)}$ are the $N_1 \times N_1$ precision matrices for the Besag and i.i.d. components on admin 1 level, respectively.

Similarly, let $k[l]$ denote the index of the admin 2 area that admin 3 area l lies in. The model that considers only spatial effects on the admin 2 level is called 'Admin 2' and models $\boldsymbol{\eta}^{(3)}$ by

$$\eta_l^{(3)} = \mu + \frac{1}{\sqrt{\kappa}}\left(\sqrt{\phi_2}u_{k[l]}^{(2)} + \sqrt{1-\phi_2}v_{k[l]}^{(2)}\right), \quad l = 1, 2, \dots N_3, \tag{4.2}$$

where $\phi_2 \in [0, 1]$ is the weight between the Besag component $\boldsymbol{u}^{(2)}$, and i.i.d. component $\boldsymbol{v}^{(2)}$. We have

$$\begin{aligned}\boldsymbol{u}^{(2)} &= (u_1^{(2)}, u_2^{(2)}, \dots u_{N_2}^{(2)}) \sim \mathcal{N}(0, (\mathbf{Q_B}^{(2)})^{-1}),\\ \boldsymbol{v}^{(2)} &= (v_1^{(2)}, v_2^{(2)}, \dots v_{N_2}^{(2)}) \sim \mathcal{N}(0, (\mathbf{Q_R}^{(2)})^{-1}),\end{aligned}$$

with $\mathbf{Q_B}^{(2)}$ and $\mathbf{Q_R}^{(2)}$ being the $N_2 \times N_2$ precision matrices for the Besag and i.i.d. components on admin 2 level.

Finally, the same is done on the last administrative level through the model 'Admin 3', defined as

$$\eta_l^{(3)} = \mu + \frac{1}{\sqrt{\kappa}}\left(\sqrt{\phi_3}u_l^{(3)} + \sqrt{1-\phi_3}v_l^{(3)}\right), \quad l = 1, 2, \dots N_3, \tag{4.3}$$

where $\phi_3 \in [0,1]$ is the weight between the Besag component $\boldsymbol{u}^{(3)}$, and i.i.d. component $\boldsymbol{v}^{(3)}$. We have

$$\begin{aligned}\boldsymbol{u}^{(3)} &= (u_1^{(3)}, u_2^{(3)}, \dots u_{N_3}^{(3)}) \sim \mathcal{N}(0, (\mathbf{Q_B}^{(3)})^{-1}),\\ \boldsymbol{v}^{(3)} &= (v_1^{(3)}, v_2^{(3)}, \dots v_{N_3}^{(3)}) \sim \mathcal{N}(0, (\mathbf{Q_R}^{(3)})^{-1}),\end{aligned}$$

with $\mathbf{Q_B}^{(3)}$ and $\mathbf{Q_R}^{(3)}$ being the $N_3 \times N_3$ precision matrices for the Besag and i.i.d. components on admin 3 level. Note that the precision matrices for the Besag components in all the models are scaled in accordance with the method described in Section 3.3.

The intercept μ is included in the models to capture the mean level of the estimates. Thus the Besag and i.i.d. components are only intended to capture the patterns of deviation from the mean level. To ensure that this is the case, a sum-to-zero constraint is introduced to $\boldsymbol{u}^{(1)}$, $\boldsymbol{v}^{(1)}$, $\boldsymbol{u}^{(2)}$, $\boldsymbol{v}^{(2)}$, $\boldsymbol{u}^{(3)}$ and $\boldsymbol{v}^{(3)}$.

4.2.2 Multi-level model

A new application of BYM models is to combine spatial smoothing effects and i.i.d. effects on multiple administrative levels. The reasoning behind this is that countries usually have different policymakers on different administrative levels, such as India that is split into states, districts and subdistricts. The idea is that the correlation between neighbouring areas on the lower levels (admin 2 and admin3) is dependent on what types of borders they share. For example, an admin 3 area is likely to have a stronger correlation with a neighbouring admin 3 area that is also in the same admin 1 area, than a neighbour for which the border also separates them into different admin 1 areas. We define the suggested multi-level model for a random vector $\boldsymbol{\eta}^{(3)}$ on admin 3 level as

$$\begin{aligned}\eta_l^{(3)} = \mu + \frac{1}{\sqrt{\kappa}} \Big(&\sqrt{w_1} \left(\sqrt{\phi_1} u_{j[l]}^{(1)} + \sqrt{1-\phi_1} v_{j[l]}^{(1)} \right) \\ &+ \sqrt{w_2} \left(\sqrt{\phi_2} u_{k[l]}^{(2)} + \sqrt{1-\phi_2} v_{k[l]}^{(2)} \right) \\ &+ \sqrt{w_3} \left(\sqrt{\phi_3} u_l^{(3)} + \sqrt{1-\phi_3} v_l^{(3)} \right) \Big), \\ &l = 1, 2, \dots N_3,\end{aligned} \tag{4.4}$$

where $w_1, w_2, w_3 \in [0,1]$ are the weights of the spatial variation on each administrative level and $w_1 + w_2 + w_3 = 1$.

The multi-level model is used with the same scaling of precision matrices for the Besag components as the single-level models. This way the weights

assigned to $\mathbf{Q_B}^{(1)}$, $\mathbf{Q_B}^{(2)}$ and $\mathbf{Q_B}^{(3)}$ can be interpreted as approximately the proportion of the marginal variance that the components represent. Additionally, the total marginal variance is controlled solely by the precision parameter and is equal to κ^{-1}. Thus the parameters of the model are easy to interpret and compare when trying to understand correlation structures in real datasets.

4.3 Inclusion of covariates

The models presented so far only include an intercept and various structured and unstructured random effects. In an attempt to include more structured effects and improve predictive accuracy, the models can contain an extra component that considers local covariates in each administrative area, as described in Section 3.5. Another multi-level model is proposed where covariates are included together with the intercept as an extension to the model in Equation (4.4). This model is called 'ML base' and is defined for a random vector $\boldsymbol{\eta}^{(3)}$ on admin 3 level as

$$\begin{aligned}\eta_l^{(3)} = \boldsymbol{g}_l^{\mathrm{T}}\boldsymbol{\beta} + \frac{1}{\sqrt{\kappa}}\Big(&\sqrt{w_1}\left(\sqrt{\phi_1}u_{j[l]}^{(1)} + \sqrt{1-\phi_1}v_{j[l]}^{(1)}\right)\\ &+\sqrt{w_2}\left(\sqrt{\phi_2}u_{k[l]}^{(2)} + \sqrt{1-\phi_2}v_{k[l]}^{(2)}\right)\\ &+\sqrt{w_3}\left(\sqrt{\phi_3}u_l^{(3)} + \sqrt{1-\phi_3}v_l^{(3)}\right)\Big),\\ &l = 1, 2, \dots N_3,\end{aligned} \tag{4.5}$$

where $w_1, w_2, w_3 \in [0, 1]$ are the weights of the spatial variation on each administrative level and $w_1 + w_2 + w_3 = 1$. $\mathbf{G} = [\boldsymbol{g}_1, \boldsymbol{g}_2, \dots \boldsymbol{g}_{N_3}]$ contains the covariate values and $\boldsymbol{\beta}$ is a parameter vector with the intercept and covariate parameters, as described in Equation (3.6). Covariates are included in the single-level models in the same way, so that they can be compared to the multi-level model both with and without covariates.

4.4 Prior distributions of model parameters

In Section 3.6 the advantages and disadvantages of setting informative priors are discussed. To find out if vague prior distributions for the parameters are good enough, or if informative priors can be significantly better, the models are tested with varying choice of priors. The main interest lies in testing the importance of priors for the weights w_1, w_2 and w_3.

In order to select appropriate priors, the attributes of these weights must be

considered. They act as weights that measure the contribution to the total variation on admin 3 level from each administrative level, and are thus defined on the unit interval and should sum up to 1. Thus it is natural to set a Dirichlet prior for the three weights, as explained in Section 3.6.

The effects of different choices of prior distribution is investigated in Chapter 5 by varying the parameters $\boldsymbol{\alpha}$ in the Dirichlet priors. Then we can observe which priors achieve the highest accuracy during estimation of prevalence and estimation of model parameters. When selecting priors during application of the models to real data, which is done in Chapter 6, the results from the simulation study are considered. It is also possible to consider knowledge about the specific country of interest when selecting $\boldsymbol{\alpha}$. For example, if it is known that local authorities in admin 3 areas have the power to rule in almost whichever way they see fit, a bias towards the w_3 weight can be introduced through the prior.

The remaining parameters in the model in Equation (4.5) are more common, and have previously been studied. For example, in the paper by Riebler et al. (2016), different priors are discussed for the weight parameter ϕ between the Besag and i.i.d. components in the BYM model. Therefore, this thesis rather focuses on the weights between variation on the different administrative levels, which have not been extensively researched before.

The parameter vector $\boldsymbol{\beta}$ contains unrestricted parameters that are set to have vague Gaussian prior distributions $\boldsymbol{\beta} \sim \mathcal{N}(\mathbf{0}, 100^2 \cdot \mathbf{I})$. The precision parameter κ is restricted to the positive half-line, so that a gamma prior is a natural choice. Again a vague prior is used so that $\kappa \sim \text{Gamma}(1, 0.01)$. Finally, beta priors are used for the weights ϕ_1, ϕ_2 and ϕ_3 that are restricted to the unit interval. We choose $\phi_i \sim \text{Beta}(1.1, 1.1),\ i = 1, 2, 3$. This is close to a uniform prior, but gives lower probability of getting values near 0 and 1, as it is assumed to be unlikely that either the Besag or the i.i.d. component completely dominates the other component.

4.5 Estimation on coarser levels

The models presented so far allow estimation on the admin 3 level. However, there is also an interest in finding estimates on the coarser levels, ideally using the same models. By using the same model for estimates on all three levels, we can ensure consistency between estimates. It also makes for more explainable estimates when they all come from the same model, and using different models usually entails questions regarding why the same model cannot be used on coarser levels.

In order to obtain prevalence estimates on admin 2 and admin 1 level, weighted averages of admin 3 estimates can be used. The weights are based on what proportion of the population in the admin 1 area or admin 2 area that lives in each of the admin 3 areas it contains. Data from WorldPop (2018) is used as a basis for these weights.

Let Adm3(j) denote the set of admin 3 areas that lie inside admin 1 area j. Also let

$$\alpha_{lj} = \frac{\text{Population}_3(l)}{\text{Population}_1(j)}$$

be the weight assigned to the estimated prevalence in admin 3 area l in admin 1 area j. Prevalence estimates on logit scale on the admin 1 level are then made using

$$\eta_j^{(1)} = \text{logit}\left(\sum_{l \in \text{Adm3}(j)} \alpha_{lj} \text{logit}^{-1}(\eta_l^{(3)})\right), \quad j = 1, 2, \ldots N_1. \tag{4.6}$$

In a similar way, let Adm3(k) denote the set of admin 3 areas that lie inside admin 2 area k, and let

$$\alpha_{lk} = \frac{\text{Population}_3(l)}{\text{Population}_2(k)}$$

be the weight assigned to the estimated prevalence in admin 3 area l in admin 2 area k. The weights are calculated in the same way, with all weights belonging to the same admin 2 area summing to 1. Then we estimate prevalence on logit scale on the admin 2 level by

$$\eta_k^{(2)} = \text{logit}\left(\sum_{l \in \text{Adm3}(k)} \alpha_{lk} \text{logit}^{-1}(\eta_l^{(3)})\right), \quad k = 1, 2, \ldots N_2. \tag{4.7}$$

Note that when covariates are not included in the models, the 'Admin 1' model (4.1) gives estimates where any two admin 2 or admin 3 areas within the same admin 1 area get the same estimates, as only spatial variation on the admin 1 level is considered. Similarly, the 'Admin 2' model gives equal estimates in any two admin 3 areas within the same admin 2 area. These results are also approximately true when covariates are included, if the covariates cannot explain a significant proportion of the variation. Therefore, it is expected that the 'Admin 1' and 'Admin 2' models are not able to make very accurate estimates on the admin 3 level when there is limited information about relevant covariates.

To further motivate the use of a multi-level model, an example simulation is

created from the model in Equation (4.4) where $\{w_1, w_2, w_3\} = \{0.9, 0.05, 0.05\}$, meaning that 90% of spatial variation comes from the admin 1 level and the remaining variation is split equally between the other two levels. When the single-level model 'Admin 3' is used to make estimates on the admin 1 level it can not capture this correlation structure, and another weakness of the 'Admin 3' model is that it will underestimate the standard deviations of the estimates. In Figure 4.1 the results from applying the 'Admin 3' model to the example simulation are shown. For each admin 1 area the estimated values are plotted together with 95% confidence intervals. Theoretically only 5% of the simulated values should fall outside these confidence intervals, but as the plot shows this is the case for ≈31% of the areas.

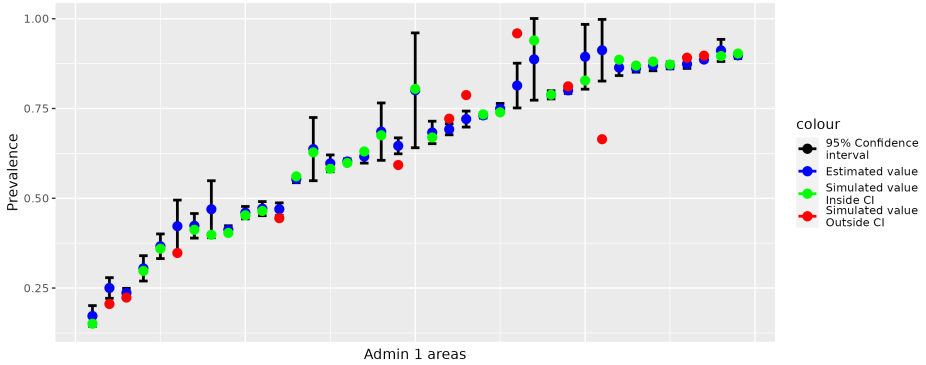


Figure 4.1: Prevalence estimates from the 'BYM - Admin 3' model on the admin 1 level. The estimates are plotted along with 95% confidence intervals and the correct, simulated values.

