Bayesian Hierarchical Modelling for Two-Dimensional Blood Pressure Data

MARIA-ZAFEIRIA SPYROPOULOU

School of Mathematics, Statistics and Actuarial Science

A thesis submitted for the degree of
Doctor of Philosophy in Statistics

PhD supervisors:
DR JAMES BENTHAM
PROF JIAN ZHANG

14 July 2022

It's a sunny day...

Unknown

Abstract

Many real-world phenomena are naturally bivariate. This includes blood pressure, which comprises systolic and diastolic levels. Here, we develop a Bayesian hierarchical model that estimates these values and their interactions simultaneously, using sparse data that vary substantially between groups and over time. A key element of the model is a two-dimensional second-order Intrinsic Gaussian Markov Random Field (IGMRF), which captures non-linear trends in the variables and their interactions. The model is fitted using Markov chain Monte Carlo methods, with a block Metropolis-Hastings algorithm providing efficient updates. Performance is demonstrated using simulated and real data. Furthermore, IGMRFs can be used to induce conditional dependence in Bayesian hierarchical models. IGMRFs have both a precision matrix, which defines the neighbourhood structure of the model, and a precision, or scaling, parameter. Previous studies have shown the importance of selecting the prior of this scaling parameter appropriately for different types of IGMRF, as it can have a substantial impact on posterior results. The focus is on the two-dimensional case, where tuning of the parameter's prior is achieved by mapping it to the marginal standard deviation of a two-dimensional IGMRF. We compare the effects of scaling various classes of IGMRF, to the application of blood pressure data.

Acknowledgements

I would like to thank James for all the discussions, his guidance and his patience throughout my PhD. I am glad that I had the chance to work with him. Furthermore, I am grateful and thankful for my family and my friends in Canterbury for standing by my side and making the PhD a joyful experience. Finally, many thanks to my sister and Fenia for their support and the constructive discussions.

Contents

Chapter 1

Introduction and background

In this chapter we provide key information on the statistical methodology to be used. A detailed explanation of the data is provided and a description on the structure of the application area is shown while the motivation for this project are presented.

1.1 Bayesian Inference

Bayesian become increasingly common modelling of complex data for various projects. Bayesian inference introduces randomness/uncertainty and this is quantified by including probability in the description of model parameters. In this way Bayesian models provide flexibility and generality for tackling complex problems, Gelman et al. [2013]. Specifically, this uncertainty is explained through a joint probability density with the assumption that the events are exchangeable meaning that the joint probability density is invariant to the permutations of the indices. In contrast to frequentist inference, parameters and hypotheses are not considered to be fixed and probabilities are assigned. In that way, probabilities of the parameters express a degree of belief which is based on prior knowledge about a particular event, i.e., conclusions drawn from previous studies or personal beliefs. In other words, the prior knowledge is expressed through prior distributions

expressing the intuition for a specific dataset, Gelman et al. [2013], McElreath [2020]. Hence, Bayesian inference combines together the personal belief and the information of the data for the final result/ output of the project. However, there are cases that a prior belief or intuition is not available and still Bayesian inference can be applied. That is the case of a non-informative prior distribution where the output of the project depends solely on the data and hence the result coincides with the classical approach.
In summary, there are two types of priors: the informative and non-informative priors. The non-informative priors are those that provide little or no information at all relative to the data. In the following sections we will talk about the priors that we are using for our modelling. The priors that we are using are informative taking both types of informative and non-informative hyperpriors instead. The non-informative hyperpriors that we are using provide little information about the data and the focus is on giving a very wide variance in the hyperprior's distribution.

For being able to combine together prior distribution and information of the data we use the Bayes' Theorem which is expressed in the following way:

$$P(\theta|x) = \frac{P(x|\theta)P(\theta)}{P(x)} = \frac{P(x|\theta)P(\theta)}{\int_\theta P(x|\theta)P(\theta)d\theta}$$

Let us assume that x are data and θ is the parameter of interest. The information provided by the data, the so-called likelihood is denoted by $P(x|\theta)$ where the prior distribution is expressed by $P(\theta)$. $P(x)$ is a probability depended on x and therefore is interpreted as normalising constant. Furthermore, $P(x)$ can be also written as a combination of the likelihood and the prior, which is the numerator while integrating out the θ parameter. Finally, the result of the Bayes' theorem expresses the posterior distribution, $P(\theta|x)$, the probability of the θ parameter conditional on the data, x. Hence, posterior distribution combines the information of the data, x and the prior information, $P(\theta)$ which can be a personal opinion or based on results of previous studies.

1.2 MCMC Methods

Markov Chain Monte Carlo (MCMC) methods are a broad class of computational algorithms which through repeated random sampling obtain numerical results, Robert et al. [1999]. Monte Carlo estimates the properties of a distribution by generating random samples from the distribution. For example we want to find the expected value of the h function in equation 1.1.

$$E_\pi[h(x)] = \int_x h(x)\pi(x)dx \tag{1.1}$$

Therefore, based on Monte Carlo integration we can randomly generate a sample, $(x_1, \dots, x_m)$ from the π density and approximate (1.1) by the empirical average $\bar{h}_m = \sum_{j=1}^{m} h(x_i)$, Metropolis and Ulam [1949]. Because of the Strong Law of Large numbers $\bar{h}_m$ converges almost surely to $E_\pi[h(x)]$.
The computational simplicity and ease that the Monte Carlo approach introduces makes it widely applicable. In more complicated distributions it is computationally more straightforward to calculate the mean of a large sample of numbers than calculating the mean directly from the distribution's equations. However, there are some limitation in Monte Carlo integration such as the cases that we cannot generate a sample from a distribution or not being able to compute the normalising constant.
Furthermore, we have also Markov chains which are stochastic models. Markov chains consist of a sequence of events and has the property that the probability of each event depends only on the event that occurred just before, Markov [2006]:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i+0) = P(X_{n+1} = j | X_n = i) = p(i,j) \tag{1.2}$$

where P is known as the transition matrix. In more detail, Markov chains were formed from a transition kernel K, a conditional probability density, (1.2), such that $x_{n+1} \sim K(x_n, x_{n+1})$. In addition a Markov chain provide a very strong stability property that is the stationary probability distribution, π and is a solution of the equation:

$$\pi^T = \pi^T P \tag{1.3}$$

where P expresses the transition matrix. In that way π allows for free moves throughout the state space with $x_n \sim \pi$ and $x_{n+1} \sim \pi$ where this property is known as *irreducibility*, i.e. all the states in Markov Chain communicate. Hence, by constructing a Markov chain with a specified stationary distribution one can obtain a sample of the desired distribution by recording states from the chain. A Markov chain can be *recurrent* which means that the average number of visits to an arbitrary state i is infinite, Robert et al. [1999]. Finally, by indicating that a Markov chain is reversible we refer to the symmetry of the transition kernel explained as:

$$\pi(i)p(i,j) = \pi(j)p(j,i), \quad i \neq j \tag{1.4}$$

Given the properties of irreducibility, recurrence and reversibility, we can combine Markov chains with the Monte Carlo approach to comprise Monte Carlo Markov Chain methods, Robert et al. [1999]. The MCMC algorithms are generally used for sampling from multi-dimensional distributions, especially when the number of dimensions is high. MCMC methods that are commonly used are the Metropolis-Hastings algorithm and Gibbs Sampler.

The intuition behind Metropolis-Hastings estimation is that we use a proportional density, $g(x)$ that is known and easier to calculate than the one that we actually want to compute, $f(x)$ due to the difficulties in analysing the latter, Chib and Greenberg [1995]. Specifically, what is difficult to calculate is the normalising constant of $f(x)$, Casella and George [1992], Robert et al. [1999].

1.2.1 Metropolis-Hastings algorithm

Let us consider now that θ is an unknown scalar parameter and y expresses the data. We are interested in computing the posterior distribution $P(\theta|y)$ however, there are cases that an analytic solution may not be available. Another distribution which has the properties to be similar to the posterior distribution is defined which is known as a proposal distribution (or kernel), $q(.)$.

The Metropolis–Hastings algorithm generates a collection of states according to a desired distribution $P(\theta|y)$. To accomplish this, the algorithm uses a Markov process, which asymptotically reaches a unique stationary distribution $\pi(\theta)$ such that $\pi(\theta) = P(\theta|y)$, Hastings [1970], Casella and George [1992]. A Markov process is uniquely defined by its transition probabilities $P(\theta^*|\theta)$, the probability of transitioning from a state θ to a state θ^* where it has a unique stationary distribution $\pi(\theta)$ when there is:

1. existence of stationary distribution. There must exist a stationary distribution $\pi(\theta)$: a sufficient but not necessary condition is detailed balance, which requires that each transition $\theta \to \theta^*$ is reversible; the probability of being in state θ and transitioning to state θ^* must be equal to the probability of being in state θ^* and transitioning to state θ, therefore:

$$\pi(\theta)P(\theta^*|\theta) = \pi(\theta^*)P(\theta|\theta^*).$$

2. uniqueness of stationary distribution. The stationary distribution $\pi(\theta)$ must be unique. This is guaranteed by ergodicity of the Markov process, which requires that every state must (1) be aperiodic, where the system does not return to the same state at fixed intervals; and (2) be positive recurrent, where the expected number of steps for returning to the same state is finite.

Now, the derivation of the algorithm starts with the condition of the detailed balance:

$$P(\theta|y)q(\theta^*|\theta) = P(\theta^*|y)q(\theta|\theta^*) \tag{1.5}$$

$$\frac{P(\theta|y)q(\theta^*|\theta)}{P(\theta^*|y)q(\theta|\theta^*)} = 1 \tag{1.6}$$

Using the condition of the detailed balance we define the condition of accepting or rejecting samples constructing the posterior distribution, Niemi [2019].

$$a = min\left\{1, \frac{\text{Posterior}^{*}\text{Proposal}^{k-1}}{\text{Posterior}^{k-1}\text{Proposal}^{*}}\right\} \tag{1.7}$$

$$a = min\left\{1, \frac{P(\theta^{*}|y)q(\theta^{(k-1)}|\theta^{*})}{P(\theta^{(k-1)}|y)q(\theta^{*}|\theta^{(k-1)})}\right\} \tag{1.8}$$

We compare the acceptance ratio a in (1.7) with the u value, randomly generated value from a uniform distribution, $u \sim U(0,1)$. In (1.7) the k term expresses the current iteration that the model is at. If a is greater than the u value then we accept θ^{*} and $\theta^{k} = \theta^{*}$ if not, then we repeat the previous value, $\theta^{k} = \theta^{(k-1)}$.

In addition, we need to calculate the acceptance rate which is the fraction of proposed samples that is accepted in a window of the last K samples, after burn-in. Therefore, for a number K of iterations we approximately want to have an acceptance rate about 50% for a one-dimensional Gaussian distribution which decreases to 20-30% accepted values for an N-dimensional Gaussian target distribution.The acceptable values will comprise the sample of the $P(\theta|y)$ posterior distribution, Smith and Roberts [1993], Gelman et al. [1996]. If we have an acceptance rate more than 50% it indicates that the chain proceeds slowly as values that are close to each other are proposed due to an existence of a small variance in the proposal distribution. The problem can be fixed by increasing the variance of the proposal distribution. On the contrary, a frequency less than 20% accepted values indicate that the chain proposes values with a large variance from each other and has as a result to reject the most of them. Again, the problem can be fixed by tuning the variance in a smaller value.
The acceptance ratio equal to one is beneficial in the case of the Gibbs Sampler where the target distribution coincides with the proposal distribution and has as a result also that the acceptance rate is 100%.

There are two important subclasses of Metropolis-Hastings algorithm, Niemi [2019]. The first is known as the independent M-H where the proposal distribution does not depend on the previous values $q(\theta^{*}|\theta^{(k-1)}) = q(\theta^{*})$ and the acceptance ratio is defined as:

$$a = min\left\{1, \frac{\frac{P(\theta^*|y)}{q(\theta^*)}}{\frac{P(\theta^{(k-1)}|y)}{q(\theta^{(k-1)})}}\right\} = min\left\{1, \frac{K(\theta^*)}{K(\theta^{(k-1)})}\right\}$$

The second one is known as a random walk M-H where the proposal is a symmetrical distribution, where in the acceptance ratio it cancels out,

$$a = min\left\{1, \frac{P(\theta^*|y)}{P(\theta^{(k-1)}|y)}\right\}$$

Lastly a special case of the Metropolis-Hastings algorithm is the Gibbs Sampler, Smith and Roberts [1993], Gilks et al. [1995]. This method is applicable when the proposal distribution is equal to the posterior distribution. Therefore, every value that is proposed is desirable since it is coming from the posterior distribution and hence, gives as a result an acceptance ratio and hence, acceptance rate equal to one. Finally, for the Gibbs sampler we use the full conditional distributions for each of the parameters.

1.2.2 Gibbs sampler

The Gibbs sampler tackles issues that occur in computing a joint distribution by creating and calculating conditional distributions of the parameters instead, Geman and Geman [1984]. In more detail, the joint distribution is the posterior distribution, $P(\boldsymbol{\theta}|y)$ where $\boldsymbol{\theta}$ is a vector of length d. The Gibbs sampler works through the full conditional distributions which for θ_1 is defined as $P(\theta_1|\theta_2, \theta_3, \ldots, \theta_d, y)$. Then the algorithm for Gibbs Sampler can be expressed as follows, Geman and Geman [1984]:

1. Set arbitrary initial values for $\boldsymbol{\theta}^{(1)} = (\theta_1^{(1)}, \theta_2^{(1)}, \ldots, \theta_d^{(1)})$.

2. For a number of iterations $n = 1 : (N-1)$ repeat:

   (a) draw $\theta_1^{(n+1)} \sim P(\theta_1^{(n)}|\theta_2^{(n)}, \ldots, \theta_d^{(n)}, y)$

   (b) ...

(c) draw $\theta_j^{(n+1)} \sim P(\theta_j^{(n)}|\theta_1^{(n+1)}, \ldots, \theta_d^{(n)}, y)$

(d) ...

(e) draw $\theta_d^{(n+1)} \sim P(\theta_d^{(n)}|\theta_1^{(n+1)}, \ldots, \theta_{d-1}^{(n+1)}, y)$

However, the Gibbs sampler can only be used when these full conditional distributions are available. Gibbs Sampler is a Markov Chain Monte Carlo (MCMC) algorithm for obtaining a sequence of observations which are approximated from a specified multivariate distribution when direct sampling is difficult. The sequence of observations can be used to approximate the marginal distribution of the parameter under study which is the parameter that the conditional or posterior distribution is referring to. Another way of extracting the marginal distribution of the parameter under study would be to integrate over the multivariate/joint distribution. However, using the conditional distributions is simpler to sample than integrating over the joint distribution.
Later on, in the analysis we also make use of block Metropolis-Hastings. Some of the full conditional distributions of model's parameters are known therefore Gibbs sampler can be easily implemented. However, due to high correlation that is observed between parameters an alternative must be used. In order to tackle this issue we apply the block-Metropolis-Hastings which updates the correlated parameters together. This method is useful and produces less correlation as it delays the update of the one parameter which then used for the other, Green and Mira [2001]. Examples of the method is shown in Chapter 3 where a more detailed description is given.

Nowadays, there is a plethora of methods to be assigned such as BUGS, JAGS, Plummer et al. [2003], STAN focusing on Hamiltonian Monte Carlo integration, Carpenter et al. [2017]. These methods are platform-specific executable and they were first created as a way for non-statisticians to use statistics. STAN specifically, has been recently created providing a bigger variety of models and analysis than its predecessors BUGS and JAGS. Specifically, you can define the prior distributions and all the distributions of the full conditionals of all the parameters and have a result of it without needing to write more details. However, due to the complexity of our model and previous success in fitting similar (albeit less complicated) models we implemented the MCMC methodology calculating the exact full conditionals using Gibbs and Metropolis-Hastings. Stan can support our model but it will be more complicated to define all the parameters. Furthermore, there is also R-INLA, Lindgren and Rue [2015] which stands for Integrated Nested Laplace Approximation using Laplace approximation and applying Intrinsic Gaussian Markov Random Fields (IGMRFs) analysed thoroughly in chapter 2. R-INLA as its name indicates does not use MCMC thus not the exact distributions of the parameters are founded but approximations. Since we were able to work with MCMC quite well, we did not resort to it. Of course the interested reader is encouraged to implement each one of the methods mentioned. Lastly, there is a possibility to use non-Bayesian modelling such as mixed-effects models which have similar properties to Bayesian hierarchical models. In more detail mixed-effects model include both fixed and random effects in the model. In that way, there are parameters that can be considered constant or fixed and others that follow a specific distribution and therefore are random, Pinheiro and Bates [2000]. However, the scope of this project was to implement Bayesian methods and take advantage of all the flexibility that Bayesian methods provide and specifically the hierarchical models with the missing data.

1.3 Bayesian Hierarchical Modelling

In the past decade, Bayesian hierarchical modelling has become an established technique in global health research, Bearak et al. [2018]. In essence, Bayesian hierarchical models capture the similarity that exists between parameters. Based on that, these parameters are drawn from the same prior distribution. The structure and the connections that hierarchical models create make these models appropriate for complex problems without necessarily using many parameters. In contrast, non-hierarchical models will need a large number of parameters which increases the risk of overfitting, Gelman [2006]. In Figure 1.1 we can see an example of

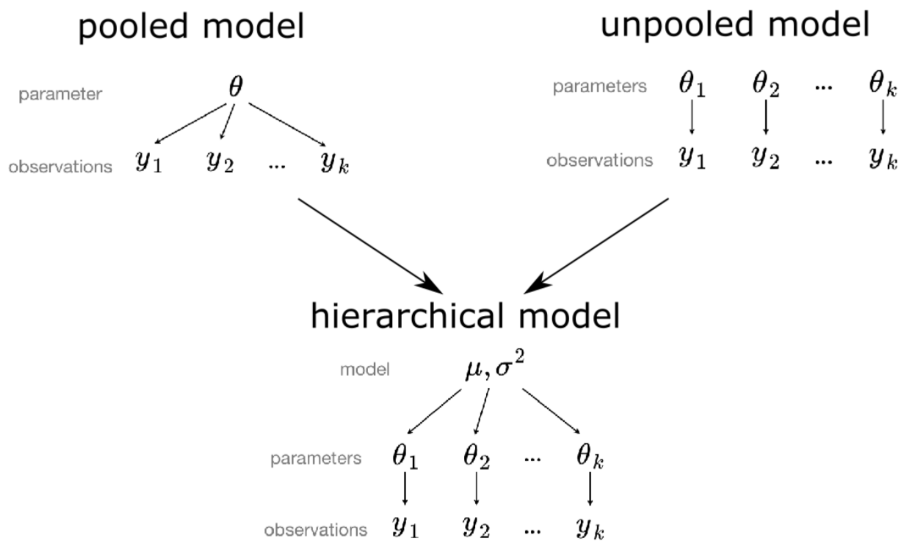

FIGURE 1.1: Comparison between pooled, unpooled and hierarchical model, Widdow Quinn [2006], Florian Wilhelm [2020]

non-hierarchical models which are known as pooled and unpooled models. Pooled and unpooled models are quite opposite models as the first considers a common parameter θ for all the observations whereas the latter considers a parameter θ_i for each observation y_i. In more detail, the prior for the pooled model is expressed through the parameter θ, and is $P(\theta)$ for observations $y_1, y_2, \ldots, y_k$. On the other hand, the unpooled model provides priors through parameters $\theta_1, \theta_2, \ldots, \theta_k$ for each of the k observations. This particular example is used to show that the hierarchical model uses one more level, which are the hyperparameters under the assumption that all θ's of the unpooled case are drawn from the same distribution and hence have the same hyperparameters.

Then hierarchical model is defined as the compromise between pooled and unpooled models as it considers a parameter θ_i for each observation y_i but taking the assumption that the θ_i parameters are exchangeable all these are drawn from the same hyperprior distribution with μ and σ^2 as the hyperparameters, Widdow Quinn [2006], Florian Wilhelm [2020]. In other words, hierarchical model is similar to the unpooled model considering one more level with hyperparameters. In that way the $\theta_1, \theta_2, \ldots, \theta_k$ will be extracted from the same distribution with common parameters, μ and σ^2. Finally, Bayesian hierarchical modelling is an asset to our particular problem as by nature the blood pressure data are hierarchical. More details about Bayesian hierarchical modelling and how it relates to our problem are presented in Section 3.1.

1.4 Non-Communicable Disease Modelling

1.4.1 Introduction

Blood pressure (BP) is the pressure of circulating blood against the walls of blood vessels, Chobanian et al. [2003]. Most of this pressure is an outcome of the heart pumping blood through the circulatory system. As a simple explanation blood pressure refers to the pressure in the large arteries. Blood pressure is usually expressed in terms of the systolic blood pressure (SBP) which is the maximum pressure during one heartbeat over diastolic blood pressure (DBP) which is the minimum pressure between two heartbeats in the cardiac cycle. Therefore, blood pressure is well summarised as a bivariate measurement consisting of diastolic and systolic blood pressure. Overall, blood pressure is one of the vital signs that evaluate a patient's health, World Health Organization [2021].

Noncommunicable diseases (NCDs), also known as chronic diseases, are of long duration which result from a combination of genetic, physiological, environmental and behavioural factors, World Health Organization [2021]. The main types of NCD are cardiovascular diseases such as heart attacks and stroke, cancers, chronic respiratory diseases, e.g., chronic obstructive pulmonary disease, asthma and similar conditions and diabetes.

On the one hand modifiable **behavioural risk factors** of NCDs such as tobacco use, physical inactivity, unhealthy diet, Bentham et al. [2020], Di Cesare et al. [2019], Vandevijvere et al. [2019] and the excess use of alcohol may result in people having raised blood pressure, Bentham et al. [2020]. On the other hand, there are also **metabolic risk factors** which contribute to four key metabolic changes that increase the risk of NCDs, Bentham et al. [2020], Di Cesare et al. [2019], Vandevijvere et al. [2019]:

- raised blood pressure
- overweight/obesity
- hyperglycemia (high blood glucose levels) and
- hyperlipidemia (high levels of fat in the blood)

Hence, as a result all these increase the risk of developing an NCD. Besides behavioural and metabolic risk factors these diseases are driven by forces that include rapid unplanned urbanization, globalization of unhealthy lifestyles and population ageing, ncd [2018]. However, in terms of attributable deaths, the leading metabolic risk factor globally is elevated blood pressure (to which 19% of global deaths are attributed), followed by overweight and obesity and raised blood glucose, Risk Factor Collaboration [2016], World Health Organization [2021]. Raised blood pressure (RBP) is defined as SBP of 140 mmHg or higher, or DBP of 90 mmHg or higher, and is estimated to affect more than one billion people worldwide (Risk Factor Collaboration [2017]).
Although NCDs are often associated with older age groups, evidence shows that more than 15 million of all deaths attributed to NCDs occur between the ages of 30 and 69 years, NCD Countdown 2030 [2018]. Children, adults and the elderly are all vulnerable to the risk factors contributing to NCDs as behavioural and metabolic risk factors take their toll in any age group. Furthermore, the access to the right medication is not available in all the countries, Fletcher et al. [2015] throughout the world which has as a result deterioration in the conditions of NCD.

1.4.2 Previous estimation on raised blood pressure

RBP is a key risk factor for non-communicable diseases such as cardiovascular conditions, cancers and diabetes, which are responsible for approximately 70% of global deaths each year (World Health Organization [2018]). In the early years of NCD risk factor monitoring publications focused on analysing trends in mean body-mass index, systolic blood pressure and cholesterol, Finucane et al. [2011], Danaei et al. [2011]. Furthermore, other studies focused on modelling associations between metabolic risk factors and national income, urbanization and Western diet, Danaei et al. [2013]. Later, a further study showed that the burden of cancer can be attributed to diabetes and high body mass index worldwide, Pearson-Stuttard et al. [2018]. Recent studies have shown that rising rural body-mass index is the main driver of the global obesity epidemic in adults, Risk Factor Collaboration [2019] while other research is concentrated on multidimensional characterization of global food supply from 1961 to 2013, Bentham et al. [2020]. Although recent and comprehensive estimates of trends in the mean values of SBP, DBP and RBP at national level are available (Risk Factor Collaboration [2017, 2016]), there is limited understanding of how the interaction between SBP and DBP varies over time, between countries, and by age and sex. While influential (World Obesity Federation [2019], Development Initiatives [2020]), these models only allow estimation of single variables at a time, e.g., solely on DBP or SBP. Given that disease risk factors have complex interactions that vary over time and between countries, the existing models therefore may fail to capture important information.