4.6 Model evaluation

Models are compared through two different criteria that measure the accuracy of the estimated $\boldsymbol{\eta}^{(a)}$ on the three levels $a = 1, 2, 3$. The first criterion is the

mean-squared error (MSE) defined as

$$\text{MSE}^{(1)}(\boldsymbol{\eta}^{(1)}, \hat{\boldsymbol{\eta}}^{(1)}) = \frac{1}{N_1}\sum_{j=1}^{N_1}(\eta_j^{(1)} - \hat{\eta}_j^{(1)})^2,$$

$$\text{MSE}^{(2)}(\boldsymbol{\eta}^{(2)}, \hat{\boldsymbol{\eta}}^{(2)}) = \frac{1}{N_2}\sum_{k=1}^{N_2}(\eta_k^{(2)} - \hat{\eta}_k^{(2)})^2,$$

$$\text{MSE}^{(3)}(\boldsymbol{\eta}^{(3)}, \hat{\boldsymbol{\eta}}^{(3)}) = \frac{1}{N_3}\sum_{l=1}^{N_3}(\eta_l^{(3)} - \hat{\eta}_l^{(3)})^2,$$

on administrative level 1, 2 and 3, respectively. Here, $\boldsymbol{\eta}^{(a)}$ are the true values and $\hat{\boldsymbol{\eta}}^{(a)}$ are the means of the estimated posterior distributions. The MSE is useful to see the accuracy of the point predictions of the models. However, it does not consider the distribution of the estimates and whether or not the real $\boldsymbol{\eta}^{(a)}$ values are deemed as plausible. For example, there can be two different estimates of a parameter with real value 2, with two different distributions such as in Figure 4.2. The model that produced the estimate with the mean closest to 2 is the preferred one if only MSE is used as a criterion. Yet the other estimate would in many cases be the preferred one as it has a much more realistic variance. This motivates another choice of criterion for model evaluation that can take into consideration the distribution of the estimates.

The second criterion is the continuous rank probability score (CRPS). It is calculated as

$$\text{CRPS}^{(1)}(\boldsymbol{\eta}^{(1)}, \hat{\boldsymbol{F}}^{(1)}) = \frac{1}{N_1}\sum_{j=1}^{N_1}\int_{\mathbb{R}}(\hat{F}_j^{(1)}(z) - 1\{z \geq \eta_j^{(1)}\})^2 dz,$$

$$\text{CRPS}^{(2)}(\boldsymbol{\eta}^{(2)}, \hat{\boldsymbol{F}}^{(2)}) = \frac{1}{N_2}\sum_{k=1}^{N_2}\int_{\mathbb{R}}(\hat{F}_k^{(2)}(z) - 1\{z \geq \eta_k^{(2)}\})^2 dz,$$

$$\text{CRPS}^{(3)}(\boldsymbol{\eta}^{(3)}, \hat{\boldsymbol{F}}^{(3)}) = \frac{1}{N_3}\sum_{l=1}^{N_3}\int_{\mathbb{R}}(\hat{F}_l^{(3)}(z) - 1\{z \geq \eta_l^{(3)}\})^2 dz,$$

where $\hat{\boldsymbol{F}}^{(a)}$ denotes the posterior cumulative distribution of $\boldsymbol{\eta}^{(a)}$. This is a proper scoring rule that is commonly used to compare predictive distributions (Matheson and Winkler, 1976). The goal is then to identify which models perform best with respect to both MSE and CRPS, where a low CRPS is the most important metric.

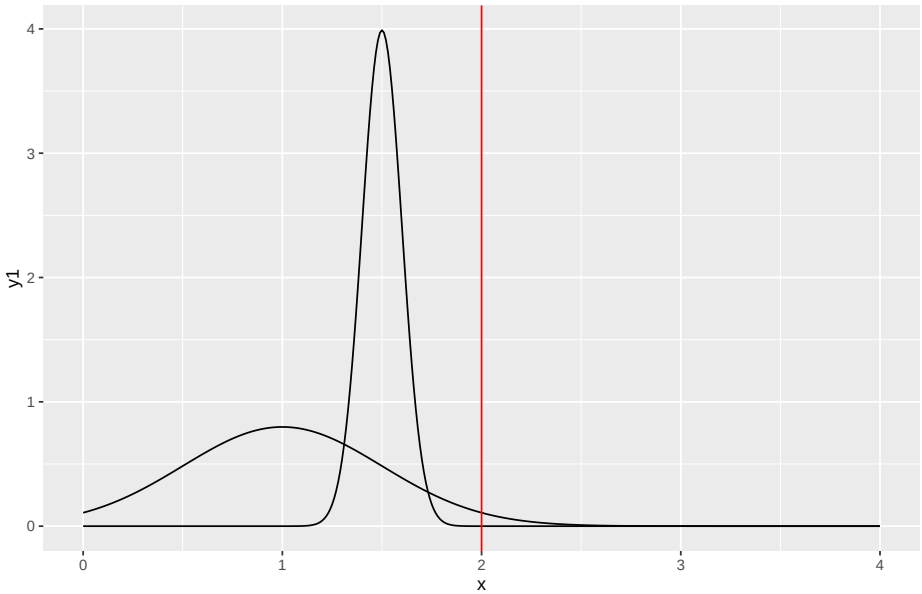


Figure 4.2: Distribution of two example estimates of a parameter where the real value is 2.

In Chapter 6 the spatial models are applied to the real DHS dataset from India. When working with real data the correct values of the model parameters are unknown, as opposed to during a simulation study. Thus the interest lies in observing the accuracy of the estimates when compared to direct estimates. This is done using 10-fold cross-validation. Traditionally, this means that the data is split into 10 groups, with each group containing the data from $\approx$10% of the admin 3 areas. The models are fitted to data from 9 of the groups, and then used to make estimates on the tenth group acting as a validation set. These estimates are compared to the 'real' values (direct estimates) using MSE or CRPS. The process is repeated 10 times, with each of the 10 groups acting as the validation set one time each.

However, we do not have reliable direct estimates in a large proportion of the admin 3 areas. Therefore, only admin 3 areas with a sufficient amount of data to use for direct estimates are included in the validation sets. 10-fold cross-validation is then performed as such:

- Split the admin 3 areas into two sets. Set 1 contains the areas that have reliable direct estimates and set 2 contains the remaining areas.
- Split set 1 into 10 groups as in traditional cross-validation.
- Fit the model to the data from 9 of the 10 groups from set 1 *and* all of set 2.
- Make estimates in the areas from the last group from set 1 and compute MSE and CRPS on these areas.
- Repeat for each of the ten groups.

The models applied to the admin 3 level in India are then evaluated by the two metrics

$$\text{MSE}^{(3)} = \frac{1}{10}\sum_{i=1}^{10}\text{MSE}^{(3)}(\boldsymbol{\eta_i}^{(3)}, \hat{\boldsymbol{\eta}}_{\boldsymbol{i}}^{(3)})$$

$$\text{CRPS}^{(3)} = \frac{1}{10}\sum_{i=1}^{10}\text{CRPS}^{(3)}(\boldsymbol{\eta_i}^{(3)}, \hat{\boldsymbol{F}}_{\boldsymbol{i}}^{(3)}),$$

where $\boldsymbol{\eta_i}^{(3)}$ are the direct estimates in the admin 3 areas in group i, and $\hat{\boldsymbol{\eta}}_{\boldsymbol{i}}^{(3)}$ and $\hat{\boldsymbol{F}}_{\boldsymbol{i}}^{(3)}$ are the means and cumulative distributions of the estimated posterior distributions in the same areas, which are estimated by a model that is fitted to all data *except* that from areas in group i.

During the process of cross-validation on admin 3 level, error estimates are also made on the admin 1 and admin 2 levels. This is done by keeping the estimated $\hat{\boldsymbol{\eta}}^{(3)}$ after each iteration of the cross-validation and performing the following steps:

- Use $\hat{\boldsymbol{\eta}}^{(3)}$ to estimate $\hat{\boldsymbol{\eta}}^{(1)}$ on the admin 1 level using Equation (4.6) and $\hat{\boldsymbol{\eta}}^{(2)}$ on the admin 2 level using Equation (4.7).
- Split the admin 1 and admin 2 areas into two sets using the same criteria as for the admin 3 areas, where set 1 contains areas with reliable direct estimates.
- For each of the 10 iterations, compute

$$\text{MSE}^{(1)}(\boldsymbol{\eta}^{(1)}, \hat{\boldsymbol{\eta}}^{(1)}),$$

$$\text{CRPS}^{(1)}(\boldsymbol{\eta}^{(1)}, \hat{\boldsymbol{F}}^{(1)}),$$

$$\text{MSE}^{(2)}(\boldsymbol{\eta}^{(2)}, \hat{\boldsymbol{\eta}}^{(2)}),$$

$$\text{CRPS}^{(2)}(\boldsymbol{\eta}^{(2)}, \hat{\boldsymbol{F}}^{(2)}),$$

where $\boldsymbol{\eta}^{(1)}$ and $\boldsymbol{\eta}^{(2)}$ are direct estimates in the admin 1 and admin 2 areas in set 1, and $\hat{\boldsymbol{\eta}}^{(1)}$, $\hat{\boldsymbol{F}}^{(1)}$, $\hat{\boldsymbol{\eta}}^{(2)}$ and $\hat{\boldsymbol{F}}^{(2)}$ are the means and cumulative distributions of the estimated posterior distributions in the corresponding areas on the admin 1 and admin 2 levels, respectively.

- Compute the final MSE and CRPS on each level as the means of the resulting MSE and CRPS from the 10 iterations.

Using the results we can find out if the models are likely to make accurate estimates of variables in small areas, where direct estimates are unfeasible, while being consistent with reliable direct estimates on the coarser administrative levels.

4.7 Implementation details

The `R` package called Template Model Builder (TMB) is used as the main tool to achieve fast parameter estimation for the binomial spatial regression models. The package combines Automatic Differentiation with the Laplace approximation along with techniques for parallel computations. AD is implemented using operator overloading in `C++` from the package CppAD. From this the gradient and Hessian are obtained for further use in the Laplace approximation to the marginal likelihood. TMB allows the user to differentiate between random and fixed effects, so that the random ones can be integrated out when evaluating the marginal likelihood.

In addition, the implementation of TMB includes automatic sparsity detection which enables much faster computations when performing operations on matrices. The sparse matrix operations are further accelerated through the use of parallel basic linear algebra subprograms (BLAS). This is especially important when working with GMRFs because of the heavy computations needed to perform Cholesky factorization and the reverse subset algorithm, which can be done in parallel.

When using the TMB package, the pre- and post processing of data has to be done in `R`, whereas the model definitions, likelihood functions and reporting of parameter estimates are implemented using `C++`. Despite the fast computations that follow from this, it is required that all precision matrices, weights and input data are generated and correctly formatted before sending them into `C++`. Thus there is a lot of preparatory, time consuming work that needs to be done before the TMB package becomes useful.

The first step of implementation is to generate the precision matrices from all three administrative levels and scale them properly. To do this the neighbourhood structures must be created, while not including island regions. Here it is important to store the indices of the removed regions, as the same regions must not be included when reading data from other sources, such as the DHS and WorldPop data.

DHS data is read to count and map the responses to the administrative areas. This also entails challenges as the survey data contains 5972 columns and over 800 000 rows. When reading this data one must be careful only to count respondents within the chosen age groups, that do not live on an island territory. The same measures are taken when reading the WorldPop data, which also consist of very large data files. India has an area of 3 287 000 km^2, meaning that on a 100m$\times$100m grid there are over 300 million grid cells that all have to be counted and mapped to their respective admin 1, admin 2 and admin 3 areas.

After having prepared all the necessary data and run the model through TMB, there is some post processing that needs to be done. This includes extracting all parameters of interest from `C++` along with their standard deviations, computing MSE and CRPS from the estimates on the three administrative levels, and storing all the results.

As a result of everything that has to be in place before applying the spatial models, it can take multiple days to get any results despite the speed of the implementation in TMB. This is not only the case in India, as the same data processing is necessary in all countries where these models are useful. On the other hand, most of the time consuming data processing only has to be done once for a single country. Thus, after everything is prepared, TMB is very useful to quickly run different models on lots of datasets. Therefore, it is possible to conduct the simulation study with hundreds of simulations in India, which is presented in the following chapter.

Chapter 5

Simulation

5.1 Purpose

In this chapter a simulation study is conducted to compare the predictive accuracy of the proposed spatial models. The goal is to determine whether or not a multi-level model should be used to make estimates on all levels rather than single-level models when there is variation on multiple administrative levels. In addition we are interested in observing if the model parameters are correctly estimated for the multi-level model, especially the parameters for the covariates. Lastly, the multi-level model is tested with different choices of prior distributions for the weights w_1, w_2 and w_3, so that the effect of the priors on predictive accuracy and parameter estimation can be observed. A set of key questions is specified as the main focus during analysis of the results:

- Do estimates from the multi-level model generally have higher predictive accuracy than from the single-level models?
- How does the choice of prior distribution of the weights affect parameter estimation?
- Are the models able to correctly estimate the covariate parameters in $\boldsymbol{\beta}$?
- How significant is the impact on predictive accuracy when using reduced amounts of simulated data to fit the models?