We use the term bivariate blood pressure as BP is widely measured using two variables, diastolic and systolic blood pressure. The bivariate distribution is of interest because both systolic and diastolic BP levels have particular medical implications, and it is not clear how medication and diet may affect these relationships, for example. Here, we extend existing methodology (Danaei et al. [2011], Finucane et al. [2014]) to the two-dimensional case, specifying a Bayesian hierarchical model that allows SBP, DBP and their interaction to be estimated simultaneously. A key development is to extend the random walk (Rue and Held [2005]) used previously Danaei et al. [2011], Finucane et al. [2014], which explained non-linear trends in individual variables, to two dimensions. To do so, we use a two-dimensional random walk or Intrinsic Gaussian Markov Random Field (IGMRF) (Rue and Held [2005], Yue and Speckman [2010], Thon et al. [2012]) for the precision matrix of a prior distribution explaining the non-linear time trends. In particular, an IGMRF has been used to make estimates at national level of mean values of risk factors including elevated blood pressure, excess body-mass index, diabetes, and sub-optimal lipid Risk Factor Collaboration [2016] profiles (Risk Factor Collaboration [2017], NCD Risk Factor Collaboration [2017], Risk Factor Collaboration [2016], NCD Risk Factor Collaboration [2020]). The models are sufficiently robust for the World Health Organization to report resulting estimates in publications such as the Global Diabetes Report (World Health Organization [2016]).

1.5 Blood Pressure Modelling

Our focus is to make estimates of blood pressure in order to observe whether the blood pressure has increased or not through the time for each country, for each age group and for each gender. The trends that are presented for each country are complex and vary over time. For example in Figure 1.2 we can see the global trends for single blood pressure variables have changed between the two genders in 2015, Risk Factor Collaboration [2016]. Moreover, in Figure 1.3 we can also make a comparison between the DBP, SBP and the RBP between the different genders and we can notice that differences occur. These data were based on studies that each country had conducted in the past and these studies can refer to national, subnational or communities' population in each country, Danaei et al. [2011], Finucane et al. [2014].

Data can be separated into individual and aggregated data. Individual data are those that selected from a specific individual. The aggregated data are coming from an average of individuals under some same characteristics. For example, in the Diastolic and Systolic Blood Pressure measurements data, we have a selection of data from a number of individuals for each ages between 18 to 80 years old. We separate the ages into 8 groups which are: 18-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79 and over 80 and for each group we are taking the summary statistics. In the whole analysis, we are making use of the aggregated data and therefore, for a specific age group we will have a mean and variance value of the Diastolic and Systolic Blood pressure respectively as well as the number of participants

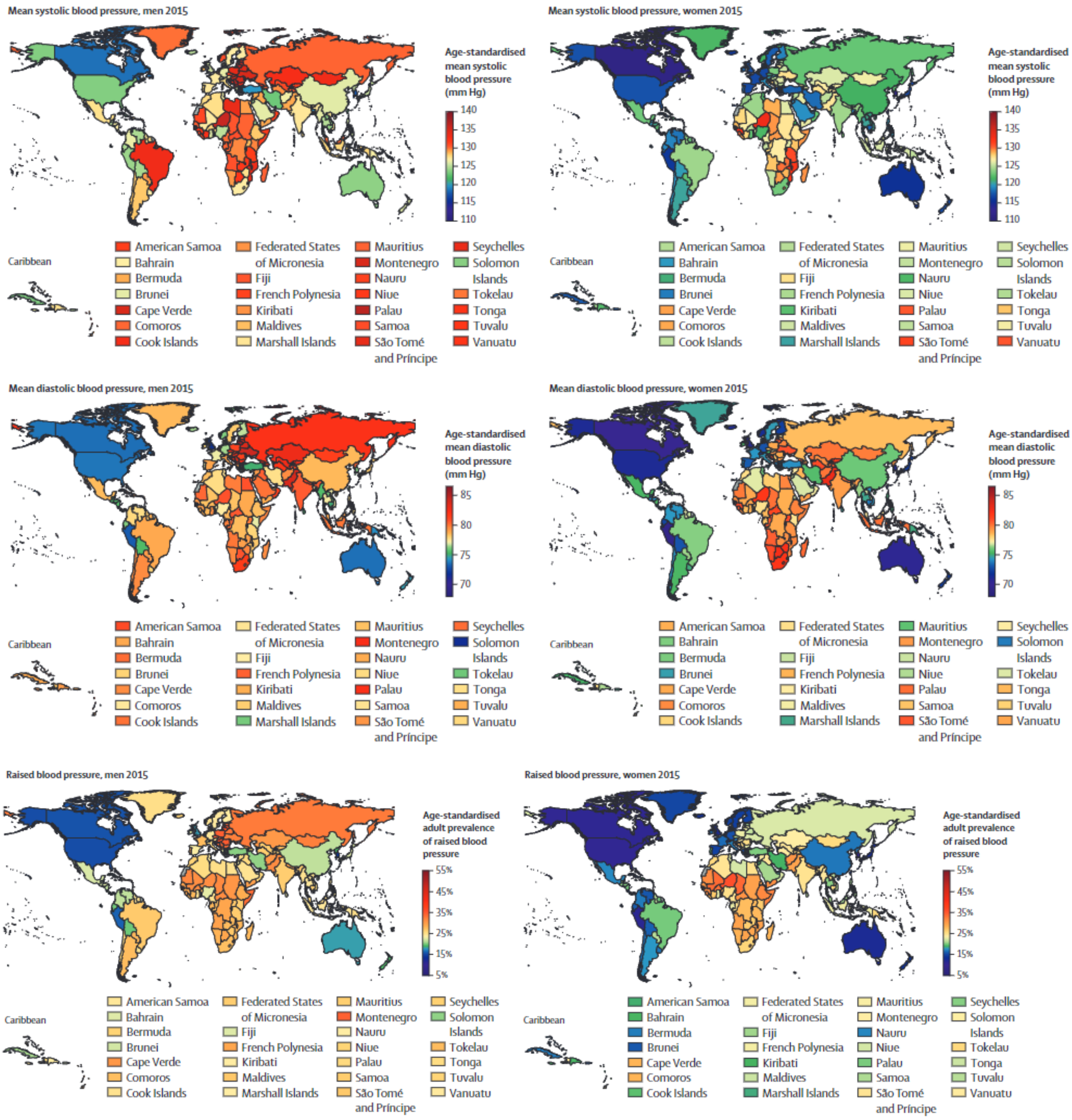

FIGURE 1.2: Age-standardised mean SBP, mean DBP, and prevalence of RBP by sex and country in 2015 in people aged 18 years and older, Risk Factor Collaboration [2017]

in each age group. Aggregated data are very much convenient to be used than the individual data as they speed up the calculations however, there are some drawbacks of using them. Firstly, we are based on the mean values of each age group as we take the mean value of all the individuals comprising the specific age group. Therefore, aggregation of data ignores individual variation as if it were a statistical noise or measurement error. In addition, differences can also

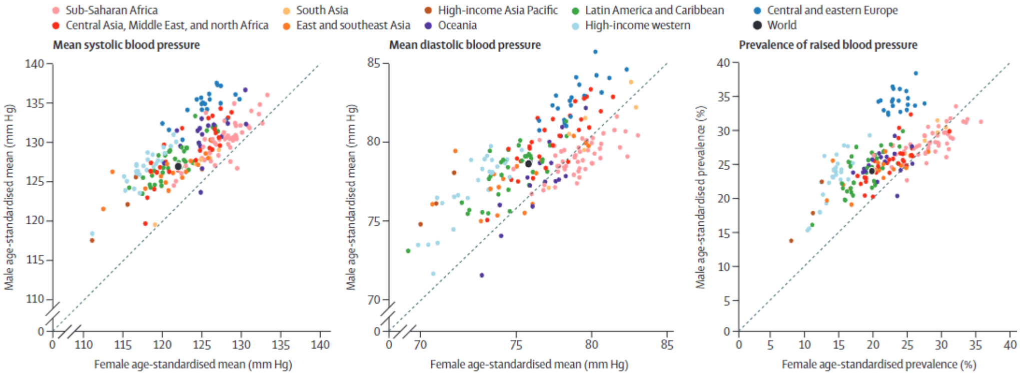

FIGURE 1.3: Comparison of age-standardised mean systolic blood pressure, mean diastolic blood pressure, and prevalence of raised blood pressure in men and women aged 18 years and older in 2015, Risk Factor Collaboration [2017]

exist between the inference that is used and the results between individual and aggregated data. In Figure 1.4 time series of DBP and SBP are presented including all the ages per year over time. What we can observe is that both DBP and SBP seem overall to get decreasing throughtout time.

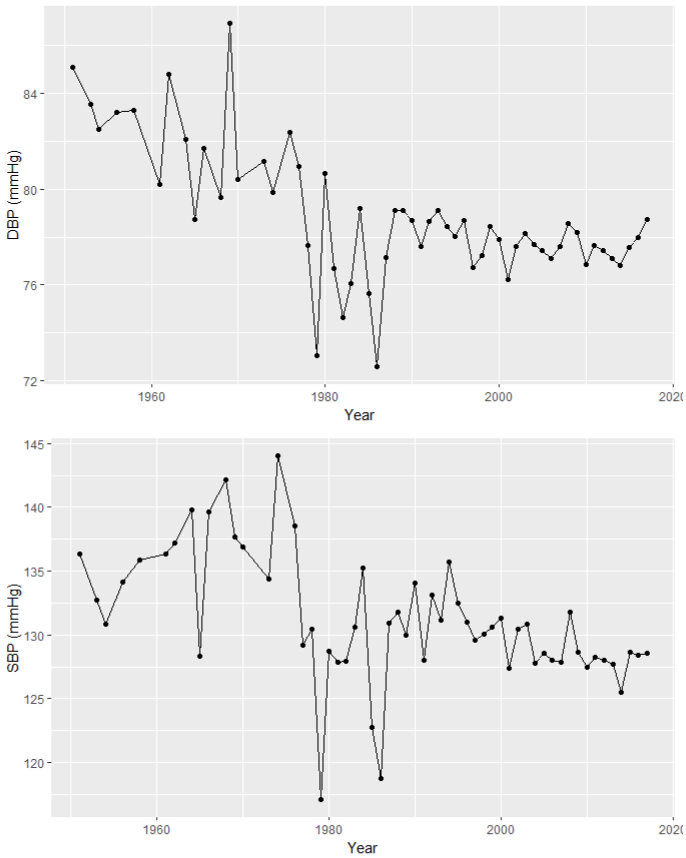

FIGURE 1.4: Time series of DBP and SBP for 1951-2017.

Although there is a large volume of data, not all the countries under study have data for each year for all the age groups and for both genders. Therefore, by using Bayesian hierarchical methodology we can overcome this obstacle, Gelman [2006]. The way of solving this problem is to provide a specific structure in the dataset and in the model. Specifically, we can combine the countries into categories. These categories will include countries that are considered similar e.g., due to similar lifestyle or culture. Then, these categories can also be categorized into larger groups and so on. Therefore, the first type of categories will separate the globe into nine groups which are known as *super-regions* and these are:

- Central and Eastern Europe
- Central Asia, Middle East and North Africa
- East and South East Asia
- High-Income Asia Pacific
- High-Income Western countries
- Latin America and Caribbean
- Oceania
- South Asia
- Sub-Saharan Africa

These super-regions can be divided into smaller groups the so-called *regions*. There are 21 regions in total which contain a smaller number of countries expressing more detail as we pay more attention to the similarities between the countries. We borrow information between the countries in the same group so we want to create groups with similar/alike countries. The groupings used in this work are the same as those used by the Non-Communicable Disease Risk Factor Collaboration, which were based on expert knowledge and have been used for around a decade. The groups were defined a priori consulting on previous studies and remain the same throughout the research, Danaei et al. [2011], Finucane et al. [2011]. The grouping is important to set similar countries in the same group as the estimation of the blood pressure is depending on the group's form. The groupings used in this work are the same as those used by the Non-Communicable Disease Risk Factor Collaboration, which were based on expert knowledge and have been used for around a decade. In Figure 1.5 a global map is shown divided into regions and super-regions and we can see clearly the separation between the countries. The shades of the same colour show the regions that comprise one super-region. For example, the majority of the African continent is depicted with shades of pink colour. The four different shades of pink declare the four different regions: 1) Central Africa, 2) East Africa, 3) West Africa, 4) Southern Africa and all these four regions together comprise the Sub-Saharan Africa super-region. Lastly, it is important to highlight that each country is also considered as a category on its own since there are differences within each country throughout the years that need to be considered.

Summing up, we have described a nested structure that the hierarchical model provides, Gelman [2006]. The entire globe can be divided into super-region groups, region groups and countries groups. By adding countries together we can overcome the problem of missing information. If one country does not have data for a specific year in a specific age group we expect a country that belongs to the same region or super-region and has data to shed light on the missing information since these two countries are expected to behave similarly. In conclusion, the hierarchical model gives more accurate predictions than the no-pooling, in which each country behaves differently and complete-pooling regressions, in which all the countries are in the same group, especially when predicting group averages, Gelman [2006].

It also needs to be stressed that in terms of borrowing information from neighbouring countries both DBP and SBP variables behave the same since if the DBP measurement is missing then the SBP measurement is missing as well.
Bayesian hierarchical models borrow strength or shrinkage through the countries in the same region and super-region but also IGMRFs express shrinkage within each country, region or super-region over time. Each country has potentially data

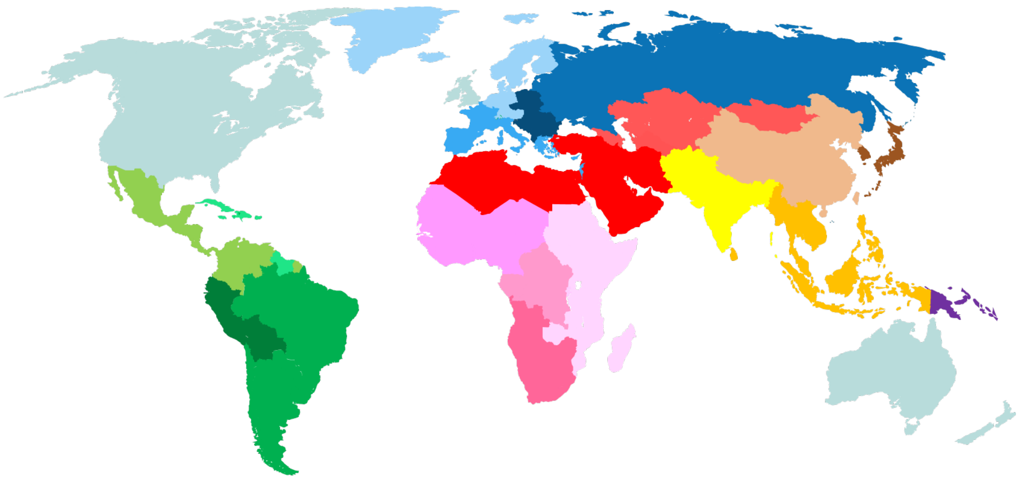

Figure 1.5: World's map divided into regions and super-regions

for each year and by the same token each region and super-region is described by data annually.

Therefore, the structure of IGMRFs provides the chance to share information between the years in this case, since it is possible to have missing data there too. Not only are IGMRFs borrowing strength over time but they also take into account the distance between the years. In other words, years closer to the year with no information will have a bigger weight than those far away.

1.5.1 Challenges

As we have mentioned before there is a large amount of data since we are including 200 countries for more than 20 years. However, there is still a problem of lack of data since not all the years, age groups or both genders have the equivalent dataset. Although we have interest to explore blood pressure measurements for 200 countries only the 165 of them provide us with data. In more detail, the 38% of those countries have data only of one year, 78% have data for less that five years whereas 90.3% have data for less than 10 years. Finally, 7.2% of the countries have data from 10 to 20 years while only 1.21% of the countries which corresponds to two countries have data between 20 and 30 years. Therefore, Bayesian hierarchical modelling is highly useful by allowing us to borrow information between countries in the same group. Although, the problem of lacking data was solved another issue has arisen. The Bayesian hierarchical model will give estimates for each possible country, age group and year for both genders and thus, makes the model computationally expensive too. Furthermore, our model presents a different prior which introduces both DBP and SBP and takes also into account the interaction of these two. In that way, the structure of the prior distribution considers more years of blood pressure and therefore more calculations are needed which means more time is needed. Finally, MCMC techniques are used, which are widely known that are computationally intensive, Gelman et al. [1996], Hastings [1970], Dellaportas and Roberts [2003]. By applying adaptive techniques in MCMC and using techniques used in biostatistics, Roberts and Rosenthal [2009], Gelman et al. [1996] we have managed to develop more efficient code Gelman and Rubin [1996].

Finally, this research could be considered impossible to apply some years ago due to lack of computational power. Nowadays computational advances have made these types of computations feasible. By using high performance computing and parallel computing we manage to minimize the time needed to produce our results. In addition techniques such as Cholesky decomposition, Higham [1990] and canonical parametrization reduced the time was needed for the calculation between matrices significantly while also techniques which specify the sparsity of the matrix proved to be highly beneficial.

1.6 Summary

This chapter provides a description of Bayesian inference. We briefly discuss the differences between frequentist and Bayesian approaches. Furthermore, the reason behind the use of Bayesian hierarchical model is explained emphasising the useful properties of Bayesian hierarchical models for the model's structure. Next, the intuition behind MCMC methods is presented with a concise definition of Monte Carlo and Markov chains as well as Metropolis-Hastings and Gibbs sampler algorithms. Following, non-communicable diseases are explained and the literature behind their modelling is discussed. Finally, the project is explained describing further the structure and the modelling of blood pressure data while highlights the motivations and the challenges that have occurred throughout this research.

1.7 Preview

The remaining chapters of the thesis have the following content:

- Chapter 2 presents the theory underlying Gaussian Markov Random Fields, the different cases that exist and also all the theory that is needed for the implementation of our Bayesian hierarchical model.

- Chapter 3 gives a detailed description of the hierarchical models and describes all the sub-models and the parameters that comprise the final model comparing them both in one and two dimensions.

- Chapter 4 presents the distribution of the sub-models and the methodology that was used to fit each of the sub-models. In addition, applications to simulated and real data are also demonstrated.

- Chapter 5 describes the scaling between the priors that needs to be made. Comparison between the different two-dimensional IGMRFs as well as between the one- and two-dimensional IGMRFs have been made. Lastly, application to the two-dimensional IGMRFs is shown.

- Finally, Chapter 6 gives a discussion of this research, conclusions are presented for the outcome of this project and future work is described.

Chapter 2

Gaussian Markov Random Fields

This chapter provides a description of various fields that are relevant to this thesis. Gaussian Markov Random Fields (GMRFs) and a special case of them, Intrinsic Gaussian Markov Random Fields (IGMRFs) are reviewed. GMRFs are often used in prior distributions to provide structure to the inverse of the covariance matrix, the precision matrix. Both GMRFs and IGMRFs are used to model non-linear relations through their precision matrix.

2.1 Definition

A GMRF is a random vector that follows a multivariate normal distribution. The particular feature of GMRFs is the conditional independence that they introduce [Rue and Held, 2005, chapter 1]. We can explain the conditional independence term by giving an example as stated by Rue and Held [Rue and Held, 2005, chapter 1]. Let us say we have a random vector $\mathbf{x} = (x_1, x_2, x_3)^T$. Here x_1 is conditionally independent from x_2 given x_3. Under this condition the joint distribution of $\mathbf{x}$ can be written as:

$$P(\mathbf{x}) = P(x_1|x_2, x_3)P(x_2|x_3)P(x_3)$$

and since x_1 is conditionally independent from x_2 it can be written as:

$$P(\mathbf{x}) = P(x_1|x_3)P(x_2|x_3)P(x_3)$$

What is beneficial with the GMRF is that the conditional independence can be expressed through the precision matrix, $\mathbf{Q}$. In more detail, when there is a dependence, non-zero values appear in the precision matrix while zero values declare an independent condition. It is useful to represent $\mathbf{Q}$ with an undirected graph.

Definition 2.1 (Gaussian Markov Random Fields)

A random vector $\mathbf{x} = (x_1, ..., x_n)^T \in \mathcal{R}^n$ is called a GMRF linked to a labelled graph $G = (V, E)$ with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q} > 0$ iff its density has the form, [Rue and Held, 2005, chapter 2]:

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

and

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E, \quad \forall i \neq j \tag{2.2}$$

A GMRF can be any Gaussian distribution with a symmetric positive definite (SPD) covariance matrix and vice versa. The properties of a GMRF are described by its precision matrix $\mathbf{Q}$, since its structure indicates connections between nodes: as shown in (2.2), non-zero values in $\mathbf{Q}$ correspond to an edge in G. When $\mathbf{Q}$ is dense, the graph is fully connected.

A connection between GMRFs and Gaussian Process (GP) can exist when we consider a finite collection of random variables for a GP. Hence, a GMRF can be considered as a discretised and sparse version of GP, Jadaliha et al. [2018], when there are zero values in the precision matrix to account for sparsity.

2.2 Intrinsic Gaussian Markov Random Fields

IGMRFs are a special case of GMRFs that we saw in Section 2.1. It is considered that the IGMRFs are GMRFs with linear constraints, [Rue and Held, 2005, chapter 3]. IGMRFs are improper GMRFs, as they have a precision matrix that is not of full rank, and since the precision matrix is not of full rank, its inverse does not exist [Rue and Held, 2005, chapter 3]. This implies that IGMRFs do not have well-defined mean or covariance matrices. However, they do have the property that the mean of an IGMRF of order T is defined up to the addition of a polynomial of order $T-1$.

Definition 2.2 (Improper Gaussian Markov Random Fields)
Let $\mathbf{Q}$ be an $n \times n$ symmetric positive semidefinite (SPSD) matrix with rank $n-k > 0$. Then $\mathbf{x} = (x_1, x_2, ..., x_n)^T$ is an improper GMRF of rank $n-k$ with parameters ($\boldsymbol{\mu}$,$\mathbf{Q}$), if its density is

$$\pi(\mathbf{x}) = (2\pi)^{\frac{-(n-k)}{2}}(|\mathbf{Q}|^*)^{1/2} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T\mathbf{Q}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{2.3}$$

Again, $\mathbf{x}$ is an improper GMRF which is represented by a graph $G = (V, E)$, where

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E, \quad \forall i \neq j \tag{2.4}$$

It needs to be stressed that the parameters ($\boldsymbol{\mu}$,$\mathbf{Q}$) do not represent mean and precision as these terms do not formally exist. However, in the following chapters we will refer to them as mean and precision for the sake of convenience [Rue and Held, 2005, chapter 2].

2.3 Circulant Matrices

There is a large variety of matrices that can be used as precision matrices, $\mathbf{Q}$ in IGMRFs [Rue and Held, 2005, chapter 2]. Therefore, it is interesting to observe

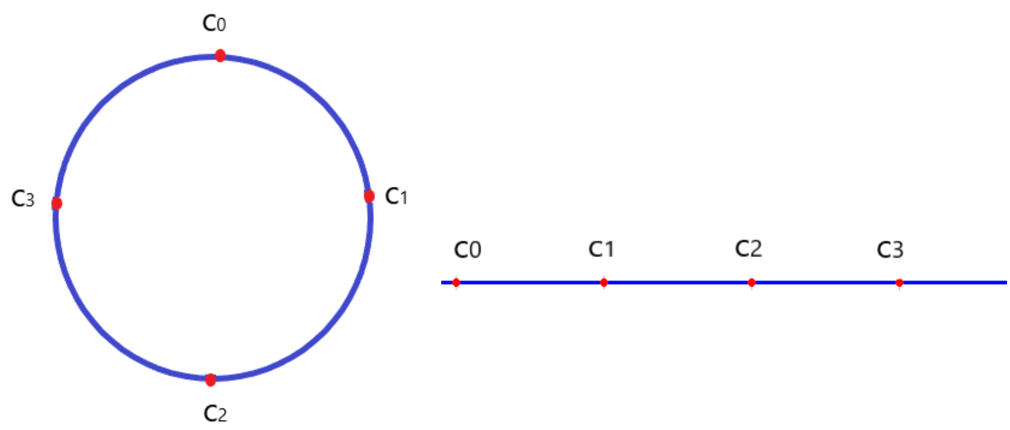

FIGURE 2.1: Nearest neighbours on a circle and a line

the definitions of circulant matrices since they are used as precision matrices too. Within the precision matrices, we can see dependences that exist between nodes which explain the structures of the graphs. The dependences between the nodes express sometimes a close distance between the nodes of a graph. However, this close distance can be interpreted in a different way. For example, we define differently the distance in time (temporal data), than the distance in spatial data. On a torus graph which explains spatial correlation, the nodes that are considered close are those which exhibit a circular behaviour. However, the circular behaviour of those nodes can be expressed differently in a matrix form. In Figure 2.1 we can see four nodes, c_0, c_1, c_2 and c_3 which are equidistant. What we can observe in the circle is that the closest neighbours of the c_0 node are the c_1 and c_3 nodes. However, when we unfold the circle in a line, only the c_1 neighbour of c_0 is close by and the c_3 seems the most distant node. In a circulant matrix the closest neighbours of a node can be on the first and last position even though they appear quite distant. In the following section we will explore further the structure and the properties of the circular matrices.