By using the TMB library to run simulations and evaluate the models, the techniques for fast computations described in Section 3.7 are also put to the test. This allows us to observe how the implementation of these techniques enables extensive testing of the spatial models, and to quantify the approximate time needed to fit the models to real data.

5.2 Simulation setup

To create realistic simulations of prevalence mappings across India, the neighbourhood structures on all three administrative levels are extracted from shapefile data obtained from the Database of Global Administrative Areas (GADM). The neighbourhood structures are used to create precision matrices for the Besag components in the multi-level model defined in (4.5) and the single-level models. After deciding upon appropriate values for the parameters $\mu, \kappa, w_1, w_2, w_3, \phi_1, \phi_2, \phi_3, \boldsymbol{\beta}$, the GMRFs $\boldsymbol{u}^{(1)}, \boldsymbol{v}^{(1)}, \boldsymbol{u}^{(2)}, \boldsymbol{v}^{(2)}, \boldsymbol{u}^{(3)}, \boldsymbol{v}^{(3)}$ are simulated and used to compute $\boldsymbol{\eta}^{(3)}$ on admin 3 level. Then a number of trials $n_l^{(3)}$ is chosen in each admin 3 area l, and the number of positive responses is simulated from $Y_l^{(3)}|\eta_l^{(3)} \sim \text{Binomial}(n_l^{(3)}, \text{logit}^{-1}(\eta_l^{(3)}))$. After simulation, the vector of numbers of responses in each admin 3 area $\boldsymbol{n}^{(3)}$, and numbers of positive responses $\boldsymbol{Y}^{(3)}$ are used as inputs to fit the models that estimate $\boldsymbol{\eta}^{(3)}$.

Selected models

There is a total of eight different models used in the simulation study. The first three are the single-level models named 'Admin 1', 'Admin 2' and 'Admin 3' that are presented in Section 4.2.1, with covariates included. The remaining five are all based on the multi-level model in (4.5) using different choices for the parameter vector $\boldsymbol{\alpha}$ that controls the prior distribution of the weights w_1, w_2 and w_3. The first of these models is called 'ML base', which has a flat prior $\boldsymbol{\alpha} = [1, 1, 1]$. Thus there is no bias towards any of the weights which means that this will serve as a 'control' model when observing the effects of other priors.

The next three models are assigned the parameters $\boldsymbol{\alpha} = [7, 1, 1]$, $\boldsymbol{\alpha} = [1, 7, 1]$ and $\boldsymbol{\alpha} = [1, 1, 7]$. This makes them have a bias towards assigning weight to the variation from the administrative level that has the highest value in $\boldsymbol{\alpha}$. They are named 'ML prior 1', 'ML prior 2' and 'ML prior 3', respectively. The parameters are chosen with the expectation that they are large enough to give observable differences in results between the models, while not introducing unrealistically high bias. Prior distributions of the weights in 'ML prior 1' are plotted in Figure 5.1. Here, the initial bias towards variation on the admin 1 level is clear, and similar plots can be made that show an equal bias towards variation on the admin 2 level in 'ML prior 2', and on the admin 3 level in 'ML prior 3'.

For the final model the parameters $\boldsymbol{\alpha} = [3, 3, 3]$ are used. This entails a bias towards having equal weights between the variation on each administrative level. Thus it is less likely to assign negligible weight to any of the three levels, and is expected to perform better when there is in fact considerable variational

effects happening on every level.

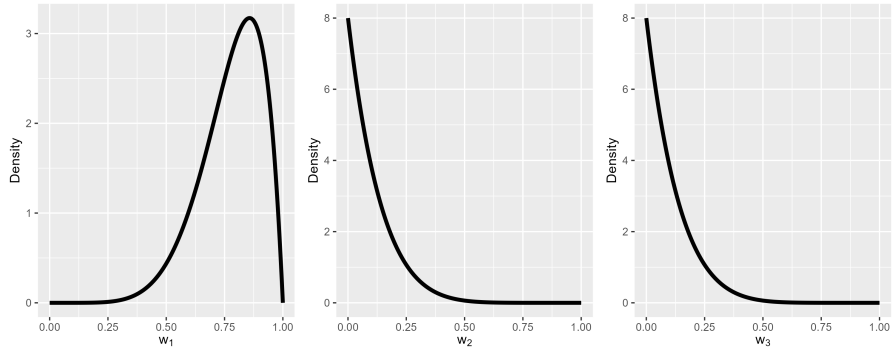


Figure 5.1: Prior marginal distributions of the weights w_1, w_2 and w_3 using the parameters $\boldsymbol{\alpha} = [7, 1, 1]$.

Variables between selected scenarios

For the weights w_1, w_2 and w_3 a set of different combinations are tested, similar to what is done for $\boldsymbol{\alpha}$. Varying these weights will firstly enable us to see how the multi-level model compares to the single-level models in different scenarios. Secondly, we can see the effect the choice of prior distribution of the weights has when accurate and inaccurate priors are used in the same scenarios. The sets of weights that are used are

$$\begin{aligned} \boldsymbol{w}_{\text{Adm1}} &= \{w_1, w_2, w_3\} = \{0.8, 0.1, 0.1\}, \\ \boldsymbol{w}_{\text{Adm2}} &= \{w_1, w_2, w_3\} = \{0.1, 0.8, 0.1\}, \\ \boldsymbol{w}_{\text{Adm3}} &= \{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}, \\ \boldsymbol{w}_{\text{Even}} &= \{w_1, w_2, w_3\} = \{0.4, 0.3, 0.3\}. \end{aligned} \tag{5.1}$$

These weights are used so that the majority (80%) of the total variance comes from either of the three administrative level, and so that the variance is more evenly distributed when using the last combination of weights. Thus each of the multi-levels models will have a prior $\boldsymbol{\alpha}$ that is fairly accurate in at least one scenario. However, the prior expected values are deliberately set to not be 100% correct for any set of weights, as choosing perfectly accurate priors is unrealistic when working with real data. Also, it is interesting to see if the slight inaccuracy of the prior is made up for during parameter estimation.

The final aspect of the simulation study that is experimented with is the amount of simulated trials within each area, $\boldsymbol{n}^{(3)}$. It is interesting to see how this affects the predictive accuracy, as the amount of available data varies a lot in the real world. For example, the DHS survey in India has significantly more responses related to education than employment among females in the 20-39 age group. Three levels of simulated data are used to research how much of an impact this difference has on predictions.

First, 100 trials are simulated in each admin 3 area as this approximately corresponds to the amount of responses on education from the DHS survey, with the main difference being that responses are not evenly distributed in the survey. Second, the number is reduced to 50 trials in each admin 3 area, with a randomly selected 10% of the areas having 0 trials. This reflects how surveys often lack responses from some areas. Lastly, the amount of data is reduced even further to 20 trials in each admin 3 area, and 25% of them having 0 trials. This leads to slightly less total trials than the amount of responses concerning employment in the DHS survey.

Choice of values for constant parameters

The remaining parameters $\mu, \kappa, \phi_1, \phi_2, \phi_3$ and $\boldsymbol{\beta}$ are all kept constant throughout the simulation study. Although experimentation with these parameters also can lead to findings of interest, this has already been done in the project by Giørtz (2022). Keeping them unchanged allows us to focus on effects of the parameters that are more important for answering the questions posed in this thesis. Values for the parameters are chosen to achieve simulations that reflect realistic data, such as the data on education from the DHS survey in India. Based on the results from the direct estimates presented in Figure 2.2, the prevalences of completed secondary education typically lie between 0.5 and 1. Thus the intercept is set to

$$\mu = 0.8,$$

where the resulting prevalence is $\text{logit}^{-1}(0.8) = 0.69$. Further, the precision parameter is set to

$$\kappa = 0.5,$$

giving a standard deviation on logit scale equal to 1.41 which is close to that of the usable direct estimates.

The parameters ϕ_1, ϕ_2 and ϕ_3 control the proportion of variance coming from the Besag component and i.i.d. component on each of the administrative levels. In the simulation study we want there to be clear spatial correlation structures, so that the weight of the i.i.d. components should be kept relatively

low. However, having no random effect would be unlikely in any realistic case. Therefore, the choice is to set

$$\phi_1 = \phi_2 = \phi_3 = 0.8,$$

meaning that 80% of the total marginal variance from the spatial components in the simulation is structured variation from the Besag components.

Finally, in order to choose $\boldsymbol{\beta}$ it is important to remember that the main part of the models presented in this thesis is the spatial components. Thus it is undesirable that the variation that is due to covariates dominates the spatial variation. However, the weight of covariates should be significant enough for it to be useful to include them in the models. The choice is to fix $\boldsymbol{\beta}$ so that approximately 25% of total variation comes from covariates, and the remaining 75% comes from spatial effects. From the model in Equation (4.5) the variance is

$$\text{Var}(\eta_l^{(3)}) = \boldsymbol{\beta}^T \text{Var}(\mathbf{G})\boldsymbol{\beta} + \frac{1}{\kappa}, \tag{5.2}$$

assuming that the covariates and spatial effects are independent, and using the empirical covariance matrix of the covariate values in $\mathbf{G}$. Through experimentation it was found that using

$$\boldsymbol{\beta} = [1, 0.75, 0.5]$$

gives the desired distribution of variance when $\kappa = 0.5$. Here, the first parameter $\beta_0 = \mu$ defaults to the weight of the intercept, and the second and third parameters β_p and β_n belong to the logarithms of the covariates 'population density' and 'nighttime light per person', respectively.

In the following sections the different models are applied to 100 simulations of each combination of the suggested amounts of data and choice of weights $\{w_1, w_2, w_3\}$. The predictive accuracy of the models are compared in each scenario based on average MSE and CRPS. Plots of the estimated model parameters from the multi-level models are also presented to see whether or not the models are able to estimate the parameters used for simulations, and help observe the effects of using different priors.

5.3 Comparison of multi-level and single-level models

We begin by looking at how the multi-level model performs in comparison to the single-level models, without using any informative priors for the model parameters. The performance of the multi-level model is assessed through measurements of mean MSE and CRPS. These are calculated based on 100 simulations of 100 trials in each admin 3 area, and repeated for all of the four weight combinations in Equation (5.1). The results on the admin 1 level, admin 2 level and admin 3 level are shown in Table 5.1, Table 5.2, Table 5.3, respectively. We can first observe that the 'Admin 1' model has on average the highest errors, especially in terms of CRPS, when the majority of variation does not happen on the admin 1 level. It is also clear that the 'Admin 1' model is not a good option for estimation on the admin 2 or admin 3 level, as the errors here are significantly higher than from the other models.

A closer look at the errors on admin 3 level from Table 5.3 shows that the multi-level model performs better than the 'Admin 2' model in all four scenarios. The 'Admin 3' model has only slightly lower MSE than the multi-level model when using the weights $\boldsymbol{w}_{\text{Adm3}}$, but is otherwise also outperformed by it. Meanwhile, there are no scenarios in which the multi-level model has significantly higher errors than any of the single-level models. This is clear evidence that a multi-level model is a more reliable choice when modelling data where the distribution of variation is unknown.

Similar comparisons are also done with reduced amounts of simulated data. In these cases the multi-level model still outperforms the single-level models, with the differences in MSE and CRPS being even more evident when applying the models to smaller sets of data. For a complete overview of the simulations that were run, results from each of the models in all different scenarios are collected in Appendix A. Further comparison of multi-level and single-level models can also be found in the project thesis Giørtz (2022).

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
Admin 1	0.47	0.83	0.83	0.68	0.91	1.38	1.44	1.21
Admin 2	0.60	0.39	0.51	0.50	0.86	0.69	0.87	0.80
Admin 3	0.90	0.45	0.38	0.49	1.06	0.72	0.68	0.74
ML Base	0.42	0.40	0.41	0.37	0.69	0.67	0.69	0.67

Table 5.1: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 1 level from the single-level models and the multi-level base model when applied to simulations with each weight combination.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
Admin 1	1.88	8.93	7.37	5.44	3.16	7.21	6.40	5.53
Admin 2	0.53	0.62	0.91	0.67	1.08	1.16	1.50	1.24
Admin 3	0.54	0.68	0.62	0.52	1.09	1.21	1.11	1.08
ML Base	0.36	0.62	0.66	0.49	0.93	1.12	1.12	1.06

Table 5.2: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 2 level from the single-level models and the multi-level base model when applied to simulations with each weight combination.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
Admin 1	2.33	9.35	11.72	6.95	3.64	7.49	8.41	6.44
Admin 2	0.89	0.94	5.73	2.26	1.86	1.89	5.25	3.16
Admin 3	0.57	0.73	0.81	0.66	1.28	1.41	1.48	1.37
ML Base	0.44	0.56	0.82	0.63	1.14	1.23	1.47	1.34

Table 5.3: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 3 level from the single-level models and the multi-level base model when applied to simulations with each weight combination.