2.3.1 Circulant and Block-Circulant Matrices

In Definition 2.3 and 2.4 we can see how the circular matrices have similar features and we can explore further the structure of each matrix, [Rue and Held, 2005, chapter 2].

Definition 2.3 (Circulant Matrix) An $n \times n$ matrix $\boldsymbol{C}$ is circulant iff for some vector $\boldsymbol{c} = (c_0, c_1, c_2, ..., c_{n-1})^T$ it has the form:

$$\boldsymbol{C} = \begin{pmatrix} c_0 & c_1 & c_2 & \dots & c_{n-1} \\ c_{n-1} & c_0 & c_1 & \dots & c_{n-2} \\ c_{n-2} & c_{n-1} & c_0 & \dots & c_{n-3} \\ \vdots & \vdots & \vdots & & \vdots \\ c_1 & c_2 & c_3 & \dots & c_0 \end{pmatrix}$$

Here $\boldsymbol{c}$ is the *base* of $\boldsymbol{C}$.

Definition 2.4 (Block-Circulant Matrix) An $Nn \times Nn$ matrix $\mathbf{C}$ is block circulant with $N \times N$ blocks, iff it can be written as:

$$C = \begin{pmatrix} \mathbf{C}_0 & \mathbf{C}_1 & \mathbf{C}_2 & \dots & \mathbf{C}_{N-1} \\ \mathbf{C}_{N-1} & \mathbf{C}_0 & \mathbf{C}_1 & \dots & \mathbf{C}_{N-2} \\ \mathbf{C}_{N-2} & \mathbf{C}_{N-1} & \mathbf{C}_0 & \dots & \mathbf{C}_{N-3} \\ \vdots & \vdots & \vdots & & \vdots \\ \mathbf{C}_1 & \mathbf{C}_2 & \mathbf{C}_3 & \dots & \mathbf{C}_0 \end{pmatrix}$$

where $\boldsymbol{C_i}$ is a circulant $n \times n$ matrix with base $\boldsymbol{c_i}$. The base of $\boldsymbol{C}$ is the $n \times N$ matrix

$$c = \begin{pmatrix} \mathbf{c}_0 & \mathbf{c}_1 & \mathbf{c}_2 & \dots & \mathbf{c}_{N-1} \end{pmatrix}^T$$

Each column has n values specifying the base of each of the block matrices, and so we can write:

$$c = \begin{pmatrix} c_{0,0} & c_{0,1} & c_{0,2} & \dots & c_{0,N-1} \\ c_{1,0} & c_{1,1} & c_{1,2} & \dots & c_{1,N-1} \\ c_{2,0} & c_{2,1} & c_{2,2} & \dots & c_{2,N-1} \\ \vdots & \vdots & \vdots & & \vdots \\ c_{n-1,0} & c_{n-1,1} & c_{n-1,2} & \dots & c_{n-1,N-1} \end{pmatrix}$$

In definitions 2.3 and 2.4 we can observe the circulant behaviour inside the matrices. The first row has a sequence from c_0 to c_{n-1} and the second row starts from the last node of the first row which is c_{n-1} and then it continues the sequence from c_0 to c_{n-2} this time. This continues until the first node of the first row will appear in the final position of the last row. Overall, in circulant matrices the last nodes of a row are the first nodes in the next row highlighting a circular behaviour. The difference between the circulant and block circulant matrices is that instead of having a simple node c_0 we will have a whole matrix, $\mathbf{C}_0$. In block matrices we do not have nodes but matrices appearing as nodes. Block circulant matrices will be used as precision matrices in Section 2.6.2 and 2.6.3.

2.4 Canonical Parametrisation and Cholesky Factorisation

In the Bayesian framework, in order to identify the posterior distribution we need to define its parameters. If we assume that the posterior distribution is a multivariate normal distribution then we want to define its mean and its covariance matrix. Although it is easy in general to distinguish the covariance matrix, there are some difficulties to find the mean parameter. For example let us assume that $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$ and if we analyse it further we have, [Rue and Held, 2005, chapter 2]:

$$
\begin{aligned}
\log p(\mathbf{x}|\boldsymbol{\mu}, \mathbf{Q}) &\propto (\mathbf{x} - \boldsymbol{\mu})^T \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) \\
&\propto \mathbf{x}^T \mathbf{Q} \mathbf{x} - \boldsymbol{\mu}^T \mathbf{Q} \mathbf{x} - \mathbf{x}^T \mathbf{Q} \boldsymbol{\mu} + \boldsymbol{\mu}^T \mathbf{Q} \boldsymbol{\mu} \\
&\propto \mathbf{x}^T \mathbf{Q} \mathbf{x} - 2\mathbf{x}^T \mathbf{Q} \boldsymbol{\mu} + \boldsymbol{\mu}^T \mathbf{Q} \boldsymbol{\mu} \\
&\propto \mathbf{x}^T \mathbf{Q} \mathbf{x} - 2\mathbf{x}^T \mathbf{Q} \boldsymbol{\mu}
\end{aligned} \tag{2.5}
$$

In (2.5), we can identify the $\boldsymbol{\mu}$ and $\mathbf{Q}$ parameters from two terms. The first term of (2.5) has solely the $\mathbf{Q}$ parameter but the second term has not solely the $\boldsymbol{\mu}$ parameter and that imposes some difficulties to extract only the value of the $\boldsymbol{\mu}$ parameter.

Canonical parametrisation introduces the mean as the product of the mean and the covariance matrix, $\boldsymbol{\mu}\mathbf{Q}$ as it exists in the second term of (2.5). If $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1})$ then with the canonical parametrisation it will be $\mathbf{x} \sim \mathcal{N}_c(\mathbf{b}, \mathbf{Q}) = \mathbf{x} \sim \mathcal{N}_c(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$. The advantage of using canonical parametrisation is that we can have updates without explicitly knowing the mean of the canonical parametrization until it becomes needed. In order to compute the mean of the canonical form we have to find the solution of $\mathbf{Q}\boldsymbol{\mu} = \mathbf{b}$ and that is computed with the Cholesky factorization. The canonical parametrization will be used extensively for the updates of some model parameters later in the thesis. Its useful properties make the computations easier and faster.

Since $\mathbf{Q}$ is an IGMRF it expresses the relations between the nodes and as it describes the conditional dependences we expect a sparse $\mathbf{Q}$ matrix. Another way of writing the precision matrix, $\mathbf{Q}$ is by using the product of two triangular matrices which are by definition sparse and also inherit the sparsity of the initial matrix, $\mathbf{Q}$. Hence, the so-called Cholesky factorization can be described as, [Rue and Held, 2005, chapter 2] :

$$\mathbf{Q} = \mathbf{L}\mathbf{L}^T \tag{2.6}$$

where $\mathbf{L}$ is a lower triangular matrix. It is a strong advantage to be able to write the $\mathbf{Q}$ matrix using a triangular sub-matrix as it gives us very fast calculations. Canonical parametrisation in combination with Cholesky factorization can give all the information that we need to define the $\boldsymbol{\mu}$ parameter of a normal distribution. The steps that we follow for the definition of the $\boldsymbol{\mu}$ parameter are presented below:

1. Canonical parametrization: $\mathbf{Q}\boldsymbol{\mu} = \mathbf{b}$
2. Implementing Cholesky factorization: $\mathbf{L}\mathbf{L}^T\boldsymbol{\mu} = \mathbf{b}$
3. Forward solve: $\mathbf{L}\mathbf{v} = \mathbf{b}$
4. Backward solve: $\mathbf{L}^T\boldsymbol{\mu} = \mathbf{v}$

By reaching the 4^{th} step $\mathbf{L}$ and $\mathbf{v}$ are known values which are used to compute $\boldsymbol{\mu}$.

2.5 One-Dimensional Random Walk

In this section we will explore different types of IGMRFs in one dimension. Their differences are observed through their precision matrices. The precision matrices will differ based on the conditional dependence that we require each time. Finally, the precision matrix in each type consists of a scalar precision parameter, λ and a structure matrix, $\mathbf{P}$. We will use the term *order* of the matrix which indicates the number of nodes that can be dependent. Following we will define the forward differences that are depending on the order. The first-order difference is defined as:

$$\Delta u_{[i]} = u_{[i+1]} - u_{[i]} \tag{2.7}$$

whereas the second order difference can be defined as:

$$\begin{aligned} \Delta^2 u_{[i]} = \Delta(\Delta u_{[i]}) &= \Delta(u_{[i+1]} - u_{[i]}) \\ &= \Delta u_{[i+1]} - \Delta u_{[i]} \\ &= u_{[i+2]} - 2u_{[i+1]} + u_{[i]} \end{aligned} \tag{2.8}$$

The higher the order the more nodes we are including. Finally, a k order difference can be defined as:

$$\Delta^k u_{[i]} = \Delta^{k-1}(\Delta u_{[i]}) \tag{2.9}$$

2.5.1 First-Order Random Walk for Regular Locations

We will firstly describe the first-order random walk making use of the IGMRFs, [Rue and Held, 2005, chapter 3.3]. We want to take into account the closest neighbours of a parameter, $\mathbf{u} = (u_{[1]}, ..., u_{[n]})^T$. So instead of examining the individual values inside the $\mathbf{u}$ vector we will construct possible pairs of values that can exist inside the $\mathbf{u}$ vector. For example, we want to compute the distance between $u_{[i]}$ and $u_{[j]}$ values of a total of n values in the $\mathbf{u}$ vector. Since we are in the case of first order it means that we will take one closest neighbour each time, hence the $u_{[j]}$ node needs to be one of the closest values to the $u_{[i]}$ node. In other words, we consider pairs of values inside a vector and make calculations for pairs which can be expressed as differences instead of calculating the individual $u_{[i]}$ or $u_{[j]}$. The way we construct the difference between the values is stated in (2.10):

$$\begin{aligned} u_{[i]} - u_{[j]} &\sim \mathcal{N}(0, \lambda^{-1}) \\ \Delta u_{[i]} &\sim \mathcal{N}(0, \lambda^{-1}) \end{aligned} \tag{2.10}$$

The density can be written as:

$$\begin{aligned} \pi(\mathbf{u}|\lambda) &\propto \lambda^{(T-1)/2} \exp\left(-\frac{\lambda}{2}\sum_{i=2}^{T-1}(\Delta u_{[i]})^2\right) \\ &\propto \lambda^{(T-1)/2} \exp\left(-\frac{\lambda}{2}\sum_{i=2}^{T-1}(u_{[i+1]} - u_{[i]})^2\right) \\ &\propto \lambda^{(T-1)/2} \exp\left(-\frac{\lambda}{2}(\mathbf{Du})^T\mathbf{Du}\right) \\ &\propto \lambda^{(T-1)/2} \exp\left(-\frac{1}{2}\mathbf{u}^T\lambda\mathbf{Pu}\right) \\ &\propto \lambda^{(T-1)/2} \exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{Qu}\right) \end{aligned} \tag{2.11}$$

Based on the above we know that $\mathbf{Q}$ can be written using the structure matrix, $\mathbf{P}$ and the scalar precision parameter, λ:

$$\mathbf{Q} = \lambda\mathbf{D}^T\mathbf{D} = \lambda\mathbf{P} \tag{2.12}$$

$$\lambda\mathbf{D}^T\mathbf{D} = \lambda \begin{pmatrix} 1 & & & \\ -1 & 1 & & \\ & -1 & 1 & \\ & & -1 & 1 \\ & & & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & 1 & -1 & \\ & & & 1 & -1 \end{pmatrix}$$

$$\mathbf{Q} = \lambda\mathbf{P} = \lambda \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} \tag{2.13}$$

Hence the vector $\mathbf{u} \sim \mathcal{N}(0, \lambda\mathbf{P}) = \{u_{[1]}, u_{[2]}, ..., u_{[n]}\} \sim \mathcal{N}(0, \lambda\mathbf{P})$

In the $\mathbf{D}$ matrix we can see the different pairs that can exist inside the $\mathbf{u}$ vector when we have $n = 5$ values inside it and therefore, $n - 1 = 4$ possible pairs. The reason that we construct the pairs is that we want to introduce correlation between the values of a vector that are the closest. Hence, we will use three nodes in total in the central rows of the $\mathbf{P}$ matrix.

The $\mathbf{P}$ matrix is expressed as the squared matrix of the $\mathbf{D}$ matrix. Hence, we expect more values to be included, the so-called neighbours. The interpretation of the first line of the $\mathbf{P}$ matrix in (2.13) is that when we are at the $u_{[1]}$ time we will borrow information from the $u_{[1+1]}$ time since we do not have other time events before. In the second row and until the second last we can borrow information from the time before, $u_{[i-1]}$ and the time after, $u_{[i+1]}$. In the first row we can only borrow information from the node ahead and in the last row from the node before since we have specific start and end points. Figure 2.2 gives an illustration of the first two rows of the structure sparse matrix $\mathbf{P}$.

A sparse matrix or sparse array is a matrix in which most of the elements are zero. In our analysis, we use a sparse matrix to describe the precision matrix which is the inverse covariance matrix. Specifically, by using the term sparse matrix we are referring to the conditional independence through the nodes of the precision matrix. Nodes that are conditionally dependent have non-zero values whereas nodes that are conditionally independent are expressed with zero values. Therefore, sparsity explains mostly the independence between nodes which thus produce zeros.

The increments of this model are defined as : $\Delta u_{[i]} = u_{[i+1]} - u_{[i]}$ which comprise the $\mathbf{D}$ matrix. Apart from corrections at the boundary, which are the first and last row, the $\mathbf{P}$ matrix consists of $-\Delta^2 u_{[i]}$ terms which are shown in (2.8). We

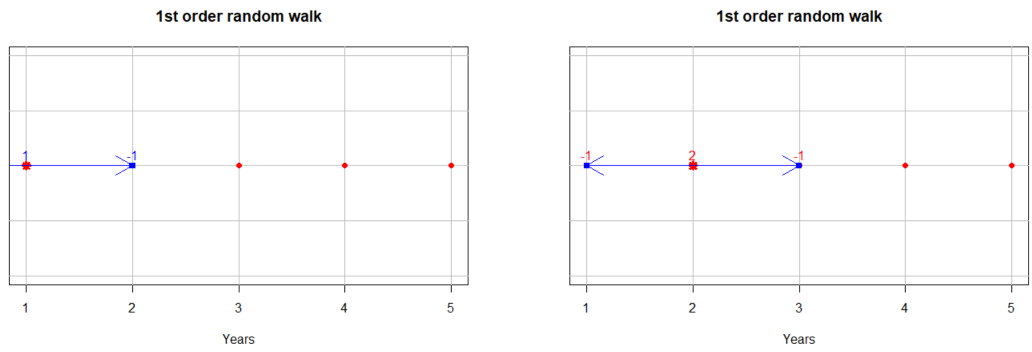

FIGURE 2.2: Nearest neighbours on a line for a first-order random walk

may interpret $(-u_{[i-1]} + 2u_{[i]} - u_{[i+1]})$ as an estimate of an underlying continuous function $u_{[i]}$ at $i = 1$ making use of the observations at $\{i-1, i, i+1\}$ positions. Due to the boundary corrections we have $T-1$ restrictions in the model. In (2.11) we can see clearly that we used $T-1$ as a linear constraint. In summary, since we have the pairs of T terms then we expect to have $T-1$ different pairs. In order to produce a proper density we impose the constraint of $\sum u_{[i]} = 0$ which has as a result a finite marginal standard deviation. Summing up, the *first order* declares the number of steps we can go in order to reach neighbours. The intuition is that we can choose up to one step in every possible direction, backwards and forward if there are available nodes. So, we can possibly have in total two neighbours plus one, the node under study.

2.5.2 Second-Order Random Walk for Regular Locations

In this case the distances that we want to take into account have the form shown in (2.14). The difference of the results in this subsection from the previous one is that we want to include more neighbours, [Rue and Held, 2005, chapter 3.4]. Therefore, the structure of the pairs between the values in the vector $\mathbf{u}$ will be different. Instead of having the difference between two values we will take the differences between three values each time. By using the term *second order* we declare how many neighbours we want to include each time, two closest neighbours and the value under study which gives us in total three values in each row of the

$\mathbf{D}$ matrix.

$$u_{[i+2]} - 2u_{[i+1]} + u_{[i]} = \Delta^2 u_{[i]} \sim N(0, \lambda^{-1}) \tag{2.14}$$

The density can be written as:

$$\begin{aligned}
\pi(\mathbf{u}|\lambda) &\propto \lambda^{(T-2)/2} \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{T-2}(\Delta^2 u_{[i]})^2\right) \\
&\propto \lambda^{(T-2)/2} \exp\left(-\frac{\lambda}{2}\sum_{i=1}^{T-2}(u_{[i+2]} - 2u_{[i+1]} + u_{[i]})^2\right) \\
&= \lambda^{(T-2)/2} \exp\left(-\frac{\lambda}{2}(\mathbf{Du})^T(\mathbf{Du})\right) \\
&= \lambda^{(T-2)/2} \exp\left(-\frac{\lambda}{2}\mathbf{u}^T\mathbf{D}^T\mathbf{Du}\right) \\
&= \lambda^{(T-2)/2} \exp\left(-\frac{1}{2}\mathbf{u}^T\lambda\mathbf{D}^T\mathbf{Du}\right) \\
&= \lambda^{(T-2)/2} \exp\left(-\frac{1}{2}\mathbf{u}^T\lambda\mathbf{Pu}\right) \\
&= \lambda^{(T-2)/2} \exp\left(-\frac{1}{2}\mathbf{u}^T\mathbf{Qu}\right)
\end{aligned} \tag{2.15}$$

Based on (2.15) we know that $\mathbf{Q}$ can be written using the $\mathbf{P}$ matrix and the λ constant:

$$\mathbf{Q} = \lambda\mathbf{D}^T\mathbf{D} = \lambda\mathbf{P} \tag{2.16}$$

$$\lambda \begin{pmatrix} 1 & & & \\ -2 & 1 & & \\ 1 & -2 & 1 & \\ & 1 & -2 & 1 \\ & & 1 & -2 \\ & & & 1 \end{pmatrix} \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & 1 & -2 & 1 & \\ & & & 1 & -2 & 1 \end{pmatrix}$$

$$\mathbf{Q} = \lambda \begin{pmatrix} 1 & -2 & 1 & & & \\ -2 & 5 & -4 & 1 & & \\ 1 & -4 & 6 & -4 & 1 & \\ & 1 & -4 & 6 & -4 & 1 \\ & & 1 & -4 & 5 & -2 \\ & & & 1 & -2 & 1 \end{pmatrix} \tag{2.17}$$

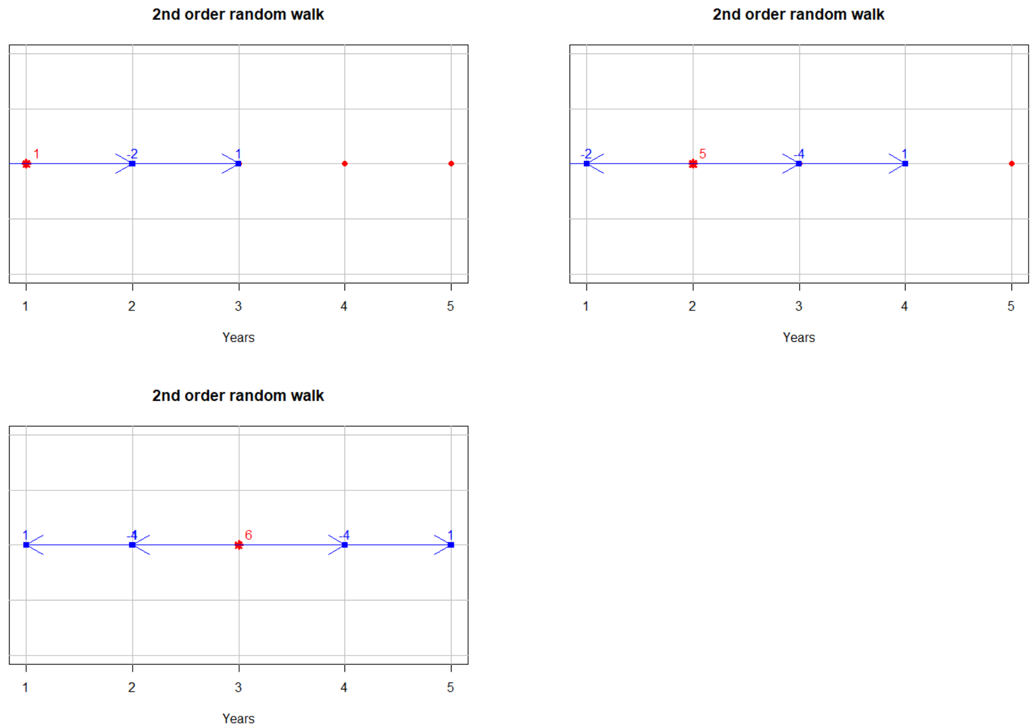

FIGURE 2.3: Nearest neighbours on a line for a second-order random walk

Hence the vector $\mathbf{u} = (u_{[1]}, u_{[2]}, .., u_{[T]})^T \sim \mathcal{N}(0, \lambda \mathbf{P})$ for all the T nodes.
The interpretation of the first line of the $\mathbf{P}$ matrix in (2.17) is that when we are at the $u_{[1]}$ node we will borrow information from the $u_{[2]}$ and the $u_{[3]}$ nodes since we do not have other nodes before. In the second row we will be at the $u_{[2]}$ node, we will borrow information from the node before, $u_{[1]}$ and two nodes forward, $u_{[3]}$ and $u_{[4]}$. From the third to fifth row the behaviour between the nodes is the same. If we are at the $u_{[i]}$ node and we will borrow information from the two nodes before $u_{[i-1]}$ and $u_{[i-2]}$ and two nodes afterwards, $u_{[i+1]}$ and $u_{[i+2]}$. We can notice that we have a standard behaviour in the central rows and two different behaviours in the first and the second row. The behaviours in the first two rows are known as corrections at the boundary. We need again to impose restrictions which this time will be $\sum_i u_{[i]} = 0$ and $\sum_i i u_{[i]} = 0$. The linear constraints give us a proper joint density and a finite marginal standard deviation. Therefore, now the linear restrictions are $T-2$ which we can also observe in (2.15). In Figure 2.3 is depicted an illustration of the first three rows in the structure matrix $\mathbf{P}$. In these three plots we can observe the different behaviour that exists in the $\mathbf{P}$ matrix. The two

first plots are those with the boundary constraints and the third one follows the behaviour of the central rows.

The increments of the $\mathbf{D}$ matrix will be $\Delta^2 u_{[i]} = u_{[i+2]} - 2u_{[i+1]} + u_{[i]}$ whereas $\mathbf{P}$ consists of -$\Delta^4 u_{[i]}$ terms apart from corrections at the boundaries as expressed in (2.9). We may interpret $u_{[i-2]} - 4u_{[i-1]} + 6u_{[i]} - 4u_{[i+1]} + u_{[i+2]}$ as an estimate of an underlying continuous function $u_{[i]}$ at $i = 1$ making use of the observations at $[i-2, i-1, i, i+1, i+2]$ positions. The middle row of the precision matrix is: $u_{[i-2]} - 4u_{[i-1]} + 6u_{[i]} - 4u_{[i+1]} + u_{[i+2]}$ and finally the linear restrictions are: $T - 2$, where T expresses the total number of nodes.

In summary, the *second order* declares the number of steps we can go in order to reach neighbours. The intuition is that we can choose up to two steps in every possible direction, backwards and forward so we can possibly have in total four neighbours plus the node under study.

2.6 Second-Order Two-dimensional Random Walk

We will now explore different types of IGMRFs in two dimensions, [Rue and Held, 2005, chapter 3.4.2]. The differences between the types are observed through their precision matrices and it is the conditional dependence inside the precision matrices that changes in each case. As previously, the precision matrix in each type consists of a scalar precision parameter, λ and a structure matrix, $\mathbf{P}$.