5.4 Effect of choice of priors on predictive accuracy

Four variations of the multi-level model are suggested as alternatives to the 'ML base' model. These have different choices of parameters to the prior distributions of the weights w_1, w_2 and w_3, and are described in Section 5.2. We want to investigate how the choice of priors affects the predictive accuracy of the model in

different scenarios. This is done by applying the models to the same simulations as in Section 5.3. The results on the admin 1 level, admin 2 level and admin 3 level are shown in Table 5.4, Table 5.5, Table 5.6, respectively.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
ML Base	0.42	0.40	0.41	0.37	0.69	0.67	0.69	0.67
ML Prior 1	0.40	0.38	0.38	0.37	0.67	0.66	0.68	0.68
ML Prior 2	0.39	0.33	0.41	0.35	0.69	0.66	0.70	0.65
ML Prior 3	0.45	0.35	0.36	0.46	0.71	0.66	0.66	0.71
ML Prior 4	0.39	0.34	0.40	0.38	0.67	0.66	0.70	0.69

Table 5.4: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 1 level from the multi-level models when applied to simulations with each weight combination.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
ML Base	0.36	0.62	0.66	0.49	0.93	1.12	1.12	1.06
ML Prior 1	0.36	0.59	0.59	0.53	0.94	1.11	1.09	1.07
ML Prior 2	0.37	0.61	0.65	0.49	0.95	1.12	1.12	1.05
ML Prior 3	0.37	0.56	0.56	0.52	0.95	1.10	1.08	1.06
ML Prior 4	0.36	0.61	0.66	0.54	0.94	1.12	1.12	1.08

Table 5.5: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 2 level from the multi-level models when applied to simulations with each weight combination.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
ML Base	0.44	0.56	0.82	0.63	1.14	1.23	1.47	1.34
ML Prior 1	0.44	0.54	0.80	0.65	1.14	1.23	1.47	1.36
ML Prior 2	0.44	0.55	0.82	0.63	1.15	1.24	1.48	1.34
ML Prior 3	0.45	0.53	0.78	0.63	1.15	1.22	1.46	1.34
ML Prior 4	0.44	0.55	0.83	0.65	1.15	1.24	1.49	1.36

Table 5.6: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 3 level from the multi-level models when applied to simulations with each weight combination.

The results show only small differences in predictive accuracy for any choice of prior distribution. For example, when measuring error on the admin 3 level from the 'ML Prior 3' model on data simulated with the weights $\boldsymbol{w}_{\text{Adm3}}$, one could expect a more significant decrease in error, as the model has an accurate prior. Table 5.6 shows a maximum relative difference in MSE between the 'ML Prior 3'

and 'ML Prior 4' models of approximately 5%. Thus it seems as though setting informative priors has a limited effect in terms of obtaining accurate predictions. Scaling up the parameter vector $\boldsymbol{\alpha}$ decreases the variance of the priors, similar to in Figure 3.3. This leads to more strict informative priors, that can have more impact on predictions. However, significantly scaling $\boldsymbol{\alpha}$ implies high certainty of the values of the weights $\{w_1, w_2, w_3\}$, which is unlikely to be justified in any realistic case. Therefore, it is not a priority to investigate this further.

On the other hand, sometimes we are not only interested in predictions, but also in parameter estimates in order to help understand the structure of the distribution of variation across different levels. Herein lies also possible explanations as to why the choice of priors have so little effect on MSE and CRPS. Thus we look further into the estimation of model parameters from the different suggested models.

5.5 Estimation of model parameters using different priors

By setting informative priors through the use of the Dirichlet parameters $\boldsymbol{\alpha}$, bias is introduced during estimation of model parameters. Not only does it affect estimation of the directly targeted parameters w_1, w_2 and w_3, but the bias can also have an added effect onto other parameters. In order to observe changes in parameter estimation from using different priors, we first look at how the 'ML Base' model without any informative priors performs.

Figure 5.2 and Figure 5.3 show parameter estimates from the 'ML Base' model in two different scenarios. The figures show that the model is able to accurately estimate the parameters μ, β_p and β_n in both cases, and this was a trend throughout the simulation study. Thus it is clear that the model is able to capture the effect of covariates when they contribute to a considerable proportion of the variation in the data. As there were only minor differences between estimates of these three parameters in all the simulations, the inclusion of covariates is not discussed further in this chapter. Instead, the importance of including covariates is explored further when working with real data in Chapter 6.

Another takeaway from the box plots is that the model struggles with estimating the weights between the Besag and i.i.d. components on the coarse administrative levels. Especially the parameter ϕ_1 is rarely accurate. This is due to there only being 39 admin 1 areas that are used, so that the amount of data points is too small to properly capture the correlation structure between them.

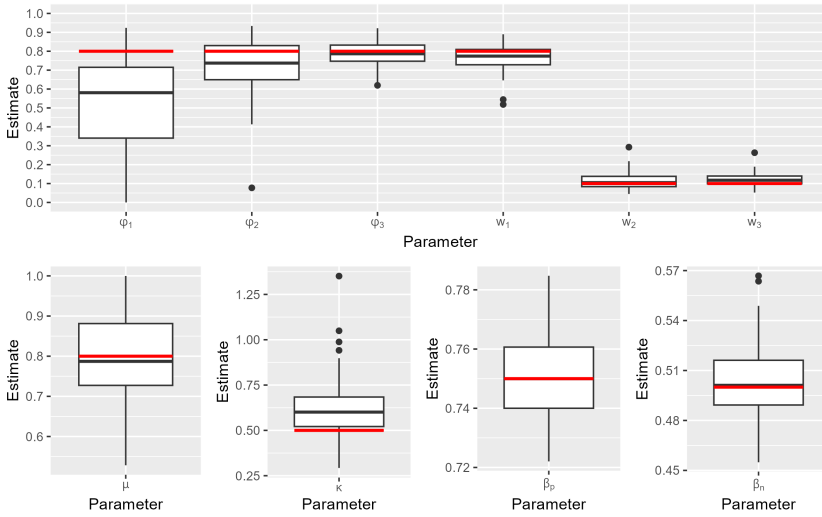


Figure 5.2: Estimated parameters using the 'ML Base' model on 100 different simulations with $\{w_1, w_2, w_3\} = \{0.8, 0.1, 0.1\}$, where 100 trials were simulated in each admin 3 area. The actual values that were used for simulation are marked with the red lines.

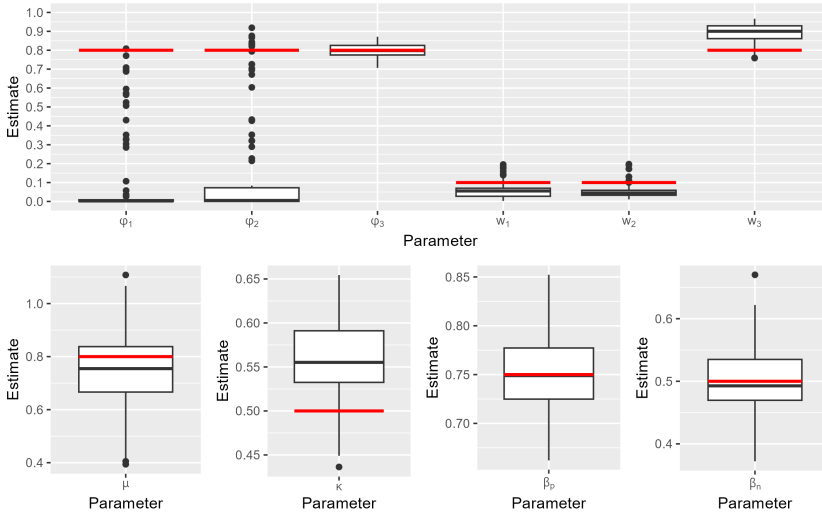


Figure 5.3: Estimated parameters using the 'ML Base' model on 100 different simulations with $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$, where 100 trials were simulated in each admin 3 area. The actual values that were used for simulation are marked with the red lines.

In Figure 5.3 the weights $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$ are used for simulation. Thus the proportion on variation coming from the admin 1 and admin 2 levels is very small, making it substantially harder to accurately estimate ϕ_1 and ϕ_2. Underestimation of these weight means that the i.i.d. components of the BYM models are assigned too much weight. Thus it seems as if the model has overestimated the total amount of randomness. However, this is compensated for through increasing the estimated value of κ, which makes the total variation decrease.

The results showed multiple ways in which estimation of κ varies depending on estimates of the various weight parameters. At first, the inverse proportionality with estimates of ϕ_1, ϕ_2 and ϕ_3 is clear. In addition, in Figure 5.4 we can see that when the prior is erroneously biased towards w_1, there is a direct impact on the estimates of w_1, w_2 and w_3, but this also leads to smaller values of κ. This occurs when too much weight is assigned to the coarse administrative levels, as variation from these levels leads to less randomness, which in turn is compensated for through underestimation of κ. The effects of the weights on the randomness

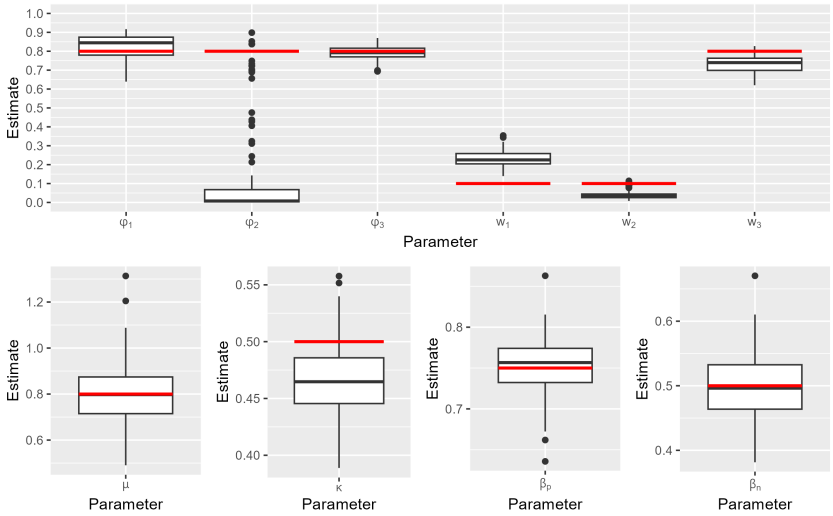


Figure 5.4: Estimated parameters using the 'ML Prior 1' model on 100 different simulations with $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$, where 100 trials were simulated in each admin 3 area. The actual values that were used for simulation are marked with the red lines.

in the data is illustrated in Figure 5.5, where two simulations are created with the same value of κ, but different sets of weights. There is clearly less total randomness between the areas in the left plot where w_1 is dominant, as the within-state variation is much lower than in the right plot where w_3 is dominant.

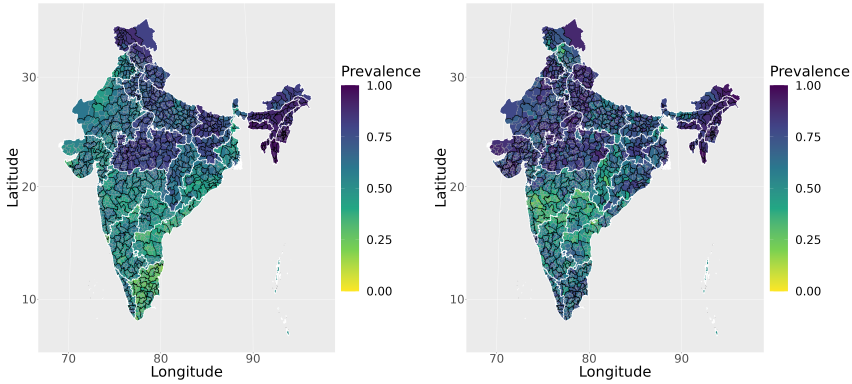


Figure 5.5: Two simulations of prevalence in India. The left plot was made with the weights $\{w_1, w_2, w_3\} = \{0.8, 0.1, 0.1\}$, and the right plot used $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$.

Finally, there is a clear trend among the models 'ML prior 1', 'ML prior 2' and 'ML prior 3' that even when the models are correctly biased towards a weight w_1, w_2 or w_3 that is dominant, they tend to overestimate the weight towards which they are biased. Thus using a prior that is expected to increase accuracy of parameter estimation in a certain scenario may have the opposite effect. An example of this phenomenon can be seen in Figure 5.6 and Figure 5.7. When using a prior with bias towards w_3, this weight is estimated too high, which was already a problem for the 'ML base' model in Figure 5.3. Thus the accuracy of parameter estimation got worse even though the choice of prior is reasonable. However, as the dominant weights were easily overestimated, using the prior $\boldsymbol{\alpha} = [3, 3, 3]$ often gave the best parameter estimates by counteracting this effect. This was especially clear in cases where the weights for fine scale level were dominant, but the 'ML prior 4' model actually seemed to give either better or just as good parameter estimates as the 'ML base' model in all cases.

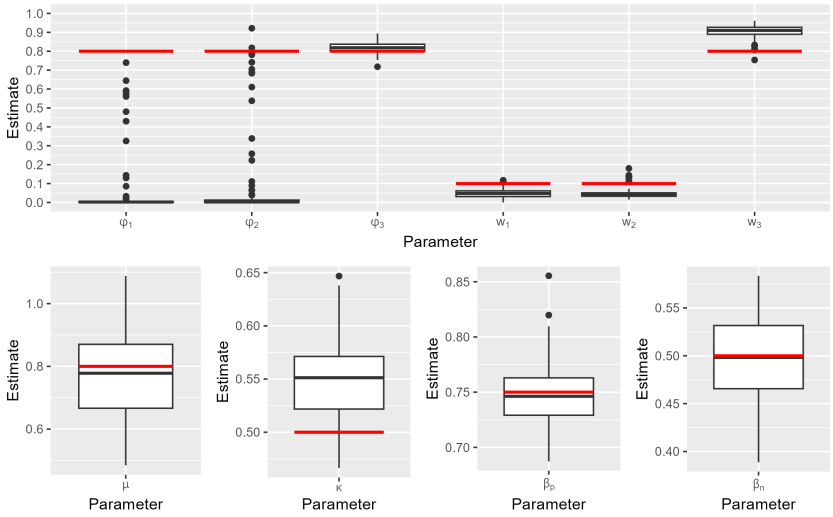


Figure 5.6: Estimated parameters using the 'ML prior 3' model on 100 different simulations with $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$, where 100 trials were simulated in each admin 3 area. The actual values that were used for simulation are marked with the red lines.