In the two-dimensional IGMRF we will account for the distances of two variables. Given that we use the *second order* term we indicate that we need the two closest neighbours to be included for each of the two variables this time. Hence, the distances are defined in the following way:

$$\Delta^2_{(1,0)} u_{[i,j]} + \Delta^2_{(0,1)} u_{[i,j]} \sim \mathcal{N}(0, \lambda^{-1}) \tag{2.18}$$

where $\Delta_{(1,0)}$ and $\Delta_{(0,1)}$ are the forward difference operators in the vertical and horizontal directions respectively, [Rue and Held, 2005, chapter 3.4.2]. We have again the precision parameter λ which represents again scalar precision parameter as we have seen in the one-dimensional case (see Section 2.5). The first-order forward differences can be written as:

$$\begin{aligned}\Delta_{(1,0)}u_{[i,j]} &= u_{[i+1,j]} - u_{[i,j]}\\ \Delta_{(0,1)}u_{[i,j]} &= u_{[i,j+1]} - u_{[i,j]}\end{aligned} \tag{2.19}$$

and the second-order forward difference for each variable is defined as:

$$\begin{aligned}\Delta^2_{(1,0)}u_{[i,j]} &= \Delta_{(1,0)}(\Delta_{(1,0)}u_{[i,j]}) = u_{[i+2,j]} - 2u_{[i+1,j]} + u_{[i,j]}\\ \Delta^2_{(0,1)}u_{[i,j]} &= \Delta_{(0,1)}(\Delta_{(0,1)}u_{[i,j]}) = u_{[i,j+2]} - 2u_{[i,j+1]} + u_{[i,j]}\end{aligned} \tag{2.20}$$

By using the definitions in (2.20) we can write (2.18) as:

$$u_{[i+2,j]} - 2u_{[i+1,j]} + 2u_{[i,j]} + u_{[i,j+2]} - 2u_{[i,j+1]} \sim N(0, \lambda^{-1}) \tag{2.21}$$

However, by changing the notation we have the equivalent:

$$\begin{aligned}u_{[i-1,j]} - 2u_{[i,j]} + u_{[i+1,j]} + u_{[i,j-1]} - 2u_{[i,j]} + u_{[i,j+1]} &=\\ u_{[i-1,j]} - 4u_{[i,j]} + u_{[i+1,j]} + u_{[i,j-1]} + u_{[i,j+1]} &\sim N(0, \lambda^{-1})\end{aligned} \tag{2.22}$$

Summing up, equations (2.21) and (2.22) are expressing exactly the same result but the difference is that they include different neighbours and give different weights to their neighbours. In (2.21) the central nodes are $u_{[i+1,j]}$ and $u_{[i,j+1]}$ whereas in (2.22) the central node is $u_{[i,j]}$ and these central nodes have the largest weights. In both (2.21) and (2.22) we can see that the summation of the weights is equal to zero. An illustration of (2.21) and (2.22) is presented in Figure 2.4.

2.6.1 Illustration of the problem

We need to find the closest neighbours for the combination of the two variables. The two variables are indicated with the indices i, j in the $u_{[i,j]}$ term. We assume that both variables have five nodes each. Hence, we want to account for all the possible combinations between the nodes of the first and the second variable. With the number of nodes for each variable being five the possible combinations will be

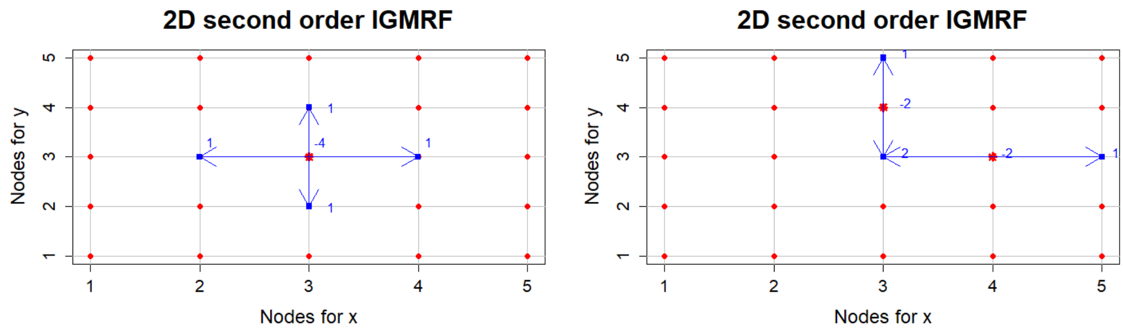

FIGURE 2.4: Nearest neighbours for a 2D second-order random walk with different centres

$5^2 = 25$. Although every combination is possible, not all of them will be used as neighbours in a specific pair and thus there will be combinations having a zero value in the precision matrix. More details can be seen by observing Figures 2.5, 2.6 and 2.7. These plots illustrate the procedure of choosing neighbours for a specific pair. We can see how the IGMRF works having as a starting point the $_{[2003,2003]}$ pair. In the $\mathbf{D}$ matrix of Figure 2.6 we include the two closest neighbours in the vertical and the two closest neighbours in the horizontal direction. The neighbours that are chosen are those one step away from the pair under study. In Figure 2.7 we can see that we take more neighbours into account. We started with two neighbours in each vertical and horizontal direction in the $\mathbf{D}$ matrix and we conclude with four neighbours in each direction in the $\mathbf{P}$ matrix. In the matrix $\mathbf{P}$, we choose neighbours that are one and two steps away from the pair under study. Also, in the $\mathbf{P}$ matrix we can see that besides the vertical and horizontal direction we can also move diagonally. The pairs from the diagonal direction are $_{[2002,2004]}$, $_{[2004,2004]}$, $_{[2004,2002]}$, $_{[2002,2002]}$. Finally, what is also interesting in these figures is that the closest neighbours of the central pair are one pair forward and one pair backwards. Not only can we go horizontally, vertically and diagonally but also we have forward and backwards steps.

$$\mathbf{u}_{D,S}(t) = \begin{pmatrix} u_{[1]} \\ \vdots \\ u_{[T^2]} \end{pmatrix} \sim \mathcal{N}(\mathbf{0}, (\lambda \mathbf{P})^{-1}) \tag{2.23}$$

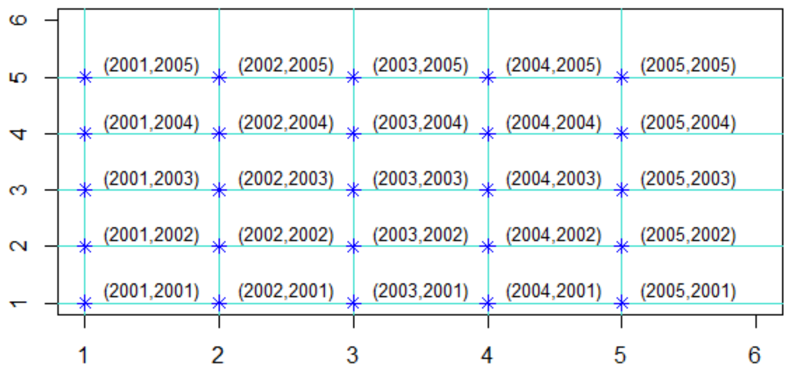

Figure 2.5: Possible combinations between the years of X and Y variable

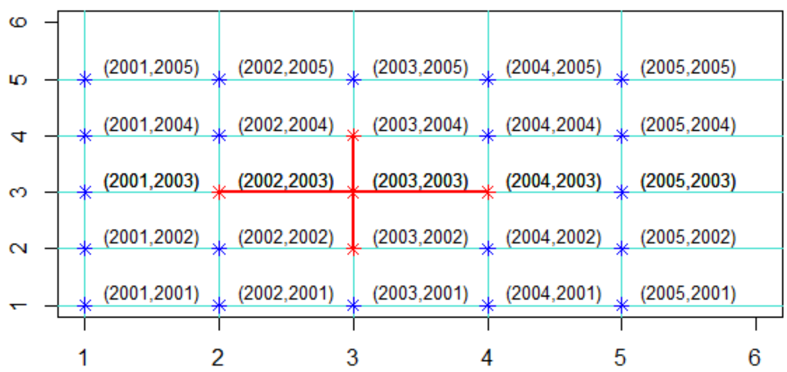

Figure 2.6: The 5 nearest neighbours for the years of X and Y variable, starting from the (2003, 2003) combination (D matrix)

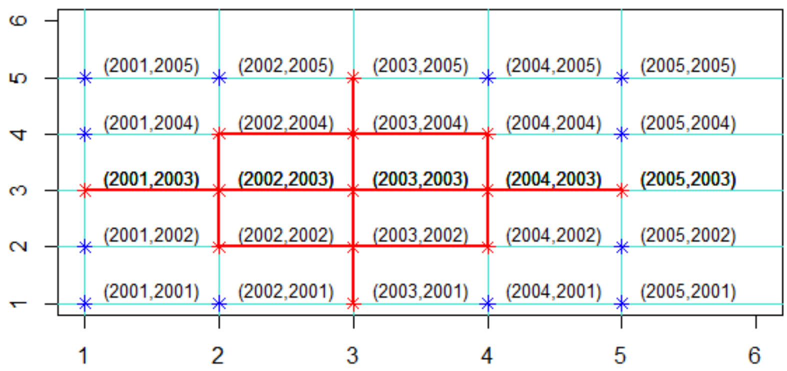

Figure 2.7: The 13 nearest neighbours for the X and Y years ($P = D^T D$ Precision matrix)

In (2.23) we state that the updated version of $\mathbf{u}$'s prior distribution has mean $\mathbf{0}$ of length T^2 and $\mathbf{P}$ which is a matrix of $T^2 \times T^2$ dimensions where T expresses the number of nodes for each variable. In conclusion, some of the combinations express the changes only in the first variable when the second variable remains constant, some others the changes only in the second variable when the first remains constant and the rest express the interaction of these two variables when both variables are changing, either both decreasing or increasing or one is increasing and the other is decreasing. The possible combinations for both variables can be written as: $\mathbf{u}_{[D,S]} = (\mathbf{u}_{[1,1]}, \mathbf{u}_{[1,2]}, \ldots, \mathbf{u}_{[2,3]}, \ldots, \mathbf{u}_{[5,5]})$
However, an alternative interpretation using coding from 1 to T^2, as the total number of combinations, also used in (2.23), is: $\mathbf{u}_{[D,S]} = (\mathbf{u}_{[1]}, \mathbf{u}_{[2]}, \ldots, \mathbf{u}_{[15]}, \ldots, \mathbf{u}_{[25]})$.
In the following chapters we will explore different types of precision matrices that can be used in the case of two dimensional second order IGMRFs. More details describing what each pair of combinations between the first and second variable's nodes expresses will be shown.

2.6.2 Precision matrix on a torus

Based on the bibliography Rue and Held [2005], the most common use of IGMRF is through the use of a torus. As we show in Section 2.6.1 the increments of the $\mathbf{D}$ matrix will be:

$$u_{[i+1,j]} + u_{[i-1,j]} + u_{[i,j+1]} + u_{[i,j-1]} - 4u_{[i,j]} \tag{2.24}$$

The summation of the coefficients that are used in (2.24) is equal to zero, [Rue and Held, 2005, chapter 3.4.2].
Bearing in mind that we have two variables as we are in a two dimensional case we need to construct accordingly the precision matrix. Defining the two variables with the indices i and j, we want to examine how these two variables will interact together. What is more, we want in every event to include information for both of these variables.

The precision matrix of the prior distribution will be defined again as $\mathbf{Q} = \lambda \times \mathbf{D}^T\mathbf{D}$ where $\mathbf{D}$ is a block-circulant matrix, there is a scalar precision parameter, λ and by $\mathbf{P} = \mathbf{D}^T\mathbf{D}$ we define the structure matrix. The definition of the precision matrix is the same as in the one-dimensional case but the structure matrix will be constructed differently in order to include two variables. The base of the $\mathbf{D}$ matrix is presented below:

$$\begin{pmatrix} -4 & 1 & 0 & 0 & 1 \\ 1 & & & & \\ 0 & & & & \\ 0 & & & & \\ 1 & & & & \end{pmatrix}$$

Analytically, the base model consists of sub-matrices. The sub-matrices that we use for the construction of the $\mathbf{D}$ matrix are presented below:

$$\mathbf{A}_1 = \begin{pmatrix} -4 & 1 & & & 1 \\ 1 & -4 & 1 & & \\ & 1 & -4 & 1 & \\ & & 1 & -4 & 1 \\ 1 & & & 1 & -4 \end{pmatrix} \mathbf{A}_2 = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}$$

so $\mathbf{D}$ will be:

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.25}$$

The $\mathbf{0}$ is presented in bold form as it stands for a 5×5 matrix full of zeros. Equation (2.25) shows the relationship of the nodes in the central rows without taking into account the corrections at the boundaries which are referring to the first and last rows. As we have seen in the first and in the second order random walk the behaviour at the boundaries is different. Observing the rows in (2.25) we

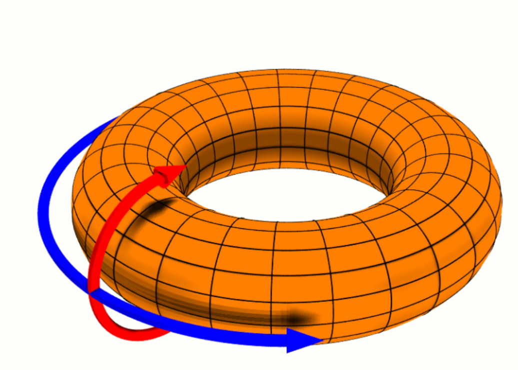

FIGURE 2.8: Graph of torus, from browser StackExchange [2021]

can see that the last node of the first row is becoming the first node of the next row and so on. The circular behaviour expresses the torus plot shown in Figure 2.8. In Figure 2.9 we can see the $\mathbf{D}$ matrix and every row shows the neighbours that are included for a specific node. In Figure 2.10 we apply the $\mathbf{P} = \mathbf{D}^T\mathbf{D}$ equation and we can see which neighbours are present in the structure matrix. Based on Figure 2.10 we can see that the middle row of the structure matrix, $\mathbf{P}$ multiplied with the $\mathbf{u}$ vector can give the non-zero values of the middle row of the precision matrix:

$$\begin{aligned} u_{[i-2,j]} + 2u_{[i-1,j-1]} - 8u_{[i-1,j]} + 2u_{[i-1,j+1]} + u_{[i,j-2]} - 8u_{[i,j-1]} + \\ 20u_{[i,j]} - 8u_{[i,j+1]} + u_{[i,j+2]} + 2u_{[i+1,j-1]} - 8u_{[i+1,j]} + 2u_{[i+1,j+1]} + u_{[i+2,j]} \end{aligned} \tag{2.26}$$

As is stated by Thon and Rue Thon et al. [2012] the structure matrix in Figure 2.9 has a torus form and therefore, does not have a special treatment at the boundaries. It includes as neighbouring nodes those that exist on opposing edges which in a torus are considered the closest neighbours. Furthermore, in the $\mathbf{Q}$ matrix we are taking into account the interactions by including the terms that exist in the diagonal direction.

It is necessary in this case too, to impose linear constraints. The linear constraints will be greater in number as we consider more variables. They will refer to the first variable $\sum iu_{[i,j]} = 0$, to the second variable, $\sum ju_{[i,j]} = 0$ and for both of them $\sum\sum u_{[i,j]} = 0$. Therefore, the linear restriction is defined as $T^2 - 3$ where T is referring to the number of nodes that each variable has and T^2 the possible combinations between the two variables. The linear restrictions are needed in

1.1	-4	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
1.2	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
1.3	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
1.4	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
1.5	1	.	.	1	-4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
2.1	1	.	.	.	.	-4	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2.2	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
2.3	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
2.4	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
2.5	.	.	.	.	1	1	.	.	1	-4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
3.1	.	.	.	.	.	1	.	.	.	.	-4	1	.	.	1	1	.	.	.	.	.	.	.	.	.
3.2	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.
3.3	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.
3.4	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.
3.5	.	.	.	.	.	.	.	.	.	1	1	.	.	1	-4	.	.	.	.	1	.	.	.	.	.
4.1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-4	1	.	.	1	1	.	.	.	.
4.2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.
4.3	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.
4.4	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.
4.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	-4	.	.	.	.	1
5.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-4	1	.	.	1
5.2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.
5.3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.
5.4	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1
5.5	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	-4

FIGURE 2.9: **D** matrix for 5 years of SBP and DBP combinations

1.1	20	-8	1	1	-8	-8	2	.	.	2	1	.	.	.	.	1	.	.	.	.	-8	2	.	.	2
1.2	-8	20	-8	1	1	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	.
1.3	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.
1.4	1	1	-8	20	-8	.	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2
1.5	-8	1	1	-8	20	2	.	.	2	-8	.	.	.	.	1	.	.	.	.	1	2	.	.	2	-8
2.1	-8	2	.	.	2	20	-8	1	1	-8	-8	2	.	.	2	1	.	.	.	.	1	.	.	.	.
2.2	2	-8	2	.	.	-8	20	-8	1	1	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.
2.3	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.
2.4	.	.	2	-8	2	1	1	-8	20	-8	.	.	2	-8	2	.	.	.	1	.	.	.	.	1	.
2.5	2	.	.	2	-8	-8	1	1	-8	20	2	.	.	2	-8	.	.	.	.	1	.	.	.	.	1
3.1	1	.	.	.	.	-8	2	.	.	2	20	-8	1	1	-8	-8	2	.	.	2	1	.	.	.	.
3.2	.	1	.	.	.	2	-8	2	.	.	-8	20	-8	1	1	2	-8	2	.	.	.	1	.	.	.
3.3	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.
3.4	.	.	.	1	.	.	.	2	-8	2	1	1	-8	20	-8	.	.	2	-8	2	.	.	.	1	.
3.5	.	.	.	.	1	2	.	.	2	-8	-8	1	1	-8	20	2	.	.	2	-8	.	.	.	.	1
4.1	1	.	.	.	.	1	.	.	.	.	-8	2	.	.	2	20	-8	1	1	-8	-8	2	.	.	2
4.2	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	.	-8	20	-8	1	1	2	-8	2	.	.
4.3	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.
4.4	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	1	1	-8	20	-8	.	.	2	-8	2
4.5	.	.	.	.	1	.	.	.	.	1	2	.	.	2	-8	-8	1	1	-8	20	2	.	.	2	-8
5.1	-8	2	.	.	2	1	.	.	.	.	1	.	.	.	.	-8	2	.	.	2	20	-8	1	1	-8
5.2	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	.	-8	20	-8	1	1
5.3	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1
5.4	.	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	1	1	-8	20	-8
5.5	2	.	.	2	-8	.	.	.	.	1	.	.	.	.	1	2	.	.	2	-8	-8	1	1	-8	20

FIGURE 2.10: Structure matrix for 5 years of SBP and DBP combinations, $\mathbf{P} = \mathbf{D}^T\mathbf{D}$

order to give a proper joint density. The rank deficiency is $T^2 - 1$. Finally, we can express the precision matrix, $\mathbf{Q}$ by using forward differences as:

$$-(\Delta^2_{(1,0)} + \Delta^2_{(0,1)})^2 u_{[i,j]} = (-\Delta^4_{(1,0)} - \Delta^4_{(0,1)} + 2\Delta^2_{(1,0)}\Delta^2_{(0,1)})u_{[i,j]} \qquad (2.27)$$

In (2.27) the last term indicates the interaction between the two variables. However, the **D** and the **Q** matrices based on Figures 2.9, 2.10 do not take into account the boundary effects, and we observe the same behaviour in each combination.

2.6.3 Isotropic Precision Matrix on a Torus

Since the precision matrix in Section 2.6.2 had some drawbacks, Rue and Held [Rue and Held, 2005, chapter 3.4.2] tried to overcome these obstacles by introducing an isotropic precision matrix on a torus. In (2.25) the structure matrix does not include the corners $u_{[1,1]}$, $u_{[1,n]}$, $u_{[n,1]}$, $u_{[n,n]}$ which is incomplete if we want to have information from each node. Therefore, a new suggestion which is an isotropic precision matrix includes every node.

In (2.11) we can observe that every row has the same number of elements following a cyclic behaviour moving from one row to the other. In (2.12) we can see that the structure matrix is not sparse but this is happening only because we used the minimum number of nodes equal to five. When the nodes are increased (above five) the matrix is sparse as with the rest of the suggestions. The sub-matrices that are used for the construction of the structure matrix, **P** are presented below:

$$\mathbf{A}_1 = \begin{pmatrix} -20 & 4 & & & 4 \\ 4 & -20 & 4 & & \\ & 4 & -20 & 4 & \\ & & 4 & -20 & 4 \\ 4 & & & 4 & -20 \end{pmatrix} \mathbf{A}_2 = \begin{pmatrix} 4 & 1 & & & 1 \\ 1 & 4 & 1 & & \\ & 1 & 4 & 1 & \\ & & 1 & 4 & 1 \\ 1 & & & 1 & 4 \end{pmatrix}$$

so **D** will be:

$$\begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{0} & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.28}$$

As stated by Rue and Held Rue and Held [2005] and Thon and Rue Thon et al. [2012] the structure matrix in Figure 2.12 also has a torus form. However, the boundaries still do not exhibit a different behaviour to the central rows and since a torus behaviour is followed the closest nodes are not included. This matrix

1.1	-20	4	.	.	4	4	1	.	.	1	.	.	.	.	.	.	.	.	.	.	4	1	.	.	1
1.2	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.
1.3	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.
1.4	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1
1.5	4	.	.	4	-20	1	.	.	1	4	.	.	.	.	.	.	.	.	.	.	1	.	.	1	4
2.1	4	1	.	.	1	-20	4	.	.	4	4	1	.	.	1	.	.	.	.	.	.	.	.	.	.
2.2	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.
2.3	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.
2.4	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.
2.5	1	.	.	1	4	4	.	.	4	-20	1	.	.	1	4	.	.	.	.	.	.	.	.	.	.
3.1	.	.	.	.	.	4	1	.	.	1	-20	4	.	.	4	4	1	.	.	1	.	.	.	.	.
3.2	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.	.
3.3	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.	.
3.4	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.	.	.	.
3.5	.	.	.	.	.	1	.	.	1	4	4	.	.	4	-20	1	.	.	1	4	.	.	.	.	.
4.1	.	.	.	.	.	.	.	.	.	.	4	1	.	.	1	-20	4	.	.	4	4	1	.	.	1
4.2	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.	.
4.3	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1	.
4.4	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.	1	4	1
4.5	.	.	.	.	.	.	.	.	.	.	1	.	.	1	4	4	.	.	4	-20	1	.	.	1	4
5.1	4	1	.	.	1	.	.	.	.	.	.	.	.	.	.	4	1	.	.	1	-20	4	.	.	4
5.2	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.	.
5.3	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4	.
5.4	.	.	1	4	1	.	.	.	.	.	.	.	.	.	.	.	.	1	4	1	.	.	4	-20	4
5.5	1	.	.	1	4	.	.	.	.	.	.	.	.	.	.	1	.	.	1	4	4	.	.	4	-20

FIGURE 2.11: **D** matrix for 5 years of SBP and DBP combinations

	1.1	1.2	1.3	1.4	1.5	2.1	2.2	2.3	2.4	2.5	3.1	3.2	3.3	3.4	3.5	4.1	4.2	4.3	4.4	4.5	5.1	5.2	5.3	5.4	5.5
1.1	468	-144	18	18	-144	-144	-8	8	8	-8	18	8	1	1	8	18	8	1	1	8	-144	-8	8	8	-8
1.2	-144	468	-144	18	18	-8	-144	-8	8	8	8	18	8	1	1	8	18	8	1	1	-8	-144	-8	8	8
1.3	18	-144	468	-144	18	8	-8	-144	-8	8	1	8	18	8	1	1	8	18	8	1	8	-8	-144	-8	8
1.4	18	18	-144	468	-144	8	8	-8	-144	-8	1	1	8	18	8	1	1	8	18	8	8	8	-8	-144	-8
1.5	-144	18	18	-144	468	-8	8	8	-8	-144	8	1	1	8	18	8	1	1	8	18	-8	8	8	-8	-144
2.1	-144	-8	8	8	-8	468	-144	18	18	-144	-144	-8	8	8	-8	18	8	1	1	8	18	8	1	1	8
2.2	-8	-144	-8	8	8	-144	468	-144	18	18	-8	-144	-8	8	8	8	18	8	1	1	8	18	8	1	1
2.3	8	-8	-144	-8	8	18	-144	468	-144	18	8	-8	-144	-8	8	1	8	18	8	1	1	8	18	8	1
2.4	8	8	-8	-144	-8	18	18	-144	468	-144	8	8	-8	-144	-8	1	1	8	18	8	1	1	8	18	8
2.5	-8	8	8	-8	-144	-144	18	18	-144	468	-8	8	8	-8	-144	8	1	1	8	18	8	1	1	8	18
3.1	18	8	1	1	8	-144	-8	8	8	-8	468	-144	18	18	-144	-144	-8	8	8	-8	18	8	1	1	8
3.2	8	18	8	1	1	-8	-144	-8	8	8	-144	468	-144	18	18	-8	-144	-8	8	8	8	18	8	1	1
3.3	1	8	18	8	1	8	-8	-144	-8	8	18	-144	468	-144	18	8	-8	-144	-8	8	1	8	18	8	1
3.4	1	1	8	18	8	8	8	-8	-144	-8	18	18	-144	468	-144	8	8	-8	-144	-8	1	1	8	18	8
3.5	8	1	1	8	18	-8	8	8	-8	-144	-144	18	18	-144	468	-8	8	8	-8	-144	8	1	1	8	18
4.1	18	8	1	1	8	18	8	1	1	8	-144	-8	8	8	-8	468	-144	18	18	-144	-144	-8	8	8	-8
4.2	8	18	8	1	1	8	18	8	1	1	-8	-144	-8	8	8	-144	468	-144	18	18	-8	-144	-8	8	8
4.3	1	8	18	8	1	1	8	18	8	1	8	-8	-144	-8	8	18	-144	468	-144	18	8	-8	-144	-8	8
4.4	1	1	8	18	8	1	1	8	18	8	8	8	-8	-144	-8	18	18	-144	468	-144	8	8	-8	-144	-8
4.5	8	1	1	8	18	8	1	1	8	18	-8	8	8	-8	-144	-144	18	18	-144	468	-8	8	8	-8	-144
5.1	-144	-8	8	8	-8	18	8	1	1	8	18	8	1	1	8	-144	-8	8	8	-8	468	-144	18	18	-144
5.2	-8	-144	-8	8	8	8	18	8	1	1	8	18	8	1	1	-8	-144	-8	8	8	-144	468	-144	18	18
5.3	8	-8	-144	-8	8	1	8	18	8	1	1	8	18	8	1	8	-8	-144	-8	8	18	-144	468	-144	18
5.4	8	8	-8	-144	-8	1	1	8	18	8	1	1	8	18	8	8	8	-8	-144	-8	18	18	-144	468	-144
5.5	-8	8	8	-8	-144	8	1	1	8	18	8	1	1	8	18	-8	8	8	-8	-144	-144	18	18	-144	468

FIGURE 2.12: Structure matrix for 5 years of SBP and DBP combinations, $\mathbf{P} = \mathbf{D}^T\mathbf{D}$

does not seem desirable for problems which do not have a circular behaviour or specifically when we want to compute relations in temporal behaviour. The increments that are used are: $(\Delta^2_{(1,0)}u_{[i,j]} + \Delta^2_{(0,1)}u_{[i,j]})$ and the precision matrix consists of: $-(\Delta^2_{(1,0)}u_{[i,j]} + \Delta^2_{(0,1)}u_{[i,j]})^2 = -\Delta^4_{(1,0)}u_{[i,j]} - \Delta^4_{(0,1)}u_{[i,j]} - \Delta^2_{(0,1)}\Delta^2_{(1,0)}u_{[i,j]}$ Finally, the middle row of the precision matrix $\mathbf{P}$ is:

$$
\begin{aligned}
&u_{[i-2,j-2]} + 8u_{[i-2,j-1]} + 18u_{[i-2,j]} + 8u_{[i-2,j+1]} + u_{[i-2,j+2]} + 8u_{[i-1,j-2]} - 8u_{[i-1,j-1]} \\
&- 144u_{[i-1,j]} - 8u_{[i-1,j+1]} + 8u_{[i-1,j+2]} + 18u_{[i,j-2]} - 144u_{[i,j-1]} + 468u_{[i,j]} \\
&- 144u_{[i,j+1]} + 18u_{[i,j+2]} + 8u_{[i+1,j-2]} - 8u_{[i+1,j-1]} - 144u_{[i+1,j]} - 8u_{[i+1,j+1]} \\
&+ 8u_{[i+1,j+2]} + u_{[i+2,j-2]} + 8u_{[i+2,j-1]} + 18u_{[i+2,j]} + 8u_{[i+2,j+1]} + u_{[i+2,j+2]}
\end{aligned}
\tag{2.29}
$$

In this type of matrix the linear restriction will be the same as in Section 2.6.2, $T^2 - 3$ having also the same rank deficiency: $T^2 - 1$.