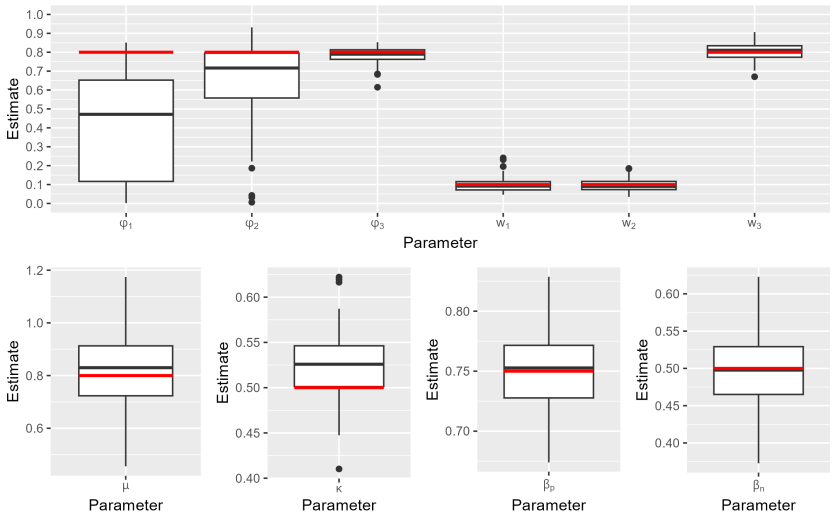


Figure 5.7: Estimated parameters using the 'ML prior 4' model on 100 different simulations with $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$, where 100 trials were simulated in each admin 3 area. The actual values that were used for simulation are marked with the red lines.

5.6 Predictions with reduced amount of simulated data

In the final part of the simulation study we want to research how the performance of the multi-level model varies when applied to data sets of different sizes. In the previous sections, each simulation has consisted of 100 simulated trials within each admin 3 area. We label this a 'large' data amount to which the model is expected to be fitted well. Two other data amounts are used for the remaining simulations. The 'medium' data amount consists of 50 simulated trials within each admin 3 area, but in a randomly selected 10% of the areas there are 0 simulated trials. This is taken a step further in the simulations with a 'small' data amount. Here, there are 20 simulated trials within each admin 3 area, with a randomly selected 25% of them having 0 trials.

The results on the admin 1 level, admin 2 level and admin 3 level from the multi-level model using different data amounts are shown in Table 5.7, Table 5.8, Table 5.9, respectively. The errors are calculated as the mean error from the five different versions of the multi-level model, each having one of the prior distributions for the parameters specified earlier. There is a clear increase in error when reducing the amount of data, which is to be expected. However, this increase is actually slower than the increase that would occur when using direct estimates on the same amount of data, even when disregarding the fact that zero trials are simulated in a proportion of the areas.

For direct estimates the expected error can be approximated through the variance

$$\mathrm{Var}(\hat{\eta}) = \mathrm{Var}(\hat{p}) \cdot \left(\frac{d}{dp}\mathrm{logit}(p)\right)^2 = \frac{1}{n}p(1-p)\left(\frac{1}{p} + \frac{1}{1-p}\right)^2.$$

To approximate this we simulate the prevalence in 10^5 independent areas, using Equation (5.2) to obtain

$$p = \mathrm{logit}^{-1}(\eta), \quad \eta \sim \mathcal{N}(\mu, \boldsymbol{\beta}^T \mathrm{Var}(\mathbf{G})\boldsymbol{\beta} + \frac{1}{\kappa}) = \mathcal{N}(0.8, 2.667).$$

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
Large	0.38	0.34	0.40	0.38	0.68	0.67	0.70	0.70
Medium	0.74	0.56	0.57	0.65	0.95	0.86	0.89	0.93
Small	1.39	1.05	1.01	1.11	1.38	1.23	1.22	1.29

Table 5.7: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 1 level aggregated from the five multi-level models when applied to simulations with each weight combination.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
Large	0.36	0.61	0.67	0.53	0.94	1.13	1.13	1.09
Medium	0.58	0.91	0.97	0.87	1.21	1.46	1.46	1.43
Small	1.00	1.71	1.65	1.56	1.61	2.07	2.02	1.99

Table 5.8: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 2 level aggregated from the five multi-level models when applied to simulations with each weight combination.

Model	MSE (10^1)				CRPS (10^1)			
	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$	$\boldsymbol{w}_{\text{Adm1}}$	$\boldsymbol{w}_{\text{Adm2}}$	$\boldsymbol{w}_{\text{Adm3}}$	$\boldsymbol{w}_{\text{Even}}$
Large	0.45	0.56	0.83	0.65	1.16	1.23	1.48	1.36
Medium	0.67	0.84	1.36	1.07	1.42	1.55	1.96	1.77
Small	1.06	1.43	2.53	1.87	1.79	2.03	2.75	2.37

Table 5.9: Error estimates of the estimated $\boldsymbol{\eta}$ on admin 3 level aggregated from the five multi-level models when applied to simulations with each weight combination.

After simulating 10^5 prevalences, computing $\text{Var}(\eta)$ for each of these and aggregating over the results, we get that for simulations with a large amount of data ($n = 100$), medium amount of data ($n = 50$) and small amount of data ($n = 20$)

$$\widehat{\text{Var}(\eta)}_{n=100} = 1.22 \cdot 10^{-1},$$

$$\widehat{\text{Var}(\eta)}_{n=50} = 2.43 \cdot 10^{-1},$$

$$\widehat{\text{Var}(\eta)}_{n=20} = 6.08 \cdot 10^{-1},$$

respectively. Clearly, the errors from the multi-level model do not increase by the same factor. This shows how using a model-based approach that can borrow informational strength in space becomes an increasingly better option when the amount of available data is small.

The tables show that the difference in predictive accuracy seems to be

relatively substantial between the different amounts of data that were tested. This is also the case when looking at a visualization of the estimated $\hat{\eta}$ in Figure 5.8. Here it is also clear that the estimates are less accurate in areas without any simulated trials. Corresponding plots when using other versions of the multi-level model or different weight combinations are very similar as well. This shows that the model is able to capture a substantial part of the correlation structure, even when there is not much available data, but the accuracy of estimates is heavily affected by the amount of data.

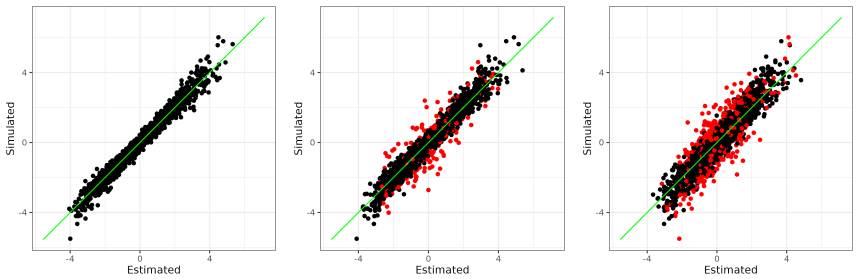


Figure 5.8: Simulated $\boldsymbol{\eta}$ on admin 3 level plotted against corresponding estimated values from the 'ML prior 4' model using different amount of simulated data as input. The simulation is made with the weights $\{w_1, w_2, w_3\} = \{0.4, 0.3, 0.3\}$. The large data amount is used for the left plot, medium amount for the middle plot and small amount for the right plot. Areas where no trials were simulated are marked in red.

Chapter 6

Case study: DHS survey in India

6.1 Purpose

Having seen that the multi-level model is able to produce accurate estimates on simulated data, the model is also applied to data from the DHS survey conducted in India between 2019 and 2021. The goal is first to see if the multi-level model is able to produce useful estimates of the variables of interest on the three administrative levels in India, where we also test the effect of removing covariates from the model. We look at both the significance of the estimated parameters for the covariates, and the difference in estimation error between the multi-level model with and without covariates. Part of the validation process here is to see the accuracy of estimates in areas where no data is available. This is done using the technique presented in Section 4.6. Additionally, while applying the models to the real DHS data we are able to measure the run time needed to fit the models. Hence we can assess whether the computational techniques presented in Section 3.7 are sufficient to make the multi-level model applicable in practice.

The DHS data is also used as a means to compare the multi-level model to the single-level models proposed in Section 4.2.1. They are compared in terms of predictive accuracy to find out whether the multi-level model is the preferred model choice when working with real data, as this can give different results than when only using simulated data. In addition, the models are applied to smaller subsets of the available data. Depending on the resulting performance, it can be argued if it is possible to obtain sufficiently accurate estimates on admin 1 and admin 2 level with less survey data. This can then be used to decide how

resources are best spent during planning of future surveys.

Finally, we also look into the insights that the estimated model parameters can give in the specific case of India, and how similar analysis can be conducted on other data sets. This is because the purpose of using a multi-level model is, besides accurate estimates, to understand the correlation structure between variables in areas across the country .

6.2 Criteria for selecting data used for validation

In Section 4.6 it is explained that the models are evaluated on the real data through cross-validation, where only areas with a sufficient amount of data are used in the validation sets. This is because the 'real' prevalences are not known, so the models are validated by using direct estimates instead. Thus the areas used for validation must have direct estimates that are reliable. In order to properly select these areas, the survey design and responses are taken into consideration.

The DHS survey data is split into responses from rural and urban parts of India. There must be at least 3 clusters from rural parts and 3 clusters from urban parts of an area for a direct estimate to be obtainable. Thus the first part of selecting the areas is to filter out the ones without this amount of clusters. In practice this can be done by computing direct estimates using the `survey` package, as described in Section 2.2, and seeing which estimates have unstable standard deviations. When there is an insufficient amount of clusters, the computed standard deviations either tend to 0 or become very high. Thus we only consider areas where the direct estimates have standard deviations between 0.01 and 1 when creating validation sets.

In addition, part of the survey design is scaling of the observations according to the method described in Section 2.2. This entails that both the number of responses and number of positive responses within each admin 3 area change significantly after scaling. To further ensure reliable direct estimates, another criterion for the selected areas is that they must have a total of more than 20 responses after scaling. As a result, 2000 out of the 2308 admin 3 areas (excluding island territories) are used for validation of the models when applied to data on educational level. For the data on employment, only 919 admin 3 areas qualify to be included in the validation sets.

6.3 Demonstration of multi-level model

The multi-level model with covariates included, specified in Equation (4.5), is first run on the complete sets of data on educational level and employment. In Figure 6.1 the estimates from the model of the prevalence of completed secondary education in India are plotted along with the coefficients of variance for each estimate on the admin 1, admin 2 and admin 3 levels. Similar plots for current employment are shown in Figure 6.2. Here it is of interest to determine whether the model-based method provides more useful estimates than commonly used methods. therefore, the plots are compared to the corresponding plots from using direct estimates, which are presented in Section 2.2.

When compared to the plots of direct estimates, the results show very little difference in the estimated prevalences on admin 1 and admin 2 level. However, the coefficients of variance are not as similar. Especially in the center part of India, where there are many small neighbouring regions, the coefficients of variance are smaller for the model-based estimates than the direct estimates. Conversely, in the north-eastern part of India there are not as many small areas, and this part of India is almost disconnected from the rest. Here, the coefficients of variance are significantly higher for the model-based estimates. This shows a major weakness of the spatial regression methods used in the model, which is that the accuracy of estimates gets worse as the sparsity of the neighbourhood structures increases. Still, the main advantage of the model-based method is that we are able to obtain estimates in all admin 3 areas, which is clear from the figures.

In Table 6.1 the error estimates resulting from cross-validation of the multi-level model on the DHS data are displayed together with corresponding results from the model without covariates. It is immediately clear from the results that the errors are significantly lower on the admin 1 and admin 2 levels than in any scenario from the simulation study, such as in Section 5.3. This happens despite the fact that approximately the same amount of data is used. The difference lies in how data is distributed across the administrative levels during the simulation study compared to how the DHS survey data is distributed.

The DHS survey conducted in India between 2019 and 2021 was designed specifically to obtain accurate estimates on admin 1 and admin 2 level. Respondents from the survey are somewhat evenly split between areas on these levels, whereas the simulated data in Chapter 5 are instead evenly split between areas on admin 3 level. These two ways of distributing data are far from equal, because the amount of admin 3 areas that make up either an admin 1 or admin 2 areas varies a lot. For example, the capital of India, New Delhi, is considered an admin 1 area,

Variable	Covariates	MSE (10^1)			CRPS (10^1)		
		Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Education	Yes	0.08	0.19	2.34	0.39	0.73	2.45
	No	0.06	0.19	2.83	0.37	0.73	2.83
Employment	Yes	0.15	0.52	4.16	0.61	1.40	3.12
	No	0.15	0.56	4.20	0.61	1.41	3.14

Table 6.1: Error estimates of the estimated $\boldsymbol{\eta}$ from the multi-level model on the three administrative levels. The estimates are made based on a selection of direct estimates with low uncertainty for each of the two variables of interest.

admin 2 area and admin 3 area on its own. Considering the part of the simulation study where a 'Large' amount of data was simulated, this means that only 100 out of the 230800 simulated responses belonged to the admin 1 area New Delhi. When having admin 1 areas with only 100 responses, it cannot be expected to get accuracies as low the ones in Table 6.1 on admin 1 level. This is not the case in the survey data, as 6387 females between the ages 20 and 39 responded to the survey in New Delhi. This is a way higher number than in any other admin 3 area, with the second most responses from an admin 3 area being 1545. This shows how the survey is designed specifically to ensure enough responses in each admin 1 area, rather than evenly distributing responses between admin 3 areas.

Shifting focus to the error estimates on admin 3 level, the results are relatively high errors compared to corresponding errors from the simulation study. Despite higher errors, the resulting estimates are still useful. The errors from the estimates regarding educational level correspond to an average error of approximately ± 6.5 percentage points when converted to a probability scale, and similarly ± 8.2 percentage points for the estimates regarding current employment. The CRPS scores also indicate that estimates with reasonable standard deviations are produced for both of the variables. Thus one can argue that the estimates on admin 3 level are useful for the purpose of making data driven decisions for small regional governments, as long as the uncertainty is also taken into consideration.

The estimated distributions of the parameters $\boldsymbol{\beta}$ in the multi-level model in Equation (4.5) can be used to assess the effect of including the covariates presented in Section 2.3. Table 6.2 shows estimates of the precision parameter and covariate parameters, along with the standard deviations of the estimates. Using Equation 5.2 to look at the two components of the variance, it can be found that, according to the estimated parameters, covariates make up approximately 9% of the total variation in the data on education, and approximately 3.7% in the data on employment.

This reflects the results in Table 6.1, where the errors on admin 3 level are higher when covariates are not included, but for employment the difference is negligible. Thus the results indicate that including the right covariates has a positive effect on the accuracy of estimates, but the spatial components of the model are still the most important when working with this data. Note that the intention in this thesis is to investigate if the inclusion of covariates works as an additional component in the multi-level model, *not* to find the ideal covariates for modelling the two variables of interest. However, this is could be done through experimentation with a wider range of covariates and significance tests using the estimated parameter values and standard deviations.

Variable	E(κ)	SD(κ)	E(β_p)	SD(β_p)	E(β_n)	SD(β_n)
Education	1.113	0.131	0.305	0.019	0.183	0.029
Employment	1.886	0.177	-0.120	0.025	0.039	0.042

Table 6.2: Estimated values and standard deviations for the precision parameter κ and parameters for the log transform of the covariates 'population density' and 'nighttime lights per person', denoted as β_p and β_n, respectively.

When applying the multi-level model to the variables of interest, the average run time needed to produce estimates in all admin 3 areas was 12 minutes and 20 seconds for the prevalence of completed secondary education, and 9 minutes and 3 seconds for the prevalence of current employment. This is definitely acceptable times to make estimates, and shows that it would not take long to run the model on similar variables from the available DHS data in a range of countries. Also, India has a complex administrative structure compared to many other countries in the world. Thus running the model on data from other countries is likely to be even faster.

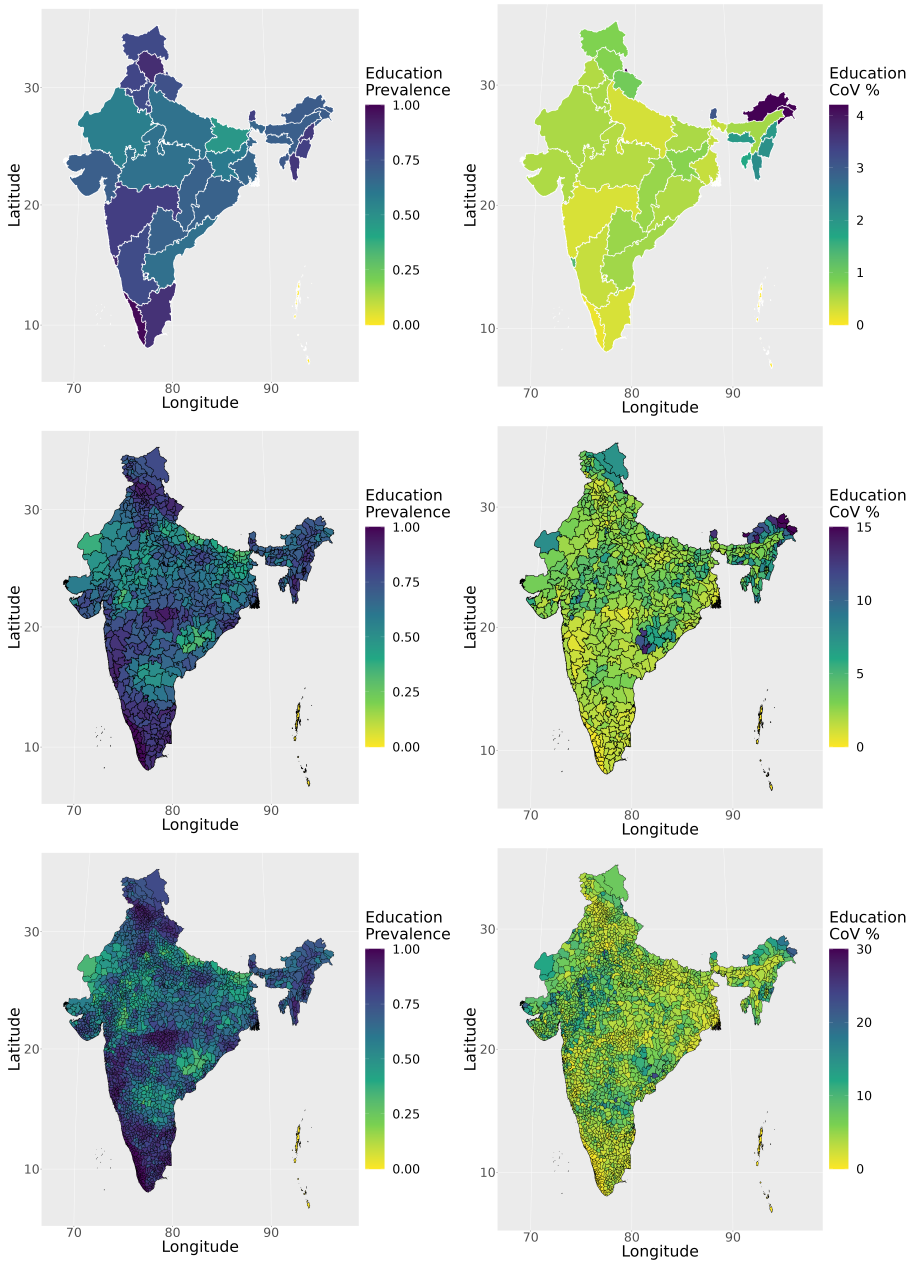


Figure 6.1: Estimates from the multi-level model of prevalences (left side) and the associated coefficients of variance (right side) of completed secondary education in all admin 1 areas (top row), admin 2 areas (middle row) and admin 3 areas (bottom row).

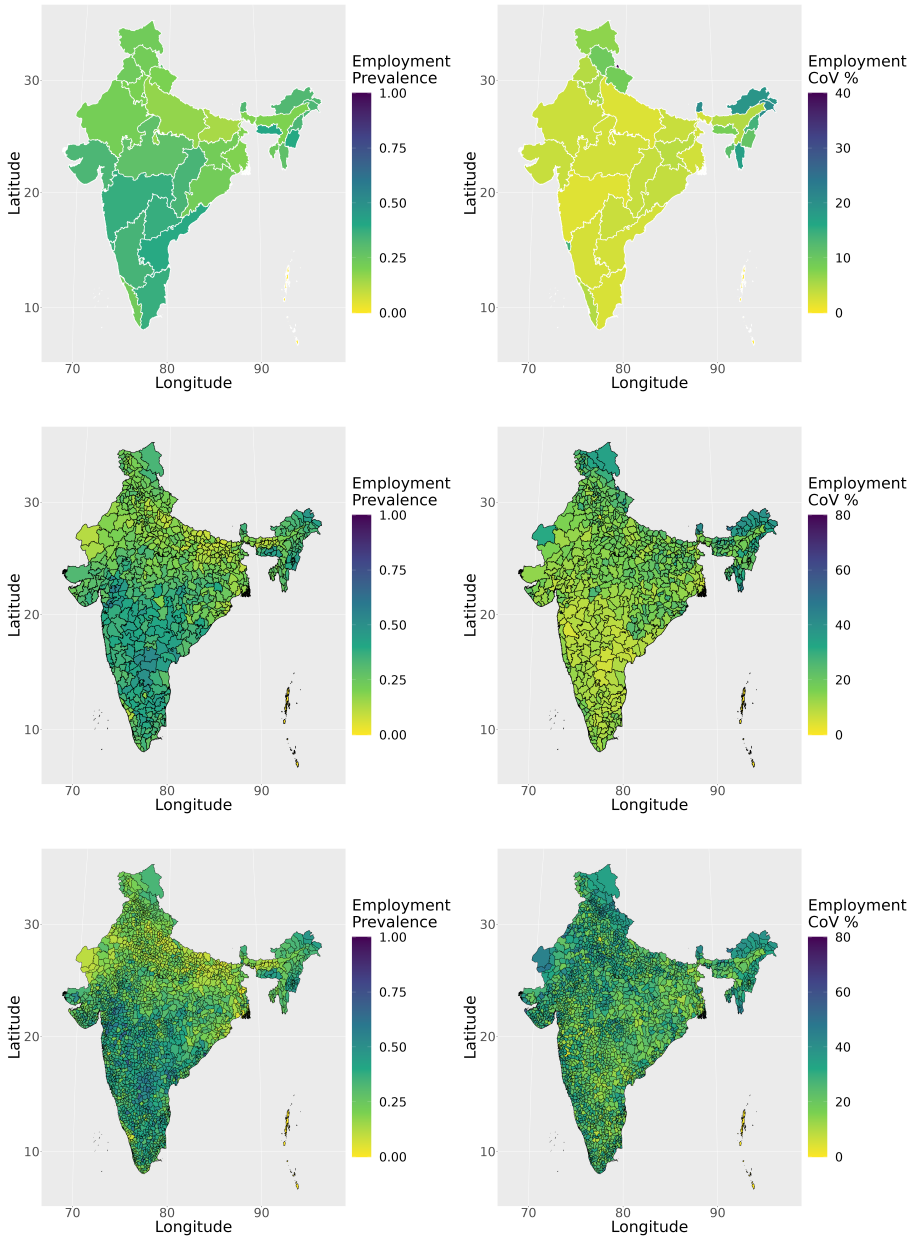


Figure 6.2: Estimates from the multi-level model of prevalences (left side) and the associated coefficients of variance (right side) of current employment in all admin 1 areas (top row), admin 2 areas (middle row) and admin 3 areas (bottom row).

6.4 Comparison to single-level models with reduced data sets

The models are compared through the accuracy of their prevalence estimates on logit scale on each administrative level, similarly to in Chapter 5. Estimates are made by fitting the models to 50% of the available DHS data on education and employment, and producing separate estimates for each administrative area on all three levels. The estimates are compared to direct estimates in the corresponding areas, which are made using the *remaining* 50% of the data. Data used for model-based estimates and direct estimates is divided through randomly selecting half of the clusters from the survey and using data from those clusters in the model-based estimates, while data from the remaining clusters is used for direct estimates. MSE and CRPS is computed using only areas with reliable direct estimates, as described in Section 6.2. To ensure accurate error estimates, this process is repeated 10 times, with the data being randomly split in half anew each iteration. Then the average MSE and CRPS from each of the models are computed.

Results from the multi-level and single-level models are presented in Table 6.3 and Table 6.4, with the tables representing results on the data on education and employment, respectively. The observant reader can notice that the errors from the multi-level model are significantly lower on admin 3 level than they were in Section 6.3. This is due to the change in selection of training and testing data. For the errors in Table 6.1, the models are fitted to training data where all observations from areas is the test set are removed. For the results in the tables below, data is selected based on randomly selected clusters instead of areas, meaning that most areas in the test set still have observations that are used to fit the models.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	0.11	2.49	4.24	0.53	3.15	3.82
Admin 2	0.21	0.40	1.97	0.68	1.05	2.29
Admin 3	0.16	0.38	0.98	0.60	1.03	1.61
Multi-level	0.12	0.36	0.97	0.53	1.02	1.61

Table 6.3: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on the real DHS data on education.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	0.21	2.27	4.29	0.76	2.67	3.54
Admin 2	0.42	1.09	2.90	0.96	1.84	2.73
Admin 3	0.33	1.08	1.52	0.91	1.88	2.11
Multi-level	0.23	1.04	1.57	0.78	1.85	2.13

Table 6.4: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on the real DHS data on employment.

Based on the error estimates, it is clear that none of the single-level models noticeably outperform the multi-level model on any level. However, the accuracy of the single-level models is very similar on the levels they are mainly intended, compared to that of the multi-level model. That is, using the 'Admin 1' model for estimates on the admin 1 level and so on, seem to give just as good estimates as the multi-level model. It is then a question of whether it is preferred to apply simple models for more explainable estimates on each level separately, or to apply a more complex model to ensure consistency between estimates on different levels and allow for more insight into the correlation structure of the data.

Finally, the multi-level model seems to be able to produce accurate estimates on the admin 1 and admin 2 levels, even when only fitted to data from 50% of the clusters in the DHS survey. This means that when only estimates on coarser levels are of interest, especially on the admin 1 level, it is possible to spend less resources on the survey and still get good results. However, this all depends on how accurate the recipients of the estimates need them to be, which has to be taken into consideration when designing the survey.