2.6.4 Precision Matrix with boundary constraints, part 1

The third suggestion of a structure matrix is presented without having a precision matrix on a torus by Terzopoulos, Terzopoulos [1988]. With torus structures we can capture spatial dependences whereas the suggestion with boundary constraints captures temporal dependencies. We compute the block matrices in order to satisfy the condition of summing-to-zero in every row and every column of the precision matrix, $\lambda\mathbf{P}$, and taking into account the closest neighbouring nodes every time. Having said that though, the rows and columns of each of the block matrices does not sum to zero. The rows and columns of the precision matrix sum to zero instead. In the first five rows we will use the sub-matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$:

$$\mathbf{A}_1 = \begin{pmatrix} 4 & -4 & 1 & & \\ -4 & 10 & -6 & 1 & \\ 1 & -6 & 11 & -6 & 1 \\ & 1 & -6 & 10 & -4 \\ & & 1 & -4 & 4 \end{pmatrix} \mathbf{A}_2 = \begin{pmatrix} -4 & 2 & & & \\ 2 & -6 & 2 & & \\ & 2 & -6 & 2 & \\ & & 2 & -6 & 2 \\ & & & 2 & -4 \end{pmatrix}$$

$$\mathbf{A}_3 = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}$$

In the sixth to tenth rows we will use the sub-matrices $\mathbf{A}_4$, $\mathbf{A}_5$ and the $\mathbf{A}_2$, $\mathbf{A}_3$ sub-matrices will be used again:

$$\mathbf{A}_4 = \begin{pmatrix} 10 & -6 & 1 & & \\ -6 & 18 & -8 & 1 & \\ 1 & -8 & 19 & -8 & 1 \\ & 1 & -8 & 18 & -6 \\ & & 1 & -6 & 10 \end{pmatrix} \mathbf{A}_5 = \begin{pmatrix} -6 & 2 & & & \\ 2 & -8 & 2 & & \\ & 2 & -8 & 2 & \\ & & 2 & -8 & 2 \\ & & & 2 & -6 \end{pmatrix}$$

$$\mathbf{A}_6 = \begin{pmatrix} 11 & -6 & 1 & & \\ -6 & 19 & -8 & 1 & \\ 1 & -8 & 20 & -8 & 1 \\ & 1 & -8 & 19 & -6 \\ & & 1 & -6 & 11 \end{pmatrix}$$

This time the sub-matrices express the structure matrix $\mathbf{P}$

$$P = \lambda \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & \mathbf{0} \\ \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_6 & \mathbf{A}_5 & \mathbf{A}_3 \\ \mathbf{0} & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_2 \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.30}$$

Specifically, observing the matrices $\mathbf{A}_4$ and $\mathbf{A}_6$ we can see that the restrictions are in the first two (and the last two) rows giving the impression of being a second-order IGMRF. Similarly, the matrix $\mathbf{A}_5$ has a different behaviour in the first (and last) row giving the impression of being a first-order IGMRF. However, since the rows and the columns do not sum to zero they cannot be considered as IGMRFs. Finally, we can see that the structure of the matrix in (2.30) looks similar to the second-order IGMRF, presented in Section 2.5.2. The constraints are based on the two rows at the boundary but now the two rows consist of sub-matrices. Each row and column sums to zero therefore, $\lambda\mathbf{P}$ can be the precision matrix of an IGMRF. Furthermore, we can see that the block matrix $\mathbf{A}_6$ in (2.30) is in the diagonal

1.1	4	-4	1	.	.	-4	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1.2	-4	10	-6	1	.	2	-6	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
1.3	1	-6	11	-6	1	.	2	-6	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
1.4	.	1	-6	10	-4	.	.	2	-6	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
1.5	.	.	1	-4	4	.	.	.	2	-4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
2.1	-4	2	.	.	.	10	-6	1	.	.	-6	2	.	.	.	1	.	.	.	.	.	.	.	.	.
2.2	2	-6	2	.	.	-6	18	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	.	.	.	.
2.3	.	2	-6	2	.	1	-8	19	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	.	.	.
2.4	.	.	2	-6	2	.	1	-8	18	-6	.	.	2	-8	2	.	.	.	1	.	.	.	.	.	.
2.5	.	.	.	2	-4	.	.	1	-6	10	.	.	.	2	-6	.	.	.	.	1	.	.	.	.	.
3.1	1	.	.	.	.	-6	2	.	.	.	11	-6	1	.	.	-6	2	.	.	.	1	.	.	.	.
3.2	.	1	.	.	.	2	-8	2	.	.	-6	19	-8	1	.	2	-8	2	.	.	.	1	.	.	.
3.3	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.
3.4	.	.	.	1	.	.	.	2	-8	2	.	1	-8	19	-6	.	.	2	-8	2	.	.	.	1	.
3.5	.	.	.	.	1	.	.	.	2	-6	.	.	1	-6	11	.	.	.	2	-6	.	.	.	.	1
4.1	.	.	.	.	.	1	.	.	.	.	-6	2	.	.	.	10	-6	1	.	.	-4	2	.	.	.
4.2	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	.	-6	18	-8	1	.	2	-6	2	.	.
4.3	.	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	19	-8	1	.	2	-6	2	.
4.4	.	.	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	18	-6	.	.	2	-6	2
4.5	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-6	.	.	1	-6	10	.	.	.	2	-4
5.1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-4	2	.	.	.	4	-4	1	.	.
5.2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-6	2	.	.	-4	10	-6	1	.
5.3	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-6	2	.	1	-6	11	-6	1
5.4	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-6	2	.	1	-6	10	-4
5.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-4	.	.	1	-4	4

FIGURE 2.13: Structure matrix for 5 years of SBP and DBP combinations, $\mathbf{P} = \mathbf{D}^T\mathbf{D}$

only in the third row whereas in the structure matrix in (2.31), Section 2.6.5 the central block matrix $\mathbf{A}_4$ is used as the diagonal in the second, third and fourth rows. That is an indication of why the rank in Section 2.6.5 is $T^2 - 1$ while the rank in Section 2.6.4 is $T^2 - 3$.

Summing up, the increments for the two-dimensional scenario are: $(\Delta^2_{(1,0)} u_{[i,j]} + \Delta^2_{(0,1)} u_{[i,j]})$ and hence, the $\mathbf{P}$'s central row consists of: $-(\Delta^2_{(1,0)} u_{[i,j]} + \Delta^2_{(0,1)} u_{[i,j]})^2 = -\Delta^4_{(1,0)} u_{[i,j]} - \Delta^4_{(0,1)} u_{[i,j]} - \Delta^2_{(0,1)} \Delta^2_{(1,0)} u_{[i,j]}$.
Therefore, the middle row of the precision matrix is: $u_{[i-2,j]} + 2u_{[i-1,j-1]} - 8u_{[i-1,j]} + 2u_{[i-1,j+1]} + u_{[i,j-2]} - 8u_{[i,j-1]} + 20u_{[i,j]} - 8u_{[i,j+1]} + u_{[i,j+2]} + 2u_{[i+1,j-1]} - 8u_{[i+1,j]} + 2u_{[i+1,j+1]} + u_{[i+2,j]}$.
Again, the linear restriction is: $T^2 - 3$ as in the previous two cases whereas the rank deficiency is $T^2 - 3$ which is the first time that coincides with the linear restrictions.

2.6.5 Precision Matrix with boundary constraints, part 2

A fourth suggestion of a structure matrix and a second without being on a torus is presented by Yue and Speckman Yue and Speckman [2010]. The structure of the precision matrix is similar to the one in Section 2.6.4 but with different block matrices which means different weights for the neighbours which we will see affects the rank deficiency. The intuition behind the choice of the matrices with boundary constraints is that we can choose how to define the matrix's boundaries. Depending on the weight that we give to the central rows of each of the block matrices we will have a different effect at the boundaries we consider. The matrix presented in Section 2.6.4 gives only to one of the nodes in the diagonal the maximum weight resulting in more boundary constraints whereas the structure matrix in Section 2.6.5 gives to more nodes the maximum, and thus same, weight which results in fewer boundary constraints. Hence, the rank deficiency of the matrices is affected by the weights we give to the nodes in the diagonal and that is why we have different rank deficiency between these two matrices.

In the first five rows we will use the sub-matrices $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$:

$$\mathbf{A}_1 = \begin{pmatrix} 6 & -5 & 1 & & \\ -5 & 12 & -6 & 1 & \\ 1 & -6 & 12 & -6 & 1 \\ & 1 & -6 & 12 & -5 \\ & & 1 & -5 & 6 \end{pmatrix} \mathbf{A}_2 = \begin{pmatrix} -5 & 2 & & & \\ 2 & -7 & 2 & & \\ & 2 & -7 & 2 & \\ & & 2 & -7 & 2 \\ & & & 2 & -5 \end{pmatrix}$$

$$\mathbf{A}_3 = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}$$

In the sixth to tenth rows we will use the sub-matrices $\mathbf{A}_4$, $\mathbf{A}_5$ and the $\mathbf{A}_2$, $\mathbf{A}_3$ sub-matrices will be used again:

$$\mathbf{A}_4 = \begin{pmatrix} 12 & -7 & 1 & & \\ -7 & 20 & -8 & 1 & \\ 1 & -8 & 20 & -8 & 1 \\ & 1 & -8 & 20 & -7 \\ & & 1 & -7 & 12 \end{pmatrix} \mathbf{A}_5 = \begin{pmatrix} -6 & 2 & & & \\ 2 & -8 & 2 & & \\ & 2 & -8 & 2 & \\ & & 2 & -8 & 2 \\ & & & 2 & -6 \end{pmatrix}$$

In the next rows these four sub-matrices will be used again. Having constructed the block matrices we can see the structure of the precision matrix, $\mathbf{Q}$:

$$\lambda \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & & \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & \\ \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 \\ & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_2 \\ & & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.31}$$

Observing the precision matrix in (2.31) we can see that the two first rows are those with the different behaviour. So, one could possibly say that we have T^2-2 constraints. However, each row contains block matrices which means that each of the block matrices present a conditional dependence and hence gives zero values to the combination going further away from the under study combination. The linear constraints are T^2-3 as in the previous cases. Furthermore, we can also observe the structure of the block matrices. The $\mathbf{A}_2$ and $\mathbf{A}_5$ matrices are similar to those for the first order random walk (2.5.1) since only the first row is changing. The $\mathbf{A}_1$ and $\mathbf{A}_4$ matrices are similar to those for the second order random walk (2.5.2) since only the first two rows are changing. However, none of the $\mathbf{A}_1$, $\mathbf{A}_2$, $\mathbf{A}_4$ and $\mathbf{A}_5$ are structure matrices of IGMRF since their rows and their columns do not sum to zero.

In the first order random walk (Section 2.5.1) we could see that we set the first (and the last) row of the matrix to have the restriction that matches the $T-1$ constraints. In the second order random walk (Section 2.5.2) we set the two first rows (and the last two) to have a different behaviour, so we had $T-2$ constraints. In the second order two-dimensional model the behaviour is slightly different. The precision matrix is now a block-circulant matrix hence, the restrictions will not be in two rows but in three block matrices. The following is an intuitive explanation of the T^2-3 constraints. Let us assume that the years we have are $T=5$ therefore, we will construct a precision matrix with $5^2 \times 5^2$ dimensions. We compute the block matrices in order to satisfy the condition of summing-to-zero in every row and every column of the precision matrix which is the first of the linear restrictions that are applied in all of the scenarios. The precision matrix $\mathbf{P}$ is defined as, Yue and Speckman [2010]:

$$\begin{aligned}&\sum_{i=2}^{T-1}\sum_{j=2}^{T-1}\{\Delta_0^2 u_{[i,j]}\}^2 + \{\Delta_1 u_{[1,1]}\}^2 + \{\Delta_2 u_{[T,1]}\}^2 + \{\Delta_3 u_{[1,T]}\}^2 + \{\Delta_4 u_{[T,T]}\}^2 \\ &+ \sum_{i=2}^{T}(\{\Delta_5 u_{[i,1]}\}^2 + \{\Delta_6 u_{[i,T]}\}^2) + \sum_{j=2}^{T}(\{\Delta_7 u_{[1,j]})\}^2 + \{\Delta_8 u_{[T,j]}\}^2)\end{aligned} \tag{2.32}$$

In more detail each of the forward differences can be written as:

$$\begin{aligned}\Delta_0^2 u_{[i,j]} &= (\Delta^2_{(1,0)} + \Delta^2_{(0,1)})u_{[i+1,j+1]} = \Delta^2_{(1,0)}u_{[i+1,j+1]} + \Delta^2_{(0,1)}u_{[i+1,j+1]} \\ (\Delta_0^2 u_{[i,j]})^2 &= (\Delta^2_{(1,0)} + \Delta^2_{(0,1)})^2 u_{[i+3,j+3]} = (\Delta^4_{(1,0)} + \Delta^4_{(0,1)} + 2\Delta^2_{(1,0)}\Delta^2_{(0,1)})u_{[i+4,j+4]}\end{aligned} \tag{2.33}$$

$$\begin{aligned}\{\Delta_1 u_{[1,1]}\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 u_{[3,3]} \\ \{\Delta_2 u_{[T,1]}\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 u_{[T,3]} \\ \{\Delta_3 u_{[1,T]}\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 u_{[3,T]} \\ \{\Delta_4 u_{[T,T]}\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 u_{[T,T]} \\ \{\Delta_5 u_{[i,1]}\}^2 &= (\Delta^2_{(1,0)} + \Delta_{(0,1)})^2 u_{[i+3,3]} \\ \{\Delta_6 u_{[i,T]}\}^2 &= (\Delta^2_{(1,0)} + \Delta_{(0,1)})^2 u_{[i+3,T]} \\ \{\Delta_7 u_{[1,j]}\}^2 &= (\Delta_{(1,0)} + \Delta^2_{(0,1)})^2 u_{[3,j+3]} \\ \{\Delta_8 u_{[T,j]}\}^2 &= (\Delta_{(1,0)} + \Delta^2_{(0,1)})^2 u_{[T,j+3]}\end{aligned} \tag{2.34}$$

The signs $(+,-)$ used in Yue and Speckman [2010] to refer to the direction of the nodes in the forward differences are omitted. Having this density we account for all the 25 different combinations. The first summation is responsible for the central values of the precision matrix: $(T-1-1)*(T-1-1) = 9$.

The nodes at the boundary in analytical form are:

- Four corners: $(_{[1,1]}, _{[1,5]}, _{[5,1]}, _{[5,5]})$.
- Boundaries on the left and right sides: $3*2 = 6$
- Boundaries at the top and bottom sides: $3*2 = 6$

1.1	6	-5	1	.	.	-5	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1.2	-5	12	-6	1	.	2	-7	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
1.3	1	-6	12	-6	1	.	2	-7	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
1.4	.	1	-6	12	-5	.	.	2	-7	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
1.5	.	.	1	-5	6	.	.	.	2	-5	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
2.1	-5	2	.	.	.	12	-7	1	.	.	-6	2	.	.	.	1	.	.	.	.	.	.	.	.	.
2.2	2	-7	2	.	.	-7	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	.	.	.	.
2.3	.	2	-7	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	.	.	.
2.4	.	.	2	-7	2	.	1	-8	20	-7	.	.	2	-8	2	.	.	.	1	.	.	.	.	.	.
2.5	.	.	.	2	-5	.	.	1	-7	12	.	.	.	2	-6	.	.	.	.	1	.	.	.	.	.
3.1	1	.	.	.	.	-6	2	.	.	.	12	-7	1	.	.	-6	2	.	.	.	1	.	.	.	.
3.2	.	1	.	.	.	2	-8	2	.	.	-7	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.
3.3	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.
3.4	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-7	.	.	2	-8	2	.	.	.	1	.
3.5	.	.	.	.	1	.	.	.	2	-6	.	.	1	-7	12	.	.	.	2	-6	.	.	.	.	1
4.1	.	.	.	.	.	1	.	.	.	.	-6	2	.	.	.	12	-7	1	.	.	-5	2	.	.	.
4.2	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	.	-7	20	-8	1	.	2	-7	2	.	.
4.3	.	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-7	2	.
4.4	.	.	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-7	.	.	2	-7	2
4.5	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-6	.	.	1	-7	12	.	.	.	2	-5
5.1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-5	2	.	.	.	6	-5	1	.	.
5.2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-7	2	.	.	-5	12	-6	1	.
5.3	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-7	2	.	1	-6	12	-6	1
5.4	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-7	2	.	1	-6	12	-5
5.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-5	.	.	1	-5	6

FIGURE 2.14: Precision matrix for 5 years of SBP and DBP combinations

Hence, the nodes are separated as: a) $\underline{(5-2)^2}$ = 9 <u>central</u> nodes and b) 4+6+6 = 16 nodes in the boundary. So, for a number of nodes $T = 5$ in each variable we have: $(5-2)^2 + 4 + (5-2) * 2 + (5-2) * 2 = 25$

In general, what are considered as central or boundary nodes can be summarised as the following: $\underline{(n_1 - 2) * (n_2 - 2)} + 4 + (n_1 - 2) * 2 + (n_2 - 2) * 2$,

In the special case when $n_1 = n_2 = T$: $\underline{(T-2)^2} + 4 + (T-2) * 2 + (T-2) * 2$.

The middle row of the precision matrix $\mathbf{P}$ is equal to:

$$\begin{aligned} &u_{[i-2,j]} + 2u_{[i-1,j-1]} - 8u_{[i-1,j]} + 2u_{[i-1,j+1]} + u_{[i,j-2]} - 8u_{[i,j-1]} + 20u_{[i,j]} \\ &\quad - 8u_{[i,j+1]} + u_{[i,j+2]} + 2u_{[i+1,j-1]} - 8u_{[i+1,j]} + 2u_{[i+1,j+1]} + u_{[i+2,j]} \end{aligned} \tag{2.35}$$

In Figure 2.14 we can see that we need three rows with block matrices in order to reach the 13^{th} line which has the maximum number of neighbours. The central row which is the one that is invariable to the boundary effects has 13 non-zero elements. In Figure 2.14 we can see the 25×25 precision matrix. The diagonal elements in $\lambda\mathbf{P}$ represent the conditional precision of DBP and SBP in the same year i given the rest of the years, while the off-diagonal elements with proper scaling show the conditional correlation between DBP and SBP in years i and j respectively given the rest of the years, Rue and Held [2005]. In Figure 2.14 the total number of nodes for each variable is five.

It would also be useful to see which and how many neighbours there are at the boundaries. Figures 2.15 and 2.16 show us the behaviour at the boundaries and

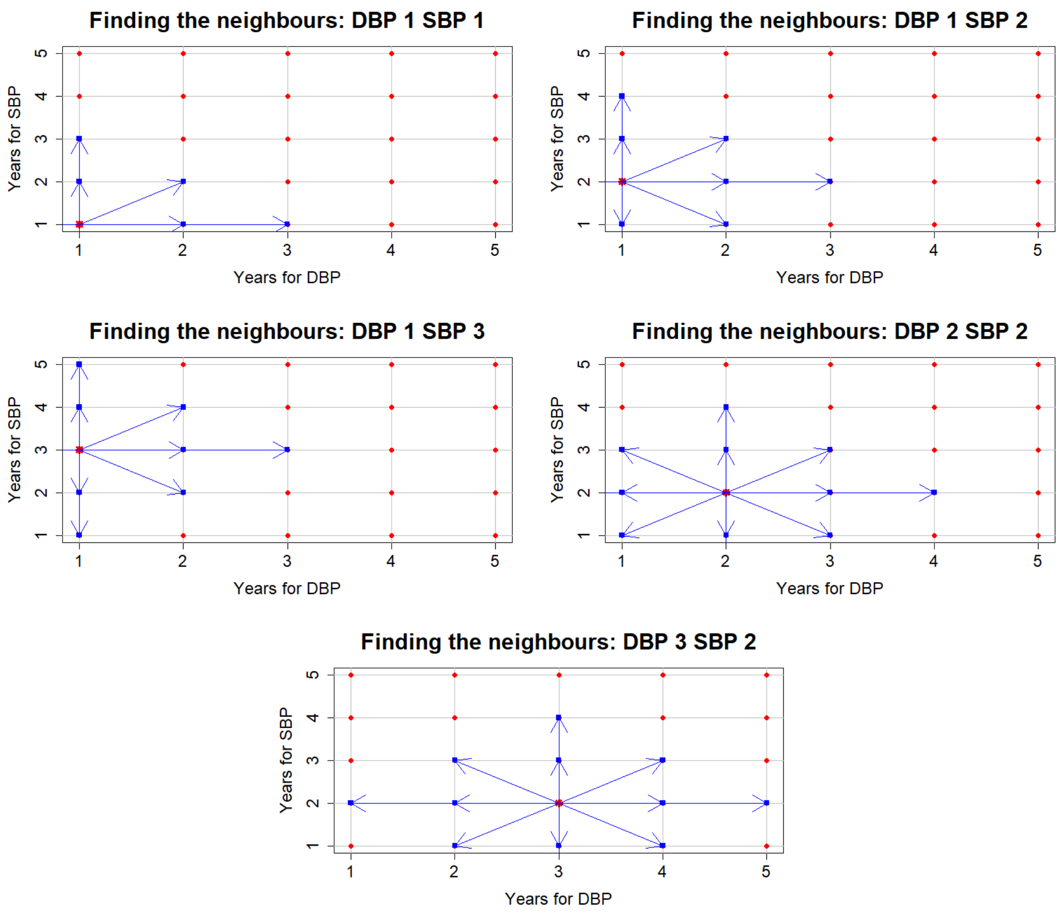

FIGURE 2.15: The neighbouring years at the boundaries for 5 years of DBP and SBP

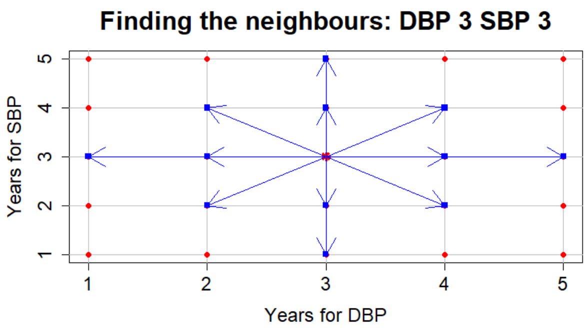

FIGURE 2.16: The neighbouring years in the centre for 5 years of DBP and SBP

in the centre. Depending on the combination of nodes the number of neighbours changes. The diagonal direction is considered as a two step distance as it is a combination of one step vertically and then one step horizontally. In general, the numbers of possible neighbours are either 6, 8, 9, 11 or 12 at the boundaries and 13 in the central years. To summarise, we have that the structure matrix $\mathbf{P}$ consists of: $-(\Delta^2_{(1,0)}u_{[i,j]} + \Delta^2_{(0,1)}u_{[i,j]})^2 = -\Delta^4_{(1,0)}u_{[i,j]} - \Delta^4_{(0,1)}u_{[i,j]} - \Delta^2_{(0,1)}\Delta^2_{(1,0)}u_{[i,j]}$

Again, the middle row of the precision matrix is:

$x_{[i,j-2]} + 2x_{[i-1,j-1]} + x_{[i-2,j]} - 8x_{[i,j-1]} - 8x_{[i-1,j]} + 2x_{[i+1,j-1]} + 20x_{[i,j]} + 2x_{[i-1,j+1]} - 8x_{[i+1,j]} - 8x_{[i,j+1]} + x_{[i+2,j]} + 2x_{[i+1,j+1]} + x_{[i,j+2]}$

Finally the linear restrictions are $T^2 - 3$ with rank deficiency $T^2 - 1$.

The **P** matrix shows all the combinations between the two variables. In the **P** matrix when the row is $_{[3,3]}$ it means that both variables are in position 3. There are 25 possible combinations but only the nearest to the $_{[3,3]}$ pair in this case will be non-zero including the $_{[3,3]}$ pair. In this case the combinations $_{[2,3]}$, $_{[3,2]}$, $_{[3,3]}$, $_{[3,4]}$, $_{[4,3]}$ have non-zero values. If we observe closely these values we can see that the $_{[2,3]}$ pair is a backward step for the first variable while the second variable remains at the same node. By the same token the $_{[3,2]}$ pair is a backward step for the second variable while the first variable remains at the same node. The $_{[3,4]}$, $_{[4,3]}$ pairs express a forward step for the second variable while the first remains at the same node and accordingly a forward step for the first variable while the second variable remains at the same node.

In the precision matrix in Figure 2.14 we can see that in the $_{[3,3]}$ pair we have the maximum number of neighbours. The $_{[3,3]}$ pair, the pair under study, takes the highest weight, 20. The $_{[2,3]}$, $_{[3,2]}$, $_{[4,3]}$, $_{[3,4]}$ pairs are the closest neighbours, one step distance, therefore, will get a weight of -8. The $_{[2,2]}$, $_{[2,4]}$, $_{[4,2]}$, $_{[4,4]}$ combinations are referring to the interactions of the two variables and correspond to $u_{[i-1,j-1]}, u_{[i-1,j+1]}, u_{[i+1,j-1]}, u_{[i+1,j+1]}$ values. These four interaction terms have the number 2 as weight. Lastly, the final four pairs, $_{[1,3]}$, $_{[3,1]}$, $_{[3,5]}$, $_{[5,3]}$, correspond to the neighbours that are 2 steps away from the $_{[3,3]}$ pair. These values are more distant than the rest of the neighbours so they take the smaller weight which is equal to one. At this point we would like to highlight that the signs in front of the weights are used for satisfying the linear constraints. The absolute value of the weights shows which of the combinations are closer to the pair under study. In summary, based on (2.35), we can notice that considering the $u_{i,j}$ the central point, the points that are closer to it have a bigger weight, $8u_{i,j-1}, 8u_{i-1,j}, 8u_{i+1,j}, 8u_{i,j+1}$, later the interaction terms follow $2u_{i-1,j-1}, 2u_{i+1,j-1}, 2u_{i-1,j+1}, 2u_{i+1,j+1}$ and finally the most distanced are the neighbours with a distance of 2 nodes: $u_{i,j-2}, u_{i-2,j}, u_{i+2,j}, u_{i,j+2}$.

Yue and Speckman Yue and Speckman [2010] also highlighted the need for deleting the first row in order to make the matrix full rank which later helps in a non-stationary matrix. Here, we will use the initial matrix (2.31) taking advantage of the stationary matrix and its non-full rank form.

2.6.5.1 More than 5 years for the second order 2D IGMRF

In the following three matrices we can see how the structure matrix, $\mathbf{P}$ changes depending on the number of years for each variable. These will help us understand how the linear constraints apply in the two dimensional case.

$$\lambda\mathbf{P} = \lambda \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & & \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & \\ \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 \\ & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_2 \\ & & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.36}$$

$$\lambda\mathbf{P} = \lambda \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & & & \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & & \\ \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & \\ & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 \\ & & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_2 \\ & & & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.37}$$

$$\lambda \mathbf{P} = \lambda \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & & & & \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & & & \\ \color{red}{\mathbf{A}_3} & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & & \\ & \color{red}{\mathbf{A}_3} & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & \\ & & \color{red}{\mathbf{A}_3} & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 \\ & & & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_2 \\ & & & & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{2.38}$$

Taking into account five nodes for each variable, $x_{i,j}$ we have a precision matrix of 25×25 dimensions and we can see that the central row is the one indicated with red colour. In (2.36) we have presented only the block matrices. When we have data for six nodes for each variable we have a precision matrix of 36×36 dimensions. In (2.37) we can see only the block matrices which can help us to notice that the central rows are the two in the middle now. Finally, having data for seven years for each variable will give us a precision matrix of 49×49 dimensions. In (2.38) we can see that the central rows are three in the middle which are denoted with red colour too. The dimensions of the block matrices are changing too. We will have 5×5, 6×6 and 7×7 respectively increasing each time the central row once, twice for the six years and three for the seven years respectively for each one of the block matrices. In conclusion, in the three cases that are shown above the boundary constraints are in the first two and last two rows. However, these rows are not numbers but block-matrices.

2.7 Summary

This chapter provides an overview of all the methods that this thesis will make use of. In Section 2.1 we see the definition of GMRFs while in Section 2.2 we analyse a case of GMRFs with linear constraints, the IGMRFs. GMRFs and IGMRFs are commonly used as prior distributions to explain dependence through the prior's

Table 2.1: Constraints for the one-dimensional 1^{st} and 2^{nd} order and for the two-dimensional 2^{nd} order IGMRF.

	1RW	2RW	RW2D
Nodes	T	T	T^2
Constraints	1	2	3
Total	$T-1$	$T-2$	T^2-3

Table 2.2: Constraints for the four different types of a two-dimensional 2^{nd} order IGMRF.

	$Torus_1$	$Torus_2$	$Bounds_1$	$Bounds_2$
Nodes	T^2	T^2	T^2	T^2
Rank Deficiency	T^2-1	T^2-1	T^2-1	T^2-3
Linear Restrictions	T^2-3	T^2-3	T^2-3	T^2-3

precision matrices. Throughout this thesis we will focus on the IGMRFs properties which are very useful for conducting the research. In Section 2.3 we examine different types of precision matrices which can be used as prior distributions satisfying the IGMRFs conditions. With the aid of canonical parametrization, Section 2.4, we are able to overcome computational difficulties of not finding the exact form of the mean parameter introducing an alternative way of using the normal distribution. Using Cholesky decomposition we manage to achieve fast calculations for the posterior distributions tackling memory issues by storing large matrices. Furthermore, in Section 2.5 we see the one-dimensional random walk which expresses the dependences between nodes. Depending on the order of the random walk, first or second order, we change the number of neighbours that we want to be dependent each time, as shown in Sections 2.5.1 and 2.5.2 whereas a comparison between these two is shown in Table 2.1. Finally, in Section 2.6 we can see the extended case of the random walk in two dimensions. Through this chapter we are able to see different types of two-dimensional random walk where each one of them captures a different relation through the nodes. Analytically, for capturing the spatial dependence between two variables, precision matrices on a torus appear more suitable, as seen in Sections 2.6.2 and 2.6.3. On the other hand, for capturing temporal dependence we have dependences on a line and we also want to include different behaviour in each node as we might not have the same information in every node under study. Suitable precision matrices are those with boundary constraints which are presented in Section 2.6.4 and 2.6.5. A comparison between the four different two-dimensional precision matrices is shown in Table 2.2. Finally, in the following chapters we will explain in detail one-dimensional and two-dimensional models making use of the one-dimensional and two-dimensional IGMRF as prior distributions.

Chapter 3

Flexible Two-Dimensional Bayesian Hierarchical Models

3.1 Introduction

Many statistical applications involve multiple parameters that can be considered as related or connected in some way by the structure of the problem Gelman et al. [2013]. Therefore, a joint probability model for these parameters can be considered appropriate in order to account for their dependence. This is achieved in a natural way if we use a prior distribution in which the group of parameters in the joint distribution is viewed as a sample from a common population distribution as explained thoroughly by Gelman Gelman et al. [2013]. Moreover, hierarchical models can even have more parameters than data points without causing overfitting by the fact that they use a population distribution to structure some dependence into the parameters.

As for Bayesian hierarchical models we need a joint probability model, it is important to comply with the requirements of exchangeability Damien et al. [2013], Gelman et al. [2013], Griffin and Brown [2017]. Exchangeability is present when the probability of a joint distribution is equal to the probability of a joint distribution having changed the order of the parameters that exist in the joint distribution,

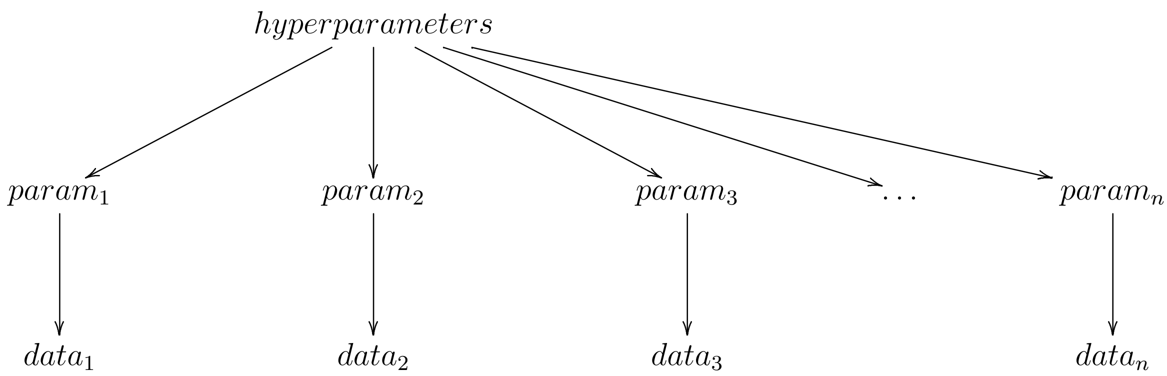

FIGURE 3.1: Structure of the hierarchical model

i.e., having permuted the parameters Damien et al. [2013]. Therefore, we have $P(X_{\pi(1)}, X_{\pi(2)}, \ldots, X_{\pi(n)}) = P(X_1, X_2, \ldots, X_n)$. However, as long as we have no information that contradicts the assumption of exchangeability, this is a logical assumption Gelman et al. [2013].

By assuming exchangeability we imply dependence between the parameters which form the joint distribution. Hence, hierarchical models provide a dependence between the parameters by expressing them in terms of the likelihood and the priors Gelman et al. [2013].

Hierarchical models can be separated into three stages. The first stage declares the distribution of the data and likelihood. The second stage accounts for the prior distribution of the parameters and in the third stage the distributions of the hyperpriors appear Damien et al. [2013]. These stages are illustrated in Figure 3.1. We present an example in (3.1) of how each of the stages could be written using distributions which is also presented in Figure 3.2.

$$\begin{aligned} \text{First stage: [data—process, parameters]} &\Rightarrow Y|X, \boldsymbol{\theta} \sim \mathcal{N}(X\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \\ \text{Second stage: [process—parameters]} &\Rightarrow \boldsymbol{\theta}|Z, \boldsymbol{\eta} \sim \mathcal{N}(Z_{\boldsymbol{\eta}}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}) \\ \text{Third stage: [(hyper)parameters]} &\Rightarrow \boldsymbol{\eta} \sim \mathcal{N}(\boldsymbol{\eta}_0, \boldsymbol{\Sigma}_{\boldsymbol{\eta}}) \end{aligned} \tag{3.1}$$

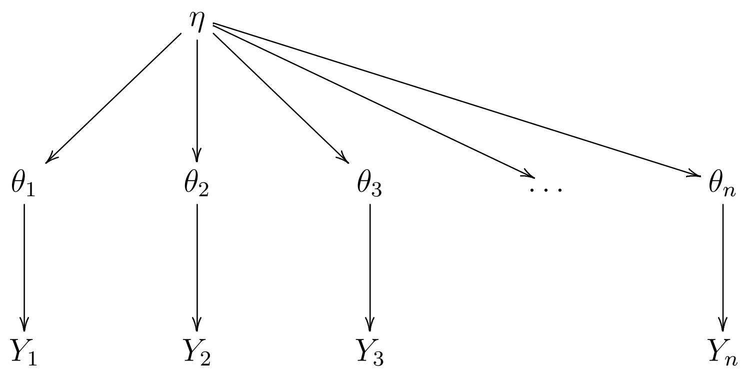

FIGURE 3.2: Example of the hierarchical model's structure

It is assumed that we have exchangeable $\boldsymbol{\theta}_i$. If $\boldsymbol{\eta}$ is fixed then we have i different/separate models. However, more often that is not the case and with unknown $\boldsymbol{\eta}$ we observe the phenomenon of shrinkage or borrowing strength across the i's. By using this term we imply that the parameter under study, $\boldsymbol{\theta}_i$ will not be based solely on the data's information, $\mathbf{Y}_i$ but also from the prior that we set on $\boldsymbol{\eta}$. That being said, we expect that the posterior will shrink from the data's distribution which is a sample estimate towards the prior's distribution which is the population distribution, the so-called shrinkage effect, as expressed in Gelman et al. [2013] and Damien et al. [2013].

In this chapter we describe two Bayesian hierarchical models, a one-dimensional and a two-dimensional model of blood pressure. The dependent variable for the one-dimensional model is SBP whereas for the two-dimensional equivalent the dependent variables are DBP and SBP taking into account the interaction(INT) between DBP and SBP as well. Both of the models have the same parameters to define the likelihood but have changes in the dimensions. In the following sections we will describe the sub-models that comprise the one- and two-dimensional models.

3.2 Models

3.2.1 The one-dimensional Model

The variable that we are interested in exploring is SBP, $\mathbf{y}$. The model has been previously defined and used by Danaei et al and Finucane et al in Danaei et al. [2011] and Finucane et al. [2011]. The likelihood for an observation of SBP in age group h from study i conducted in country j is, Danaei et al. [2011], Finucane et al. [2011]:

$$y_{h,i} = a_{j[i]} + b_{j[i]}t_i + u_{j[i],t_i} + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\gamma}_i(\mathbf{z}_h) + e_i + \epsilon_{h,i} \tag{3.2}$$

$$y_{h,i} \sim \mathcal{N}(a_{j[i]} + b_{j[i]}t_i + u_{j[i],t_i} + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\gamma}_i(\mathbf{z}_h) + e_i, SD^2_{h,i}/n_{h,i} + \tau_i^2) \tag{3.3}$$

The general model of blood pressure can be separated into five different sub-models, Danaei et al. [2011], Finucane et al. [2011]:

- Linear change over time with intercept and trend, a_j, b_j
- Non-linear change over time, u_j
- Covariate effects, $\boldsymbol{\beta}$
- Age model, $\boldsymbol{\gamma}_i(\mathbf{z}_h)$
- Study-specific effects, e_i:
  measure the difference between the average blood pressure measurement in study i and the average blood pressure measurement in the entire dataset. An intercept that changes from study to study which also has its own prior distribution.
- Error term/study-age specific random effects, $\epsilon_{h,i}$:
  follow a Gaussian distribution with zero mean and variance $SD^2_{h,i}/n_{h,i}$. $SD^2_{h,i}$ is the standard deviation of blood pressure values observed in the h age group in study i while $n_{h,i}$ expresses the sample size of individuals that each

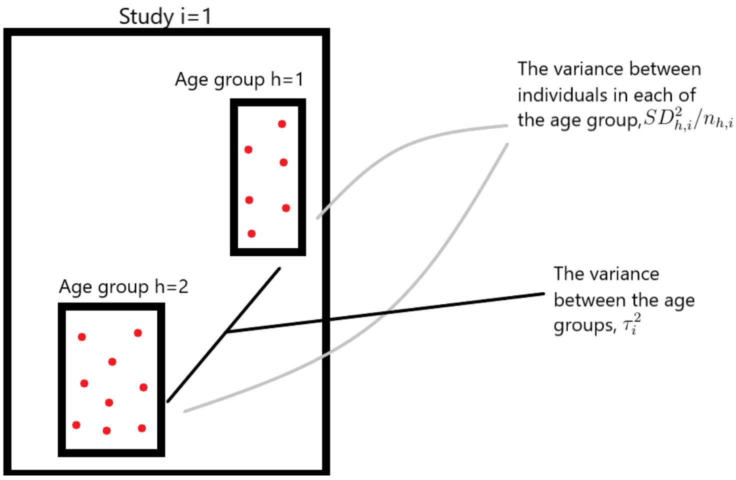

FIGURE 3.3: Interpretation of the variances in study i

h age group has in the study i. The $y_{h,i}$ terms express the mean values of a specific h age group and study i and that is why we have $SD^2_{h,i}/n_{h,i}$ and not solely $SD^2_{h,i}$. Each study has several age groups and each age group consists of a number of individuals. Here, $SD^2_{h,i}/n_{h,i}$ is the variance between the individuals in the same h age group of the same study i. In other words, the variance *within* a particular h age group in a specific study i. Also through the residuals we account for the study-specific deviation which is the deviation of blood pressure measurements of the h age group from the average in the study i. Hence, the second component of the likelihood variance, $\boldsymbol{\tau}_i^2$ expresses the variance that exists within a study i, the variance that exists *between* all the age groups of the same study i. In Figure 3.3 we can observe in detail the two variances that exist in the likelihood.

3.2.2 The two-dimensional model

A key contribution of this thesis is to extend the one-dimensional model to two dimensions. The one-dimensional model is presented in (3.2.1) and analytically explained in Danaei et al. [2011], Finucane et al. [2011]. The new dependent variables are the DBP, SBP and the interaction(INT) of these two variables. Changes in the construction of the model are necessary. We are expressing the two-dimensional model as having three dependent variables and not two considering the interaction of DBP and SBP as the third dependent variable. By introducing three dependent variables we achieve a diagonal covariance matrix whereas otherwise we would have a dense covariance matrix. A diagonal covariance matrix provides computational efficiency which is useful for our analysis. We have a model

$$\begin{bmatrix} y_{h,i,D} \\ y_{h,i,S} \\ y_{h,i,I} \end{bmatrix} = \begin{bmatrix} a_{j[i],D} \\ a_{j[i],S} \\ a_{j[i],I} \end{bmatrix} + \begin{bmatrix} b_{j[i],D}t_i \\ b_{j[i],S}t_i \\ b_{j[i],I}t_i \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{j[i],(D,S),t_i} \end{bmatrix} + \begin{bmatrix} X_{i,D}\beta \\ X_{i,S}\beta \\ X_{i,I}\beta \end{bmatrix} \\ + \begin{bmatrix} \gamma_{i,D}(z_h) \\ \gamma_{i,S}(z_h) \\ \gamma_{i,I}(z_h) \end{bmatrix} + \begin{bmatrix} e_{i,D} \\ e_{i,S} \\ e_{i,I} \end{bmatrix} + \begin{bmatrix} w_{h,i,D} \\ w_{h,i,S} \\ w_{h,i,I} \end{bmatrix} + \begin{bmatrix} \epsilon_{h,i,D} \\ \epsilon_{h,i,S} \\ \epsilon_{h,i,I} \end{bmatrix} \tag{3.4}$$

where

$$y_{h,i,I} = y_{h,i,D} \times y_{h,i,S}$$

Therefore $\mathbf{y}$ follows a normal distribution:

$$\mathbf{y}_{h,i} \sim N(\mathbf{a}_{j[i]} + \mathbf{b}_{j[i]}t_i + \mathbf{u}_{j[i]}t_i + X_i\beta + \gamma(z_h) + \mathbf{e}_i, \mathbf{SD}^2_{h,t}/n_{h,i} + \boldsymbol{\tau}^2_i) \tag{3.5}$$

where

$$\mathbf{SD}^2_{h,i}/n_{h,i} = \begin{bmatrix} SD^2_{h,i,D}/n_{h,i} & & \\ & SD^2_{h,i,S}/n_{h,i} & \\ & & SD^2_{h,i,I}/n_{h,i} \end{bmatrix} \tag{3.6}$$

$$\boldsymbol{\tau}^2_i = \begin{bmatrix} \tau^2_{i,D} & & \\ & \tau^2_{i,S} & \\ & & \tau^2_{i,I} \end{bmatrix} \tag{3.7}$$

Now, the $\mathbf{y}$ vector will provide three blood pressure measurements for a study i in the age group h in a country j corresponding to DBP, SBP and INT denoted by D, S, I. The $\mathbf{y}$ will still be a vector but the length of it will be three times longer than before. Although each of the parameters in (3.4) and (3.5) is repeated three times, it can be seen that there is a single non-linear sub-model $\mathbf{u}_{j[i],t_i}$. That is because $\mathbf{u}_{j[i],t_i}$ provides simultaneous estimates of non-linear trends in SBP, DBP and INT which are provided implicitly by using an IGMRF as described in Section 2.2. In the following sections we will explore analytically the behaviour of all the sub-models in one and two dimensions.

3.3 Linear sub-model

3.3.1 One-dimensional Linear sub-model

In the linear model, there are two formulas that are used. One describes the intercept of the linear model and the second is referring to the trend of the linear model. These are expressed in (3.8) and (3.9) respectively Danaei et al. [2011], Finucane et al. [2011].

$$a_j = a_j^c + a_{k[j]}^r + a_{l[k]}^s + a^g \tag{3.8}$$

$$b_j = b_j^c + b_{k[j]}^r + b_{l[k]}^s + b^g \tag{3.9}$$

Both the intercept and the slope of the linear model follow a hierarchical structure Danaei et al. [2011], Finucane et al. [2011]. Our dataset has information on countries which are categorised into regions and super-regions Danaei et al. [2011], Finucane et al. [2011]. Finally, all the super-regions comprise the globe Danaei et al. [2011], Finucane et al. [2011]. The hierarchical structure shows a nested form of our data which can be explained as subsections:

$$\text{c: countries} \subset \text{r: regions} \subset \text{s: super-regions} \subset \text{g: global}$$

Equation (3.8) describes that the intercept of a specific country is defined as the summation of the country's intercept, the region's intercept that the country belongs to, the intercept of the country's super-region and the globe's intercept. The hierarchical structure applies to each of the terms, a_j^c, $a_{k[j]}^r$ and $a_{l[k]}^s$. The a_j^c is part of the joint distribution of the intercept for all the J countries, the $a_{k[j]}^r$ term is part of the joint distribution of the intercepts for all the K regions and $a^s_{l[k]}$ is part of the joint distribution of the intercepts for all the L super-regions.

In the same way Equation (3.9) expresses the summation of the slope of a particular country, the slope of the country's region, the slope of the country's super-region and the globe's slope. Again, each of the terms expresses a hierarchical structure. The b^c_j parameter is part of the joint distribution of the slopes for all the J countries, the $b^r_{k[j]}$ parameter is part of the joint distribution of the slopes for all the K regions whereas the $b^s_{l[k]}$ parameter is part of the joint distribution of the slopes for all the L super-regions. The a^g and b^g do not express a hierarchical structure since they account solely for one group, the globe. The hierarchical structure of the intercept and the slope can also be seen in (3.10) which expresses prior distributions.
The countries, regions, super-regions and the globe levels are in a nested form. Therefore there is an impact between the levels as well. As the notation of the terms in (3.8), (3.9) indicate the country's coefficient affects the region's coefficient that this particular country belongs to and in the same way the region's coefficient affects the super-region's coefficient that this particular region belongs to.