6.5 Interpretation of estimated model parameters

An advantage of applying the multi-level model to real data sets is that the estimated model parameters can be used to get a better understanding of the structure of the variation of a variable across a country or region. For example, when applied to the data on education, the multi-level model estimated the weights $\mathrm{E}(\{w_1, w_2, w_3\}) = \{0.339, 0.051, 0.610\}$ with standard deviations $\mathrm{SD}(\{w_1, w_2, w_3\}) = \{0.072, 0.026, 0.073\}$. The weights represent the proportional contribution of each administrative level to the total marginal variance of the spatial model components. This can then be interpreted as a strong indication that most of the variation happens on the admin 3 level, while variation on the

admin 1 level also has a significant impact. However, the simulation study showed that the parameter estimates are likely to be somewhat inaccurate, despite the relatively low estimated standard deviations of the weights. This is important to take into consideration before drawing any conclusions. Still the results are reliable enough to create an overall picture of how educational levels varies across different administrative borders.

The parameters κ and $\boldsymbol{\beta}$ can be used to estimate the marginal variance of the variable of interest using Equation (5.2). This is a way to measure how large the differences between different parts of the country currently are, which is useful as many countries strive to reduce differences among the population. The covariate parameters β_p and β_n also provide a simple way to see if the covariates are negatively or positively correlated with a variable, and to what degree. For example, Table 6.2 shows that a high population density is positively correlated with educational level, meaning that investing in the expansion of urban areas can help boost educational level across areas.

Chapter 7

Discussion

We have been able to develop a fast multi-level model for small area estimation that is a good alternative to traditional methods. Both the simulation study and case study on India provide results that clearly show that the multi-level produces estimates that are as good or better than the single-level models when the data has variation on multiple spatial scales. None of the single-level models could consistently compete with the multi-level model in all scenarios, and seem to only be preferable model choices when a large emphasis is placed on using models of lower complexity. However, the accuracy of estimates from the multi-level model varied significantly between different scenarios. It was demonstrated that the errors increase when a large proportion of variation happens on the fine-scale administrative levels, and that the amount of data used to fit the model has a great impact on the accuracy of the estimates.

On the other hand, having small amounts of available data is a big part of the motivation behind developing the spatial model in the first place. Therefore, it is interesting to see that the multi-level model can produce useful estimates with the small amounts of data, depite the reduction in accuracy that comes from scarce training data. In Section 2.2 it is illustrated that direct estimates are insufficient when making estimates on a fine scale when many areas have little to no available data. We also saw in Section 5.6 that as the amount simulated data decreases, the expected error of direct estimates increases *faster* than the observed errors from the multi-level model. Thus using the model-based approach is a good option when there is a need for estimation in areas with small amounts of available data, even if direct estimates are obtainable.

Another aim of the simulation study was to investigate what effect differ-

ent prior distributions of the weight parameters w_1, w_2 and w_3 have on predictive accuracy and parameter estimation. The effect on predictive accuracy turned out to be almost negligible. It seems that the level of bias introduced through the parameter vector $\boldsymbol{\alpha}$ must be unreasonably high for it to make a significant difference on predictions. However, the prior does have a noticeable effect on parameter estimation. The models tended to overestimate the most dominant of the three weight parameters w_1, w_2 and w_3. As a result, the prior $\boldsymbol{\alpha} = [3, 3, 3]$ introduces a bias that can compensate for this effect, making it the prior with the most accurate parameter estimates throughout the simulation study. Thus this seems to generally be a reasonable choice of prior, even when the bias it introduces does not accurately reflect reality.

Results from the simulation study show that the parameters in $\boldsymbol{\beta}$ are very accurately estimated. Thus the inclusion of covariates worked well as an extra component in the multi-level model. This was also the case when applying the model with covariates to the real DHS data from India during the case study. The estimated errors on admin 3 level from the cross-validation increased when removing covariates from the model, especially for the data on education. For employment the results do not show much of a difference, which could be due to the lower amount of data or simply that the covariates are not as relevant for this variable. However, even for education the covariates only made up approximately 9% of variation in the estimates. Thus the inclusion of covariates can help increase accuracy, but the spatial regression components of the model should still be the main focus and takeaway when interpreting the results. This shows an important part of making the model-based estimates, which is to properly analyze the estimated model parameters, in order to understand which components are the most impactful.

Another interesting finding from the simulation study is the interaction between the precision parameter κ and the different weight parameters. For example, the weights between the Besag components and i.i.d. components on admin 1 and admin 2 level, ϕ_1 and ϕ_2 are easily underestimated, meaning that too much weight is assigned to the random components. This is then compensated for through overestimation of κ, which in turn reduces the estimated total variability from the model. Especially on the admin 1 level, accurate estimation of the weight ϕ_1 is difficult as there are only 39 admin 1 areas included when fitting the model. That is not a sufficient amount of areas to get accurate estimates of this weight, which should be carefully carefully considered when trying to interpret the weight itself and the precision parameter. This is likely to be a recurring problem when applying the model in other countries as well, as there are usually few areas on the coarsest administrative level.

From the case study on India it is clear that the multi-level model can produce useful results when applied to a real data. It is able to make use of the survey design to get estimates on the admin 1 and admin 2 level that in the worst cases are only 2 percentage points off the direct estimates. It also produces admin 3 estimates where the estimated prevalences are in general around 6-8% off. This shows much room for improvement, but is accurate enough to be useful insight when considering the lack of alternative ways of making reliable estimates, especially in areas where direct estimates are unfeasible as we saw in Section 2.2. Through the results from the multi-level model, local policymakers in admin 3 areas can get the information they need to make more data-driven decisions, even in areas where little actual data is available.

In order to get a better picture of how the multi-level model performs on DHS data, there are changes that could be made to the simulation study that will make it more representative for the available datasets. For example, the simulated responses can be distributed more based on admin 1 or admin 2 borders, instead of simulating the same amount of responses within each admin 3 area. This will help get more realistic error estimates on the admin 1 and admin 2 level, which is an important part of displaying the applicability of the model. In addition, there are many more parameter combinations that can be experimented with to see the predictive accuracy in more scenarios. That was partly done in Giørtz (2022), where a simulation study was conducted with different values of the intercept μ and precision parameter κ, but this should still be researched even further.

Section 3.7 presents computational techniques that have been a crucial part of running the models fast and being able to apply the models through an extensive simulation study and the case study. The complexity of the multi-level model and administrative structure in India entail large computational challenges, which demonstrates the power of the used techniques. By using the implementations of the Laplace approximation, Automatic Differentiation and sparse matrix operation provided by the TMB library in `R`, we were able to run and fit the models to hundreds of simulations in different scenarios and real DHS data, all within a reasonable time frame. On average, fitting the multi-level model took approximately 5 minutes to perform, with the amount of data included being the main factor that affected the run time. Fitting the model to the real DHS data on education took slightly longer, but still only around 10 minutes, which is considered to be an acceptable result.

On the other hand, an extensive amount of data collection and prepara-

tion must be completed before the multi-level model can be applied. This includes downloading and mapping population and covariate data from WorldPop (2018) to administrative areas on each level, extracting complex shapefiles of the country in question and using this to create neighbourhood matrices, and reading and cleaning massive data files from the DHS surveys. Going through this process takes significantly more than 5 minutes, which makes it a time consuming task to apply the model for the first time to a new country. However, this only has to be done once for each country, meaning that when all the preparations are done, the multi-level model can be used to make estimates of a range of variables of interest across a country.

The multi-level is useful in other scenarios than the case study presented in this thesis, as it is designed to generalize well to a wide range of datasets. For example, the DHS program has performed over 400 surveys in more than 90 countries, which means that there is already a lot of available data similar to the data used in Chapter 6 that is well suited for the new multi-level model. Additionally, the model can be applied to data on other scales, for example by including a national administrative level above the admin 1 level, or a fine-scale grid of 10×10 kilometers below the admin 3 level. In that case the model can be compared to continuously indexed models or other discrete models such as the Leroux model (Leroux et al., 2000).

In addition, results from the model can be used to determine how to better perform surveys in the future. Through analysis of the model estimates, it can be determined in which areas the estimated variables have high and low uncertainties, and then make adjustments to upcoming surveys thereafter. For example, the estimated coefficients of variance in the north-eastern part of India are slightly higher when using the multi-level model compared to the direct estimates, which we can see from the case study. If the intention is to use the model-based approach to make estimates on future DHS survey data, it would be wise to prioritize acquiring more data from this part of India, and in the island territories that the model can not currently include.

The development of a new multi-level model for accurate small area estimates has shown to be successful. The model provides a way to make estimates across administrative levels where there are areas with little to no available data, while producing estimates on different levels that are consistent with each other. The inclusion of covariates in the model worked well and gave improved results when applied to real sets of data, whereas the results of using informative priors for the model parameters did not lead to much difference in predictive accuracy. Along with useful estimates, the model

parameters provide good insight into the correlation structure of the data and increased interpretability of covariate effects. It can also be adjusted to include more or fewer administrative levels, other covariates or even new spatial components. The fact that estimates on multiple different levels all come from the same model is also a desirable trait of the multi-level model, as this increases simplicity compared to when different methods are used for estimates on each level.

To further improve the performance and usefulness of the multi-level model, an effort should be put into researching optimal survey design for surveys that are specifically intended to be used as input data for the model. Thus surveys can be conducted to achieve accurate estimates while spending as little resources as possible. Also, a larger focus can be put on collecting appropriate covariate data, which could be worth investing more resources into to achieve better estimates, rather that having more comprehensive surveys. Future work with multi-level models is bound to result in improved performance during estimation, which can be crucial for making a difference in developing countries.

Bibliography

Besag, Julian (1974). 'Spatial Interaction and the Statistical Analysis of Lattice Systems'. In: *Journal of the Royal Statistical Society: Series B (Methodological)* 36, pp. 192–225.

Besag, Julian, Jeremy York and Annie Mollié (1991). 'Bayesian image restoration with two applications in spatial statistics'. In: *Annals of the Institute of Statistical Mathematics* 43, pp. 1–59.

Dirichlet, J. P. G. (1850). 'Über die Reihen, welche durch die Potenzreihen $\sum(a, b, c, ..., z)^n$ entstehen'. In: *Journal für die reine und angewandte Mathematik* 40, pp. 209–227.

Fisher, RA (1936). 'The Use of Multiple Measurements in Taxonomic Problems'. In: *Annals of Eugenics* 7.2, pp. 179–188.

Fournier, David A. et al. (2012). 'AD Model Builder: using automatic differentiation for statistical inference of highly parameterized complex nonlinear models'. In: *Optimization Methods and Software* 27.2, pp. 233–249. DOI: 10.1080/10556788.2011.597854. eprint: https://doi.org/10.1080/10556788.2011.597854. URL: https://doi.org/10.1080/10556788.2011.597854.

Fuglstad, Geir-Arne, Zehang Richard Li and Jon Wakefield (2021). 'The Two Cultures for Prevalence Mapping: Small Area Estimation and Spatial Statistics'. In: DOI: 10.48550/ARXIV.2110.09576. URL: https://arxiv.org/abs/2110.09576.

GADM – Global Administrative Areas (2023). URL: https://gadm.org/ (visited on 24th Feb. 2023).

Giørtz, Petter Jeppesen (2022). 'Fast Spatial Multilevel Models for Small Area Estimation'. In: Unpublished.

Kristensen, Kasper (2023). *TMB: Template Model Builder*. Comprehensive R Archive Network (CRAN). URL: https://cran.r-project.org/web/packages/TMB/TMB.pdf.

Leroux, Brian G., Xingye Lei and Norman Breslow (2000). 'Estimation of Disease Rates in Small Areas: A new Mixed Model for Spatial Dependence'. In: ed. by M. Elizabeth Halloran and Donald Berry, pp. 179–191. DOI: https://doi.org/10.1007/978-1-4612-1284-3_4.

Lumley, Thomas (2004). 'Analysis of Complex Survey Samples'. In: *Journal of Statistical Software* 9.8, pp. 1–19. DOI: 10.18637/jss.v009.i08. URL: https://www.jstatsoft.org/index.php/jss/article/view/v009i08.

Matheson, James E. and Robert L. Winkler (1976). 'Scoring Rules for Continuous Probability Distributions'. In: *Management Science* 22.10, pp. 1087–1096.

Mercer, Laina et al. (2014). 'A comparison of spatial smoothing methods for small area estimation with sampling weights'. In: *Spatial Statistics* 8, pp. 69–85. ISSN: 2211-6753.

Mercer, Laina D. et al. (2015). 'Space–time smoothing of complex survey data: Small area estimation for child mortality'. In: *The Annals of Applied Statistics* 9.4, pp. 1889–1905. DOI: 10.1214/15-AOAS872. URL: https://doi.org/10.1214/15-AOAS872.

Population Sciences - IIPS/India, International Institute for and ICF (2022). 'India national family health survey NFHS-5 2019-21'. In: URL: https://www.dhsprogram.com/pubs/pdf/FR375/FR375.pdf.

Riebler, Andrea et al. (2016). 'An intuitive Bayesian spatial model for disease mapping that accounts for scaling'. In: *Statistical Methods in Medical Research* 25.4, pp. 1145–1165. DOI: 10.1177/0962280216660421.

Rue, Håvard and Leonhard Held (2005). *Gaussion Markov Random Fields - Theory and Applications*. Chapman and Hall/CRC.

Skaug, Hans J. and David A. Fournier (2006). 'Automatic Approximation of the Marginal Likelihood in Non- Gaussian Hierarchical Models'. In: *Computational Statistics & Data Analysis* 56, pp. 699–709.

Sørbye, Sigrunn Holbek and Håvard Rue (2014). 'Scaling intrinsic Gaussian Markov random field priors in spatial modelling'. In: *Spatial Statistics* 8. Spatial Statistics Miami, pp. 39–51. ISSN: 2211-6753. DOI: https://doi.org/10.1016/j.spasta.2013.06.004. URL: https://www.sciencedirect.com/science/article/pii/S2211675313000407.