It should be stressed that the shrinkage or borrowing strength applies between the J countries, between the K regions and between L super-regions for each of the intercept and the slope respectively by setting the same prior for all the J countries, the same prior for all the K regions and the same prior for all the L super-regions. Each one of the equations' terms of (3.8), (3.9) has its own priors presented in (3.10). Equation (3.10) shows us the prior distribution for each of the terms which is the same distribution across countries, regions and super-regions defined differently for intercept and slope. The hierarchical structure of the linear sub-model is also presented in Figures 3.4 and 3.5 for the case of the linear's sub-model intercept and slope between countries. In the same way regions and

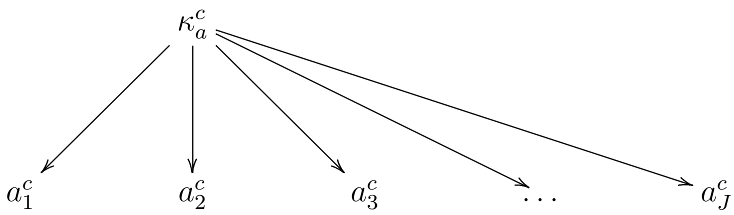

FIGURE 3.4: Structure of the hierarchical model for the intercept of countries in the linear sub-model

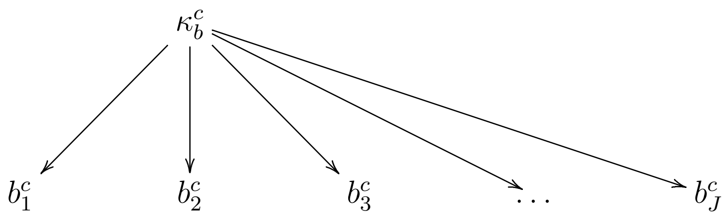

FIGURE 3.5: Structure of the hierarchical model for the slope of countries in the linear sub-model

super-regions can be described.

$$
\begin{aligned}
a_j^c &\sim \mathcal{N}(0, \kappa_a^c), b_j^c \sim \mathcal{N}(0, \kappa_b^c) \\
a_k^r &\sim \mathcal{N}(0, \kappa_a^r), b_k^r \sim \mathcal{N}(0, \kappa_b^r) \\
a_l^s &\sim \mathcal{N}(0, \kappa_a^s), b_l^s \sim \mathcal{N}(0, \kappa_b^s)
\end{aligned}
\tag{3.10}
$$

$$
\begin{aligned}
a^g, b^g &\sim (5, 0.01) \\
\kappa_a^c, \kappa_a^r, \kappa_a^s, \kappa_b^c, \kappa_b^r, \kappa_b^s &\sim (5, 0.01)
\end{aligned}
\tag{3.11}
$$

The κ terms follow a non-informative hyperprior whereas a^g and b^g a non-informative prior.

3.3.2 Two-dimensional Linear sub-model

Since we are interested in the variables DBP, SBP and their interaction, INT we need to construct the linear sub-model accordingly. We introduce the terms of intercept and slope for each of DBP and SBP and INT for the two-dimensional linear sub-model. We will therefore have three different intercepts and three different slopes depending on the i study and j country. Therefore, we use the same modelling of SBP for DBP and INT as well. As in the one-dimensional model we also have a nested structure but this time for each of the three variables. The hierarchical form indicates that countries are nested in regions, regions are nested in super-regions and all these comprise the globe.

$$a_{j,V} = a^c_{j,V} + a^r_{k[j],V} + a^s_{l[k],V} + a^g_V \tag{3.12}$$

$$b_{j,V} = b^c_{j,V} + b^r_{k[j],V} + b^s_{l[k],V} + b^g_V, \quad \text{where} \quad V = D, S, I \tag{3.13}$$

As in the one-dimensional case described in Section 3.2.1, for each of the a and b terms a prior distribution will be assigned without including any dependence between the a's and between the b's.

$$\begin{aligned} a^c_{j,V} \sim \mathcal{N}\left(0, \kappa^c_{(a,V)}\right), b^c_{j,V} \sim \mathcal{N}\left(0, \kappa^c_{(b,V)}\right), a^r_{k,V} \sim \mathcal{N}\left(0, \kappa^r_{(a,V)}\right) \\ b^r_{k,V} \sim \mathcal{N}\left(0, \kappa^r_{(b,V)}\right), a^s_{l,V} \sim \mathcal{N}\left(0, \kappa^s_{(a,V)}\right) \\ b^s_{l,V} \sim \mathcal{N}\left(0, \kappa^s_{(b,V)}\right), \quad \text{where} \quad V = D, S, I \end{aligned} \tag{3.14}$$

It should be noted that in the two-dimensional case the hierarchical structure will exist for each of the variables, DBP, SBP and INT. More specifically, the term $\alpha^c_{(j,D)}$ refers to the intercepts of all the J countries solely for the DBP variable. The same will apply for $\alpha^c_{(j,S)}$ and $\alpha^c_{(j,I)}$ which refer to the intercepts of all the J countries solely for the SBP and INT variables respectively. Therefore, for the intercepts in the two-dimensional linear sub-model we will have three different joint distributions each referring to the joint distribution of countries, regions and super-regions for each of the three variables, DBP, SBP and INT as is depicted for the intercept and slope of the countries level in Figures 3.6 and 3.7. Hence, in total we will have nine different joint distributions.

Finally, we set a non-informative hyperprior for the κ terms and a non-informative prior for the a^g and b^g parameters.

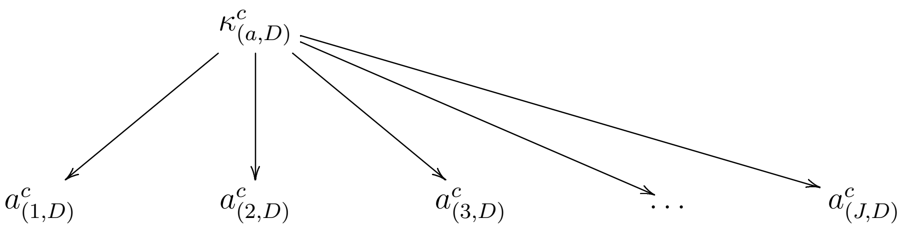

FIGURE 3.6: Structure of the hierarchical model for the DBP's intercept of countries in the linear sub-model

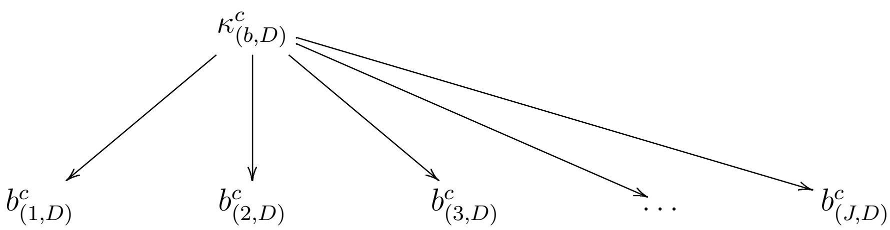

FIGURE 3.7: Structure of the hierarchical model for the DBP's slope of countries in the linear sub-model

3.4 Non-Linear sub-model

3.4.1 One-dimensional Non-Linear sub-model

The non-linear sub-model is used to capture non-linear changes over time. In this model again a hierarchical structure is present where countries are regions' subdivisions, regions are super-regions' subdivisions and all these are subdivisions of the globe. In equation (3.15) the non-linearity of a specific country is expressed as a summation of the country's non-linearity, the non-linearity that exists in the region that the country belongs to, the non-linearity of the country's super-region and finally, by the non-linearity that exists in globe Danaei et al. [2011], Finucane et al. [2011]. As in the linear sub-model, presented in Section 3.3.1, the non-linear sub-model exhibits a hierarchical structure. In more detail, the $\mathbf{u}^c_j$ parameter is part of the joint distribution of all the J countries, the $\mathbf{u}^r_{k[j]}$ parameter is part of the joint distribution of all the K regions whereas the $\mathbf{u}^s_{l[k]}$ parameter is part of the joint distribution of all the L super-regions.

$$\mathbf{u}_j = \mathbf{u}^c_j + \mathbf{u}^r_{k[j]} + \mathbf{u}^s_{l[k]} + \mathbf{u}^g \tag{3.15}$$

In matrix form, the non-linear model for all country-year combinations can be written as Danaei et al. [2011], Finucane et al. [2011]:

$$\mathbf{M}\mathbf{u} = \mathbf{M}^c\mathbf{u}^c + \mathbf{M}^r\mathbf{u}^r + \mathbf{M}^s\mathbf{u}^s + \mathbf{M}^g\mathbf{u}^g \tag{3.16}$$

Here, $\mathbf{M}^c$ is a translation matrix of $I \times JT$ dimensionality that assigns the elements of $\mathbf{u}^c$ to I studies observed in particular country-year combinations, $J \times T$. Since we do not have studies for every year and for every country the $\mathbf{M}^c$ will be a sparse matrix. The $\mathbf{M}^r$, $\mathbf{M}^s$ and $\mathbf{M}^g$ matrices correspond to region-year, super-region-year and globe-year combinations and the dimensions of those are $I \times KT$, $I \times LT$, $I \times T$ respectively. For each of the $\mathbf{u}$ parameters in Equation (3.15) we introduce prior distributions Danaei et al. [2011], Finucane et al. [2011]. From Equation (3.15) we can identify the phenomenon of shrinkage between the J countries since all the J countries will have the same distribution. In a similar way, shrinkage will exist between the K regions and between the L super-regions as all the regions will have the same distribution and also all the super-regions will be assigned to the same distribution. Lastly, we have a distribution for the non-linearity in the globe but since it is not a joint distribution we do not expect to observe the phenomenon of shrinkage between the countries. In Figure 3.9 we can observe the hierarchical structure between the countries and between the regions. It is necessary to stress that each $\mathbf{u}_j^c$, $\mathbf{u}_j^r$, $\mathbf{u}_j^s$ and $\mathbf{u}_j^g$ terma are vectors of T values.

$$\begin{aligned} \mathbf{u}_j^c \sim \mathcal{N}(\mathbf{0}, \lambda_c\mathbf{P}), \ \ \mathbf{u}_{k[j]}^r \sim \mathcal{N}(\mathbf{0}, \lambda_r\mathbf{P}), \\ \mathbf{u}_{l[k]}^s \sim \mathcal{N}(\mathbf{0}, \lambda_s\mathbf{P}), \ \ \mathbf{u}^g \sim \mathcal{N}(\mathbf{0}, \lambda_g\mathbf{P}) \end{aligned} \tag{3.17}$$

In addition, the hyperprior for each of the scalar parameters λ's is defined in (3.18).

$$\lambda_c \sim \mathcal{N}(a_5, b_5)\lambda_r \sim \mathcal{N}(a_6, b_6)\lambda_s \sim \mathcal{N}(a_7, b_7)\lambda_g \sim \mathcal{N}(a_8, b_8) \tag{3.18}$$

In order to allow the model to differentiate among the degrees of non-linearity that exist in country, region, super-region and global levels, we assign each of $\mathbf{u}$'s components a second order IGMRF, as explained in Section 2.5.2. In particular,

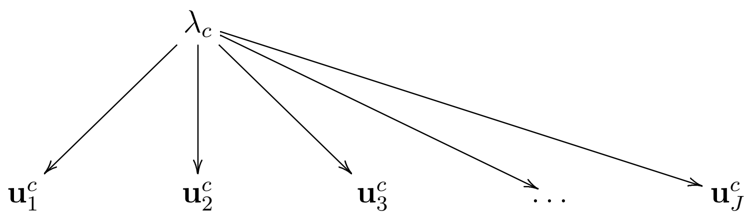

FIGURE 3.8: Structure of the hierarchical model for the non-linear parameters in countries

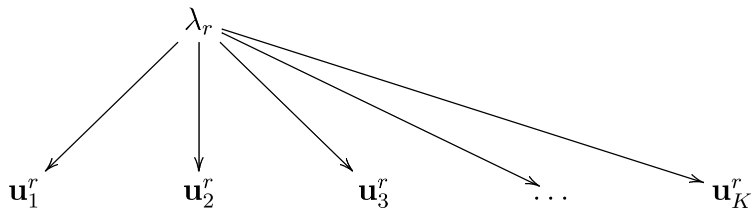

FIGURE 3.9: Structure of the hierarchical model for the non-linear parameters in regions

the T-vectors $\mathbf{u}_j^c(j = 1, ..J)$, $\mathbf{u}_k^r(k = 1, ..K)$, $\mathbf{u}_l^s(l = 1, ..L)$ and $\mathbf{u}^g$ each has as a prior a Gaussian distribution. Specifically, it should be noted that T expresses the years of observations and therefore the $\mathbf{u}_j^c$ will be a vector of length T for each country.

3.4.1.1 Interpretation of the Non-Linear Model

The intercept and the slope of the non-linear model express the same functions as the intercept and the slope of the linear model. However, when the intercept of one of the models, linear and non-linear, changes in time, the other should be equal to zero. In other words, we cannot have intercepts of linear and non-linear change in time at the same time. Only one of these relations should be applied. These are known as identifiability constraints Danaei et al. [2011], Finucane et al. [2011].

3.4.1.2 Identifiability connection with the Linear Model

The definition of identifiability is equivalent to saying that different values of the parameters must generate different probability distributions of the observable variables Lehmann and Casella [2006]. In other words, it means that distinct values of $\boldsymbol{\phi}$ should correspond to distinct probability distributions. Therefore, a model is identifiable if the parameter values, $(\boldsymbol{\phi}_0, \boldsymbol{\phi}) \in \boldsymbol{\Phi}$ uniquely determine the probability distribution of the data, $\mathbf{y} \in S_{\mathbf{y}}$ and the probability distribution of the data uniquely determines the parameter values:

$$F(\mathbf{y}; \boldsymbol{\phi}_0) = F(\mathbf{y}; \boldsymbol{\phi}) \quad \text{iff} \quad \boldsymbol{\phi}_0 = \boldsymbol{\phi}, \tag{3.19}$$

where $\boldsymbol{\Phi}$ denotes the set of all possible parameter values and $S_{\mathbf{y}}$ is the set of all possible values of data Huang [2005]. Identifiability constraints need to be applied between the linear and non-linear parameters which are $\mathbf{a}$, $\mathbf{b}$ and $\mathbf{u}$ Danaei et al. [2011], Finucane et al. [2011]. The restriction we will apply here is to restrict the last group, the non-linear sub-model to zero. For that reason the mean of the intercept and slope for each $\mathbf{u}_c$, $\mathbf{u}_r$, $\mathbf{u}_s$ and $\mathbf{u}_g$ is equal to zero. This is equivalent to saying that between two groups in order to have identifiability we need to set the second group equal to zero, expressed in Lehmann and Casella [2006]. A different identifiability constraint would be the sum of the linear and non-linear intercepts being equal to zero and by the same token the sum of the linear and non-linear slopes being equal to zero as well. More details can be found in Lehmann and Casella [2006].

Many times Bayesian models present identifiability problems that are caused due to collinearity or overparameterization etc. One solution to identifiability problems of this nature is the proposal of weakly informative priors. However, references given in the bibliography have stated that a formal definition of Bayesian non-identifiability is equivalent to a lack of identifiability in the likelihood. Therefore, this equivalence also implies that identifiability does not depend on the nature of the prior specification. Of course, priors that are too informative i.e. too precise, strong, will limit Bayesian learning from the data. With regard to prior specification in nonidentifiable models, the middle ground between too informative and insufficiently informative is the solution, Gelfand and Sahu [1999].

3.4.2 Two-dimensional non-linear sub-model

The two-dimensional non-linear model has a different behaviour from the rest of the sub-models. The interaction of DBP and SBP, INT is already integrated in the model. Precisely, by having the DBP and SBP we account implicitly for the interaction of these two variables. By considering every possible combination between DBP and SBP, we actually account for changes in DBP, for changes in SBP and every simultaneous change for both DBP and SBP which is the interaction of these two variables. By using in the prior the $\mathbf{P}$ matrix that was presented in Section 2.6.5 we are able to take into account the already implicit variable of interaction. Throughout this section we will use the precision matrix of Section 2.6.5 which describes the second-order polynomial IGMRF in two dimensions. The non-linear term $\mathbf{u}_{j[i]}$ for country j and study i, present in the likelihood in (3.4) and (3.5) can be represented as:

$$\mathbf{u}_{j,(D,S)} = \mathbf{u}^c_{j,(D,S)} + \mathbf{u}^r_{k[j],(D,S)} + \mathbf{u}^s_{l[k],(D,S)} + \mathbf{u}^g_{(D,S)} \tag{3.20}$$

In matrix form, the non-linear model for all country-year combinations can be written as:

$$\mathbf{M}\mathbf{u}_{(D,S)} = \mathbf{M}^c\mathbf{u}^c_{(D,S)} + \mathbf{M}^r\mathbf{u}^r_{(D,S)} + \mathbf{M}^s\mathbf{u}^s_{(D,S)} + \mathbf{M}^g\mathbf{u}^g_{(D,S)} \tag{3.21}$$

Here, $\mathbf{M}^c$ is a translation matrix of $I \times JT^2$ dimensionality that assigns the elements of $\mathbf{u}^c_{(D,S)}$ to studies observed in particular country-year combinations. The $\mathbf{M}^r$, $\mathbf{M}^s$ and $\mathbf{M}^g$ matrices correspond to region-year, super-region-year and globe-year combinations with $I \times KT^2$, $I \times LT^2$ and $I \times T^2$ dimensions respectively. For each of the $\mathbf{u}$ variables we have a prior distribution

$$\begin{aligned} \mathbf{u}^c_{j,(D,S)} &\sim \mathcal{N}(\mathbf{0}, \lambda_c \mathbf{P}), \mathbf{u}^r_{k,(D,S)} \sim \mathcal{N}(\mathbf{0}, \lambda_r \mathbf{P}) \\ \mathbf{u}^s_{l,(D,S)} &\sim \mathcal{N}(\mathbf{0}, \lambda_s \mathbf{P}), \mathbf{u}^g_{(D,S)} \sim \mathcal{N}(\mathbf{0}, \lambda_g \mathbf{P}) \end{aligned} \tag{3.22}$$

where $\mathbf{P}$ is a precision matrix defined in Section 2.6.5. By the structure of the priors in (3.22) we assume dependence between DBP, SBP and INT which is expressed through the precision matrix $\mathbf{P}$ as also explained in Yue and Speckman [2010]. We can notice that given the structure of the priors in (3.22) we have the same number of hierarchical structures as in the one dimensional case, Section 3.4.1. In more detail, the $\mathbf{u}^c_{j,(D,S)}$ term expresses the non-linearity in a two dimensional model of DBP and SBP for each of the J countries having as a result that all the J countries follow the same distribution. This actually means that there will be shrinkage of all the j values. The same will appear in each of the $\mathbf{u}^r_{k,(D,S)}$ and $\mathbf{u}^s_{l,(D,S)}$ where shrinkage will exist between the k and l values respectively. Since, the $\mathbf{u}^g_{(D,S)}$ is related to only one group we are not referring to a joint distribution and therefore the shrinkage phenomenon cannot be applied. The hierarchical structure of the two-dimensional non-linear sub-model is presented in Figures 3.10 and 3.11 for the country and regional level respectively. It should be noted that $u_{j(D,S)}$ will be a vector of size T^2. Lastly, hyperpriors for the scalar parameters λ's are introduced

$$\lambda_c \sim \mathcal{N}(a_5, b_5), \lambda_r \sim \mathcal{N}(a_6, b_6), \lambda_s \sim \mathcal{N}(a_7, b_7), \lambda_g \sim \mathcal{N}(a_8, b_8) \tag{3.23}$$

In equation (3.22) $\mathbf{0}$ is a vector of length T^2, $\mathbf{P}$ is a $T^2 \times T^2$ matrix, the detailed form of which is shown in Section 2.6.5, and each of the non-linear terms is a vector that comprises all possible combinations of years for DBP and SBP. For instance, for the country-level non-linear term, we have a vector of length T^2, $\mathbf{u}^c_j = (u_{[1,1]}, u_{[1,2]}, \ldots, u_{[3,2]}, u_{[3,4]}, u_{[3,5]}, \ldots, u_{[T,T]})$ for a specific country. The precision parameters, λ_c, λ_r, λ_s, λ_g, are each scalars. Each will follow a Normal distribution as shown in (3.23) with each of the λ parameters having a larger variance as in the one-dimensional equivalent in (3.18) .

$$
\begin{tikzcd}
 & \lambda^c \arrow[dl] \arrow[d] \arrow[dr] \arrow[drr] \arrow[drrr] & & & \\
\mathbf{u}^c_{1,(D,S)} & \mathbf{u}^c_{2,(D,S)} & \mathbf{u}^c_{3,(D,S)} & \cdots & \mathbf{u}^c_{J,(D,S)}
\end{tikzcd}
$$

FIGURE 3.10: Structure of the hierarchical model for DBP, SBP and INT of countries in the non-linear sub-model

$$
\begin{tikzcd}
 & \lambda^r \arrow[dl] \arrow[d] \arrow[dr] \arrow[drr] \arrow[drrr] & & & \\
\mathbf{u}^r_{1,(D,S)} & \mathbf{u}^r_{2,(D,S)} & \mathbf{u}^r_{3,(D,S)} & \cdots & \mathbf{u}^r_{K,(D,S)}
\end{tikzcd}
$$

FIGURE 3.11: Structure of the hierarchical model for DBP, SBP and INT of regions in the non-linear sub-model

While the model allows estimation for T^2 combinations, we can only add information to the likelihood when data are present, and this is only the case when both DBP and SBP are in the same year because the measurements are taken for individuals simultaneously. Here, we describe the $\mathbf{u}^c_j$ parameter at the country level of the hierarchy, but the $\mathbf{u}$ parameter for the other levels is analogous.

3.4.2.1 Explicit use of INT in non-linear sub-model

The non-linear model will be an extension of the one dimensional non-linear model that we presented in Section 3.4.1. For each of the DBP, SBP and INT variables the model will be the same, a second order random walk for the precision of the prior without considering dependence among the three variables, DBP, SBP and

INT.

$$\mathbf{u}_{(j,V)} = \mathbf{u}^c_{(j,V)} + \mathbf{u}^r_{(k[j],V)} + \mathbf{u}^s_{(l[j],V)} + \mathbf{u}^g_V \tag{3.24}$$

$$\mathbf{M}\mathbf{u}_V = \mathbf{M}^c\mathbf{u}_{(c,V)} + \mathbf{M}^r\mathbf{u}_{(r,V)} + \mathbf{M}^s\mathbf{u}_{(s,V)} + \mathbf{M}^g\mathbf{u}_{(g,V)}, \quad \text{where} \quad V = D, S, I \tag{3.25}$$

In that way each of the DBP, SBP and INT will have its own prior distributions. For each of the $\mathbf{u}$ we have the prior distributions that are presented in Section 2.5.2.The prior distributions is defined as follow:

$$\begin{aligned} \mathbf{u}^c_{j,V} \sim \mathcal{N}(\mathbf{0}, \lambda_c\mathbf{P}), \mathbf{u}^c_{k,V} \sim \mathcal{N}(\mathbf{0}, \lambda_r\mathbf{P}), \mathbf{u}^c_{l,V} \sim \mathcal{N}(\mathbf{0}, \lambda_s\mathbf{P}), \\ \mathbf{u}^c_V \sim \mathcal{N}(\mathbf{0}, \lambda_g\mathbf{P}), \quad \text{where} \quad V = D, S, I \end{aligned} \tag{3.26}$$

Finally, following this modelling we will have four precision parameters for each of the DBP, SBP, INT variables whereas before we had four precision parameters for all the three variables together. Hence, the hyperpriors for the scalar parameters λ's solely for the DBP variable is presented in equation (3.27).

$$\lambda_c \sim \mathcal{N}(a_5, b_5), \lambda_r \sim \mathcal{N}(a_6, b_6), \lambda_s \sim \mathcal{N}(a_7, b_7), \lambda_g \sim \mathcal{N}(a_8, b_8) \tag{3.27}$$

3.5 Covariate effects sub-model

3.5.1 One-dimensional covariate effects sub-model

The covariate effects sub-model includes two types of effects, the country and study level effects as defined by Danaei et al and Finucane et al in Danaei et al. [2011], Finucane et al. [2011].