United Nations General Assembly (2015). 'Transforming our world: The 2030 agenda for sustainable development'. In: URL: https://sdgs.un.org/sites/default/files/publications/21252030%20Agenda%20for%20Sustainable%20Development%20web.pdf.

Utazi, C. Edson et al. (2021). 'District-level estimation of vaccination coverage: Discrete vs continuous spatial models'. In: *Statistics in Medicine* 40.9, pp. 2197–2211. DOI: https://doi.org/10.1002/sim.8897. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/sim.8897. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/sim.8897.

WorldPop (2018). 'Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076)'. In: DOI: https://dx.doi.org/10.5258/SOTON/WP00646.

Appendix A

Simulation results

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	0.564	2.349	2.535	1.481	3.226	3.676
Admin 2	0.598	0.528	0.892	0.855	1.080	1.855
Admin 3	0.895	0.538	0.569	1.058	1.088	1.276
ML base	0.422	0.356	0.435	0.693	0.933	1.140
ML prior 1	0.404	0.362	0.438	0.671	0.938	1.142
ML prior 2	0.393	0.366	0.444	0.692	0.949	1.152
ML prior 3	0.448	0.367	0.445	0.712	0.947	1.153
ML prior 4	0.389	0.356	0.442	0.674	0.937	1.150

Table A.1: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.8, 0.1, 0.1\}$. A total of 100 observations were simulated in each admin 3 area in this case.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	1.242	8.947	9.344	1.554	7.215	7.487
Admin 2	0.391	0.621	0.938	0.687	1.158	1.891
Admin 3	0.447	0.675	0.725	0.718	1.206	1.413
ML base	0.403	0.619	0.556	0.669	1.117	1.234
ML prior 1	0.382	0.586	0.544	0.656	1.110	1.230
ML prior 2	0.333	0.612	0.554	0.659	1.122	1.236
ML prior 3	0.350	0.561	0.529	0.656	1.096	1.219
ML prior 4	0.336	0.608	0.553	0.663	1.123	1.237

Table A.2: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.1, 0.8, 0.1\}$. A total of 100 observations were simulated in each admin 3 area in this case.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	3.345	8.001	1.910	2.094	6.436	8.415
Admin 2	0.509	0.909	5.734	0.869	1.496	5.250
Admin 3	0.379	0.618	0.808	0.677	1.107	1.478
ML base	0.407	0.664	0.818	0.692	1.117	1.469
ML prior 1	0.380	0.586	0.797	0.681	1.094	1.470
ML prior 2	0.405	0.649	0.819	0.702	1.124	1.482
ML prior 3	0.364	0.559	0.778	0.658	1.079	1.463
ML prior 4	0.400	0.660	0.826	0.704	1.125	1.485

Table A.3: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$. A total of 100 observations were simulated in each admin 3 area in this case.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	0.679	5.439	6.948	1.209	5.531	6.439
Admin 2	0.498	0.673	2.256	0.796	1.242	3.161
Admin 3	0.487	0.522	0.659	0.739	1.080	1.371
ML base	0.370	0.493	0.627	0.666	1.057	1.343
ML prior 1	0.370	0.526	0.648	0.677	1.068	1.355
ML prior 2	0.348	0.493	0.628	0.652	1.052	1.341
ML prior 3	0.457	0.516	0.633	0.711	1.060	1.344
ML prior 4	0.384	0.535	0.653	0.691	1.079	1.361

Table A.4: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.4, 0.3, 0.3\}$. A total of 100 observations were simulated in each admin 3 area in this case.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	1.053	1.966	2.375	1.211	3.159	3.635
Admin 2	1.364	0.959	1.161	1.380	1.497	2.000
Admin 3	1.572	1.056	1.057	1.571	1.575	1.740
ML base	0.606	0.557	0.671	0.896	1.204	1.430
ML prior 1	0.758	0.557	0.662	0.953	1.194	1.417
ML prior 2	0.831	0.573	0.663	0.956	1.199	1.415
ML prior 3	0.775	0.595	0.678	0.975	1.218	1.426
ML prior 4	0.739	0.575	0.668	0.946	1.205	1.420

Table A.5: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.8, 0.1, 0.1\}$. In each simulation a total of 50 observations were simulated in a random selection of 90% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	1.272	9.243	9.648	1.591	7.227	7.531
Admin 2	0.850	1.318	1.288	1.043	1.685	2.076
Admin 3	0.909	1.566	1.600	1.106	1.898	2.079
ML base	0.521	0.931	0.842	0.848	1.464	1.549
ML prior 1	0.536	0.904	0.836	0.863	1.456	1.548
ML prior 2	0.549	0.890	0.821	0.851	1.439	1.534
ML prior 3	0.621	0.952	0.848	0.885	1.470	1.551
ML prior 4	0.558	0.914	0.835	0.856	1.462	1.548

Table A.6: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.1, 0.8, 0.1\}$. In each simulation a total of 50 observations were simulated in a random selection of 90% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	7.002	10.779	13.230	3.827	7.303	8.759
Admin 2	0.938	1.670	6.120	1.191	2.025	5.264
Admin 3	0.786	1.408	2.039	1.045	1.726	2.275
ML base	0.659	1.044	1.393	0.922	1.482	1.976
ML prior 1	0.578	0.985	1.371	0.891	1.469	1.968
ML prior 2	0.615	1.008	1.375	0.896	1.469	1.970
ML prior 3	0.623	1.049	1.367	0.915	1.487	1.963
ML prior 4	0.571	0.967	1.355	0.886	1.456	1.961

Table A.7: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$. In each simulation a total of 50 observations were simulated in a random selection of 90% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	3.018	5.574	7.008	2.554	5.538	6.412
Admin 2	1.017	1.250	2.576	1.192	1.719	3.238
Admin 3	1.047	1.250	1.524	1.196	1.696	2.044
ML base	0.653	0.862	1.055	0.916	1.426	1.762
ML prior 1	0.600	0.842	1.061	0.895	1.420	1.768
ML prior 2	0.755	0.855	1.044	0.939	1.410	1.750
ML prior 3	0.739	0.887	1.074	0.958	1.437	1.773
ML prior 4	0.653	0.872	1.072	0.929	1.425	1.770

Table A.8: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.4, 0.3, 0.3\}$. In each simulation a total of 50 observations were simulated in a random selection of 90% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	1.864	2.156	2.449	1.660	3.122	3.555
Admin 2	2.771	1.897	1.714	2.160	2.178	2.323
Admin 3	3.104	1.982	1.912	2.439	2.262	2.371
ML base	1.158	0.906	1.025	1.287	1.574	1.776
ML prior 1	1.165	0.922	1.028	1.294	1.581	1.777
ML prior 2	1.230	0.945	1.031	1.321	1.586	1.774
ML prior 3	1.373	0.967	1.045	1.357	1.597	1.784
ML prior 4	1.393	0.996	1.060	1.381	1.611	1.793

Table A.9: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.8, 0.1, 0.1\}$. In each simulation a total of 20 observations were simulated in a random selection of 75% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	2.645	9.484	9.599	2.163	7.113	7.368
Admin 2	1.594	2.566	2.075	1.566	2.506	2.507
Admin 3	1.798	3.132	3.048	1.674	2.892	2.971
ML base	1.131	1.824	1.497	1.277	2.125	2.068
ML prior 1	1.077	1.788	1.473	1.256	2.100	2.055
ML prior 2	0.978	1.725	1.441	1.217	2.079	2.040
ML prior 3	0.986	1.763	1.464	1.209	2.095	2.049
ML prior 4	1.051	1.708	1.429	1.233	2.072	2.033

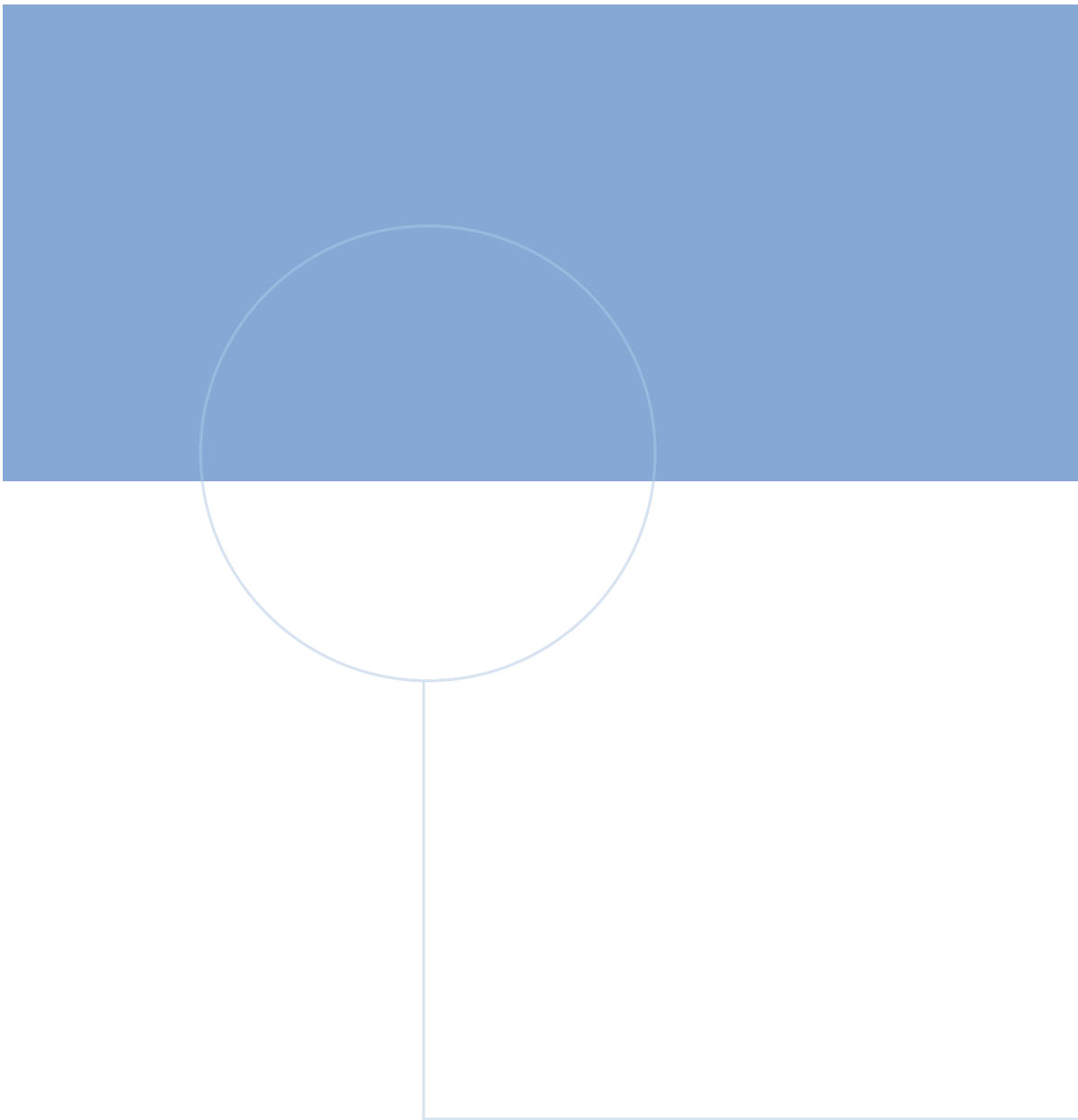
Table A.10: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.1, 0.8, 0.1\}$. In each simulation a total of 20 observations were simulated in a random selection of 75% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	2.315	7.678	11.769	2.111	6.228	8.205
Admin 2	1.920	3.100	6.980	1.754	2.825	5.348
Admin 3	1.689	2.702	4.067	1.624	2.561	3.371
ML base	1.027	1.738	2.520	1.229	2.047	2.742
ML prior 1	1.030	1.721	2.550	1.243	2.046	2.756
ML prior 2	0.982	1.739	2.555	1.238	2.053	2.761
ML prior 3	1.007	1.800	2.595	1.239	2.069	2.770
ML prior 4	1.008	1.652	2.525	1.224	2.019	2.747

Table A.11: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.1, 0.1, 0.8\}$. In each simulation a total of 20 observations were simulated in a random selection of 75% of the admin 3 areas in this case, and 0 observations in the remaining areas.

Model	MSE (10^1)			CRPS (10^1)		
	Admin1	Admin2	Admin 3	Admin 1	Admin2	Admin3
Admin 1	2.197	5.745	7.119	1.941	5.412	6.304
Admin 2	2.168	2.429	3.328	1.819	2.481	3.468
Admin 3	2.158	2.576	2.932	1.877	2.563	2.919
ML base	1.194	1.489	1.833	1.315	1.961	2.359
ML prior 1	1.170	1.502	1.845	1.304	1.966	2.368
ML prior 2	1.129	1.500	1.839	1.293	1.960	2.359
ML prior 3	1.284	1.587	1.878	1.347	1.992	2.375
ML prior 4	1.110	1.560	1.867	1.290	1.987	2.371

Table A.12: Error estimates of the estimated $\boldsymbol{\eta}$ from each of the models on the three administrative levels. The errors in this table were calculated from running the models on 100 separate simulations using the weights $\{w_1, w_2, w_3\} = \{0.4, 0.3, 0.3\}$. In each simulation a total of 20 observations were simulated in a random selection of 75% of the admin 3 areas in this case, and 0 observations in the remaining areas.



NTNU

Norwegian University of
Science and Technology