At the country level we include two time-varying covariates Danaei et al. [2011], Finucane et al. [2011]. The first is country-level urbanization, defined as the proportion of a country's population that lived in urban areas. The second one is national availability of multiple food types ($\mathbf{X}^{food}$). A principal component analysis (PCA) was used to summarize food availability data from which we used the first four components: $X^{food} = (\mathbf{X}^{food_1}, \mathbf{X}^{food_2}, \mathbf{X}^{food_3}, \mathbf{X}^{food_4})$ Bentham et al. [2020]. The purpose of PCA is to capture what are the food types of consumption for each country.

At the study level we include two covariates Danaei et al. [2011], Finucane et al. [2011]. The first is coverage, $\mathbf{X}^{cvrg}$, a three-level factor variable indicating whether the data were national, subnational or were from individual communities. The second is study-level urbanization, $\mathbf{X}^{s.urb}$, a three level factor indicating whether the study population was rural, urban or in between and expressed as rural-and-urban.

In equation (3.28) we represent the country-level urbanization having an intercept and a term depending on time whereas the food types are solely depending on time. Study-level covariates exist to tackle possible biases that exist in the data. Previous studies have shown the need to consider possible differences from national studies as subnational or community's studies may exhibit different behaviour than the national studies. Therefore, we have terms for the subnational and community coverage both having an intercept and terms depending on time, Danaei et al. [2011], Finucane et al. [2011]. Furthermore, we want to compare both country's and study's urbanization when these two terms have an antithetic behaviour. Moreover, we have terms in (3.28) that combine a country's urbanization with rural study's urbanization and on the other hand terms that shows country's lack of urbanization with a study in an urban area, as was explained in Danaei et al. [2011] and Finucane et al. [2011].

$$
\begin{aligned}
\mathbf{X}_i\boldsymbol{\beta} = {} & \beta_1\mathbf{X}^{c.urb}_{j[i],t[i]} + \beta_2\mathbf{X}^{c.urb}_{j[i],t[i]}t_i + \\
& \beta_3\mathbf{X}^{food_1}_{j[i],t[i]}t_i + \beta_4\mathbf{X}^{food_2}_{j[i],t[i]}t_i + \beta_5\mathbf{X}^{food_3}_{j[i],t[i]}t_i + \beta_6\mathbf{X}^{food_4}_{j[i],t[i]}t_i + \\
& \beta_7\mathbf{I}\{\mathbf{X}^{cvrg}_{j[i],t[i]} = \text{subnational}\} + \beta_8\mathbf{I}\{\mathbf{X}^{cvrg}_{j[i],t[i]} = \text{subnational}\}t_i + \\
& \beta_9\mathbf{I}\{\mathbf{X}^{cvrg}_{j[i],t[i]} = \text{community}\} + \beta_{10}\mathbf{I}\{\mathbf{X}^{cvrg}_{j[i],t[i]} = \text{community}\}t_i + \\
& \beta_{11}\mathbf{X}^{c.urb}_{j[i],t[i]}\mathbf{I}\{\mathbf{X}^{s.urb}_{j[i],t[i]} = \text{rural}\} + \beta_{12}\mathbf{X}^{c.urb}_{j[i],t[i]}\mathbf{I}\{\mathbf{X}^{s.urb}_{j[i],t[i]} = \text{rural}\}t_i + \\
& \beta_{13}(1 - \mathbf{X}^{c.urb}_{j[i],t[i]})\mathbf{I}\{\mathbf{X}^{s.urb}_{j[i],t[i]} = \text{urban}\} + \beta_{14}(1 - \mathbf{X}^{c.urb}_{j[i],t[i]})\mathbf{I}\{\mathbf{X}^{s.urb}_{j[i],t[i]} = \text{urban}\}t_i
\end{aligned} \tag{3.28}
$$

Observing countries with urbanization under an urban study means that $\mathbf{X}_i$ will have a zero value. Likewise, countries not being characterized with urbanization under a rural study means that $\mathbf{X}_i$ will have a zero value.

3.5.2 Two-dimensional covariate effects sub-model

As before, the covariate effects sub-model includes two types of effects, the country and study level effects as defined by Finucane et al. [2011] and Danaei et al. [2011]. The countries and the studies are the same for both one- and two- dimensional models, therefore we expect the $\mathbf{X}$ matrix to contain the same information for all the three dependent variables. However, we will set different β terms for each of our variables. In (3.29) we define the $\boldsymbol{\beta}_1$ but the same applies for the rest of the $\boldsymbol{\beta}_k$ terms where $k = 1, \ldots, 14$.

$$
\boldsymbol{\beta}_1 = \begin{bmatrix} \beta_{1D} \\ \beta_{1S} \\ \beta_{1I} \end{bmatrix} \tag{3.29}
$$

The $\mathbf{X}$ matrix for each of the covariate effects will be a matrix of three columns from now on, each column of which will contain the same information for DBP, SBP, INT, so the columns will be identical multiplying with the parameters in (3.29).

3.6 Age sub-model

3.6.1 One-dimensional Age sub-model

The mean of SBP may be non-linearly associated with age. Therefore, a choice of a non-parametric model is more appropriate. B-splines then were used for the age sub-model. In the work of Danaei et al and Finucane et al cubic splines were used to fit the age sub-model Danaei et al. [2011], Finucane et al. [2011]. The use of B-splines instead is because of their greater flexibility. The cubic spline that was used in Danaei et al. [2011], Finucane et al. [2011] has the form explained in (3.30):

$$\gamma_i(z_h) = \gamma_{1i} z_h + \gamma_{2i} z_h^2 + \gamma_{3i} z_h^3 + \gamma_{4i}(z_h - 45)_+^3 + \gamma_{5i}(z_h - 60)_+^3 \tag{3.30}$$

In contrast, the equation for a B-spline curve of degree k is defined by

$$S(t) = \sum_{i=0}^{n} N_{i,k}(t) P_i \tag{3.31}$$

where $(P_0, P_1, \ldots, P_n)$ are control points and $N_{i,k}(t)$ are the **basis functions** which are defined using the **Cox-de Boor recursion formula** shown in (3.32).

$$\begin{aligned} N_{i,0}(t) &= \begin{cases} 1 & \text{if} \quad t_i \leqslant t < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \\ N_{i,j}(t) &= \frac{t - t_i}{t_{i+j} - t_i} N_{i,j-1}(t) + \frac{t_{i+j+1} - t}{t_{i+j+1} - t_{i+1}} N_{i+1,j-1}(t) \end{aligned} \tag{3.32}$$

The values of t_i are taken from a sequence called a knot vector

$$T = (t_0, t_1, \ldots, t_m)$$

The number of knots in T, $m + 1$ is related to the degree k and the number of control points $n + 1$ by

$$m = k + n + 1$$

For example, in our analysis we will use a cubic B-spline, $k = 3$ with number of control points $n+1 = 4+1$ therefore $m = k+n+1 = 3+4+1 = 8$ knots with a knot vector

$$T = (t_0, t_1, \ldots, t_8)$$

Therefore, B-splines can be expressed as:

$$S(t) = \sum_{i=0}^{4} N_{i,3}(t) P_i \tag{3.33}$$

Hence, in equation (3.34) we express the age sub-model with cubic B-splines

$$\begin{aligned} S(t) = \gamma_i(N(t)) &= \sum_{j=1}^{5} N_{j-1,3}(t)\gamma_{ji} \\ &= \gamma_{1i}N_{0,3}(t) + \gamma_{2i}N_{1,3}(t) + \gamma_{3i}N_{2,3}(t) + \gamma_{4i}N_{3,3}(t) + \gamma_{5i}N_{4,3}(t) \end{aligned} \tag{3.34}$$

where each i expresses a particular study.

In order to reduce dependence among model parameters we use a middle-aged group, specifically 50-year-old as the baseline age, centring the age variable by subtracting 50 from all values. Each of the $\gamma_{1i}, \gamma_{2i}, \ldots, \gamma_{5i}$ values are defined in (3.35) and also stated in Danaei et al. [2011], Finucane et al. [2011]:

$$\gamma_{ki} = \psi_k + \phi_k \mu_i + c_{kj[i]} \quad \text{where} \quad k = 1, 2, \ldots, 5 \tag{3.35}$$

Generally, blood pressure can increase more steeply with age where blood pressure in a baseline age group is higher. In order to allow the shape of the age curve to vary with level, the spline coefficients for study i $(\gamma_{1i}, \ldots, \gamma_{5,i})$ are permitted to depend on μ_i, the estimated 50-year-old SBP value for i study Danaei et al. [2011], Finucane et al. [2011]:

$$\mu_i = a^c_{j[i]} + b^c_{j[i]} t + \mathbf{X}_i \boldsymbol{\beta} + u_{j[i]} t_i + e_i$$

Age association not only can vary due to the differences in country means but also due to random effects. Therefore, we include country-specific random spline

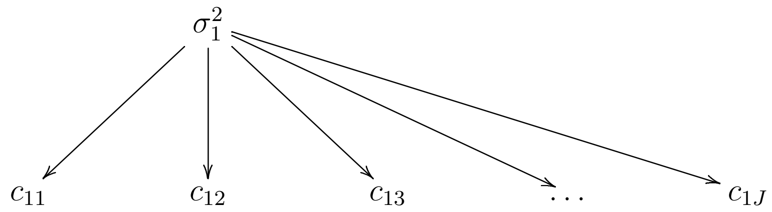

FIGURE 3.12: Structure of the hierarchical model for the random effects of the age sub-model for the σ_1^2 case

coefficients in our specification of the γ's as stated in Danaei et al. [2011], Finucane et al. [2011]:

$$c_{kj} \sim \mathcal{N}(0, \sigma_k^2), \quad \text{where} \quad k = 1, 2, \dots, 5 \tag{3.36}$$

with a flat improper prior placed on each of the five σ^2's. It should be noted that we have a hierarchical structure here too. For example, c_{1j} for all the J countries will follow the same distribution with zero mean and variance σ_1^2. An example of how the hierarchical model is structured is presented in Figure 3.12 for the $c_{11}, \dots, c_{1J}$. The same will apply for each of the other $c_{2j}, \dots, c_{5j}$ terms where each is a term of a joint distribution for all the J countries. Finally, the $\psi_1, \dots, \psi_5$ parameters describe the age pattern for an "average" country so, $c_{1j} = \cdots = c_{5j} = 0$ at an average SBP level corresponding to $\mu_i = 0$.

It is necessary to stress that in this sub-model we express the age as a continuous variable although the blood pressure measurements are reported for age-groups, Danaei et al. [2011], Finucane et al. [2011]. For the simplicity of the calculations, we used the midpoint of each age-group as a temporary replacement for the unobserved continuous age value.

3.6.2 Two-dimensional Age sub-model

There might be non-linearity between age and each of the variables, DBP, SBP and INT. Therefore, for each of the relations we use B-splines. In (3.37) we construct three age sub-models each expressing the association between dependent variable and age.

$$
\begin{aligned}
S(t) = \gamma_{i,V}(N(t)) &= \sum_{j=1}^{5} \gamma_{ji,V} N_{j-1,3}(t) \\
&= \gamma_{1i,V} N_{0,3}(t) + \gamma_{2i,V} N_{1,3}(t) + \gamma_{3i,V} N_{2,3}(t) + \gamma_{4i,V} N_{3,3}(t) + \gamma_{5i,V} N_{4,3}(t), \\
\text{where} \quad V &= D, S, I
\end{aligned} \tag{3.37}
$$

The γ's are given by:

$$
\begin{aligned}
\boldsymbol{\gamma}_{1i} = \begin{bmatrix} \gamma_{1i,D}(z_h) \\ \gamma_{1i,S}(z_h) \\ \gamma_{1i,I}(z_h) \end{bmatrix} &= \begin{bmatrix} \psi_1 \\ \psi_2 \\ \psi_3 \end{bmatrix} + \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \end{bmatrix} \mu_i + \begin{bmatrix} c_{1j[i]} \\ c_{2j[i]} \\ c_{3j[i]} \end{bmatrix} \\
&\cdots \\
\boldsymbol{\gamma}_{5i} = \begin{bmatrix} \gamma_{5i,D}(z_h) \\ \gamma_{5i,S}(z_h) \\ \gamma_{5i,I}(z_h) \end{bmatrix} &= \begin{bmatrix} \psi_{13} \\ \psi_{14} \\ \psi_{15} \end{bmatrix} + \begin{bmatrix} \phi_{13} \\ \phi_{14} \\ \phi_{15} \end{bmatrix} \mu_i + \begin{bmatrix} c_{13j[i]} \\ c_{14j[i]} \\ c_{15j[i]} \end{bmatrix}
\end{aligned} \tag{3.38}
$$

Again as we want the shape of the age curve to vary with level, the spline coefficients for study i, $(\boldsymbol{\gamma}_{1i}, \ldots, \boldsymbol{\gamma}_{5i})$ are allowed to depend on μ_i, the estimated value of SBP, DBP and their interaction for 50 years old for that study:

$$
\boldsymbol{\mu}_i = \mathbf{a}^c_{j[i]} + \mathbf{b}^c_{j[i]} t + \mathbf{X}_i \beta + u_{j[i]} t + \mathbf{e}_i \tag{3.39}
$$

Also **a** and **b** express the intercept and the slope of the linear model for each of the three variable, DBP, SBP and their interaction. We also include the random spline coefficients for the j country in our specification of the γ's as well:

$$
c_{kj} \sim N(0, \sigma_k^2) \quad \text{where} \quad k = 1, 2, \ldots, 15 \tag{3.40}
$$

3.7 Age-related random effects

3.7.1 One-dimensional $\boldsymbol{\sigma}^2$'s sub-model

In this section we are allowing for the variance of the random effects of the age sub-model. We introduce the hyperparameter of the age sub-model, σ_k^2, where $k = 1, \ldots, 5$. Following we define the prior and hyperprior distributions that were used for the parameters and the hyperparameters of the age sub-model. More specifically, we will account for the variance parameter of the age sub-model's random effects Danaei et al. [2011], Finucane et al. [2011]. Defining the random effects as $\mathbf{c}_k$ and the variance of those, σ_k^2, hence the hyperprior for σ_1^2 can be written as presented in 3.41. An analogous procedure is followed for $k = 2, \ldots, 5$. These two variables can be defined as:

$$\begin{aligned} \mathbf{c}_1|\sigma_1^2 &\sim \mathcal{N}(0, \sigma_1^2 \mathbf{I}_J) \\ \sigma_1^2 &\sim \mathcal{B}eta(c, d) \end{aligned} \tag{3.41}$$

In 3.41, $\mathbf{c}_1$ represents a vector of J values and $\mathbf{I}_J$ a $J \times J$ diagonal matrix. In the same way, $\sigma_2^2, \ldots, \sigma_5^2$ are defined.

3.7.2 Two-dimensional $\boldsymbol{\sigma}^2$'s sub-model

In the two-dimensional case the age model has 15 sub-models and hence, 15 random effects of the age model, c terms will exist. This occurs because there will be five sub-models for each of the three variables represented in (3.38). In this section we describe the hyperparameters of the c random effects. The σ^2 term is related to the random variable c of a specific country j as shown in (3.42).

$$\begin{bmatrix} c_{1j} \\ c_{2j} \\ \cdots \\ c_{15j} \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \mathrm{diag} \begin{bmatrix} \sigma_1^2 \\ \sigma_2^2 \\ \cdots \\ \sigma_{15}^2 \end{bmatrix} \right) \tag{3.42}$$

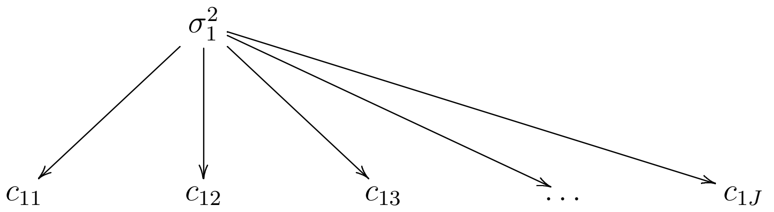

FIGURE 3.13: Structure of the hierarchical model for the DBP's random effects of the age sub-model between countries

For all the countries the distribution in matrix form can be expressed as:

$$\mathbf{c}_k \sim \mathcal{N}(\mathbf{0}, \sigma_k^2 \mathbf{I}_J) \qquad \text{for} \qquad k = 1, 2, \dots, 15 \tag{3.43}$$

In a joint Gaussian distribution without accounting for correlation between the $\mathbf{c}_k$ terms we can write:

$$\begin{bmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \cdots \\ \mathbf{c}_{15} \end{bmatrix} \sim \mathcal{N} \left(\mathbf{0}, diag \begin{bmatrix} \sigma_1^2 \mathbf{I}_J \\ \sigma_2^2 \mathbf{I}_J \\ \cdots \\ \sigma_{15}^2 \mathbf{I}_J \end{bmatrix} \right) \tag{3.44}$$

where $\mathbf{I}_j$ is a diagonal matrix of $J \times J$ dimensions. Once again, each of the $\mathbf{c}_k$ terms expresses a hierarchical structure in which shrinkage is observed between the J countries. The difference with the one-dimensional case is that we will have a hierarchical structure for each of the 15 $\mathbf{c}_k$ terms whereas before there were only 5 different $\mathbf{c}_k$ terms.

The distribution for each σ_k^2 hyperparameter, where $k = 1, 2, \dots, 15$, is expressed as:

$$\sigma_k^2 \sim \mathcal{B}eta(c, d) \tag{3.45}$$

Concluding, each of the σ_k^2's hyperparameters is independent of one another.

3.8 Study-specific random effects, e_i^2

3.8.1 One-dimensional e_i^2

In each study there are effects from study to study the so-called e_i Danaei et al. [2011], Finucane et al. [2011]. Therefore, for each study and for all the age groups within the study there is the same e_i effect which means that the study-specific e_i term does not allow differences between the age groups of the i study which has as a result that all age groups can have an unusually high or an unusually low mean after accounting for other terms in the model. Moreover, we introduce a prior on the e_i terms which depends on the national, subnational or community coverage of the i study. In this section we are interested in estimating the variance of e_i's prior, the hyperparameter ν_i Danaei et al. [2011], Finucane et al. [2011]. Each e_i is assigned a Gaussian prior with variance depending on the coverage of study i Danaei et al. [2011], Finucane et al. [2011]:

$$\nu_i = var(e_i) = \begin{cases} \nu_{nat} & \text{if study } i \text{ is national} \\ \nu_{subn} & \text{if study } i \text{ is subnational} \\ \nu_{com} & \text{if study } i \text{ is community level} \end{cases}$$

$$\begin{aligned} e_i &\sim \mathcal{N}(0, \nu_i) \\ \nu_i &\sim \mathcal{B}eta(c, d) \end{aligned} \tag{3.46}$$

The ν_{nat} term expresses the fact that even after accounting for sampling variability, national studies may still not represent the country's mean SBP level with perfect accuracy. Random effects from national studies are constrained to have smaller variance than random effects from subnational studies and so forth ($\nu_{nat} < \nu_{subn} < \nu_{com}$) as explained in Danaei et al. [2011], Finucane et al. [2011].

3.8.2 Two-dimensional e_i^2

In this section we describe the distribution of the hyperparameter for the study-specific random effect $\mathbf{e}_i$. Each $\boldsymbol{\nu}$ is the hyperparameter under study which expresses the variance of the $\mathbf{e}_i$ terms. Each of the ν_D, ν_S, ν_I terms can take the value of ν_{nat}, ν_{subn} and ν_{com} if the study i is national, subnational or community respectively. Since D, S, I are expressing the measurements of the same object all the three variables will have the same coverage in a specific study i. By denoting in bold the national, subnational and community categories (such as $\boldsymbol{\nu}^2_{nat}$) we have a vector of three values, one for each of the variables. In equation (3.47) we observe how the national coverage is defined. The procedure is analogous for the subnational and community coverage.

$$\boldsymbol{\nu}_{nat} = \begin{bmatrix} \nu_{nat,D} \\ \nu_{nat,S} \\ \nu_{nat,I} \end{bmatrix} \tag{3.47}$$

$$\boldsymbol{\nu}_i = \begin{bmatrix} \nu_{i,D} \\ \nu_{i,S} \\ \nu_{i,I} \end{bmatrix} = \begin{cases} \boldsymbol{\nu}^2_{nat} & \text{if study } i \text{ is national} \\ \boldsymbol{\nu}^2_{subn} & \text{if study } i \text{ is subnational} \\ \boldsymbol{\nu}^2_{com} & \text{if study } i \text{ is community level} \end{cases}$$

Since, we no longer have only SBP but also DBP and their interaction we will have a vector of three variables expressing the study-specific random effect $\mathbf{e}_i$.

$$e_{i,V} \sim \mathcal{N}\left(0, \nu_{i,V}\right), \quad \text{where} \quad V = D, S, I \tag{3.48}$$

Therefore, we will have two more variances for the study-specific random effect $\mathbf{e}_i$ which will not be correlated with each other.

$$\nu_{i,D}, \nu_{i,S}, \nu_{i,I} \sim \mathcal{B}eta(c, d) \tag{3.49}$$

For each of the hyperparameters a hyperprior distribution is assigned. Finally, we assume there is less variability in national studies than in subnational studies and again less than in the community studies.

3.9 Residual age-by-study variability, τ^2

3.9.1 One-dimensional τ^2

Not taking into account the effect of baseline-age SBP, the age patterns within a given study may differ and not be consistent with their country's age pattern. For each $y_{h,i}$ measurement there is a standard deviation for the sample of individuals in a specific age group and study ($SD^2_{h,i}/n_{h,i}$). We want to account for the variance of residuals that exists between the age groups within a study denoted as τ_i^2. It should be stressed that this cannot be captured by the e's, which are equal across all age groups in any given study, so we include an additional variance component for each study. As in the case of the e's, τ_i^2 depends on the study's coverage. A prior distribution is defined for $\epsilon_{h,i}$ which depends on the study's coverage and is defined as is explained in Danaei et al. [2011], Finucane et al. [2011]:

$$\tau_i^2 = \begin{cases} \tau^2_{nat} & \text{if study } i \text{ is national} \\ \tau^2_{subn} & \text{if study } i \text{ is subnational} \\ \tau^2_{com} & \text{if study } i \text{ is community level} \end{cases}$$

$$\begin{aligned} \epsilon_{h,i} &\sim \mathcal{N}(0, \tau_i^2) \\ \tau_i^2 &\sim \mathcal{B}eta(c, d) \end{aligned} \tag{3.50}$$

We again assume that there is less residual variability in national studies than in subnational studies and so forth, with $\tau^2_{nat} < \tau^2_{subn} < \tau^2_{com}$.

3.9.2 Two-dimensional τ^2

The τ^2 term expresses the variance of the age-by-study residuals, $\epsilon_{h,i}$. The residuals express the deviation of the blood pressure measurement from the average in study i. In the two-dimensional model the blood pressure measurement is represented by three values, DBP, SBP and their interaction so we do not expect each of these variables to have the same deviation from the average in study i. The coverage of the i study is the same for each of the three variables, DBP, SBP and INT but since the $\mathbf{y}_{h,i}$ has a different range of values in each of the variables we expect to have a different variance as well. By denoting with bold the national, subnational and community categories (such as $\boldsymbol{\tau}^2_{nat}$) we express a vector of three values, one for each of the variables. In equation (3.51) we observe how the national coverage is defined. The procedure is analogous for the subnational and community coverage.

$$\boldsymbol{\tau}^2_{nat} = \begin{bmatrix} \tau^2_{nat,D} \\ \tau^2_{nat,S} \\ \tau^2_{nat,I} \end{bmatrix} \tag{3.51}$$

$$\boldsymbol{\tau}^2_{i} = \begin{bmatrix} \tau^2_{i,D} \\ \tau^2_{i,S} \\ \tau^2_{i,I} \end{bmatrix} = \begin{cases} \boldsymbol{\tau}^2_{nat} & \text{if study } i \text{ is national} \\ \boldsymbol{\tau}^2_{subn} & \text{if study } i \text{ is subnational} \\ \boldsymbol{\tau}^2_{com} & \text{if study } i \text{ is community} \end{cases}$$

We define different $\epsilon_{h,i}$ for each of the three variables and hence, three different values for τ_i. Since, we have three variables the τ_i will be a vector of three values.

$$\epsilon_{h,(i,V)} \sim \mathcal{N}\left(0, \tau^2_{i,V}\right), \quad \text{where} \quad V = D, S, I \tag{3.52}$$

Next we present the hyperprior distributions for the parameter $\boldsymbol{\tau}_i$. Each of the $\boldsymbol{\tau}_i$ hyperparameters are independent of one another and each of them follows a Beta distribution:

$$\tau^2_{i,D}, \tau^2_{i,S}, \tau^2_{i,I} \sim \mathcal{B}eta(c, d) \tag{3.53}$$

We again assume there is less variability in national studies than in subnational studies and so forth. In summary, residuals, $\boldsymbol{\epsilon}_{h,i}$ have two types of variance, and in this section we describe the variance between the age groups in each study.

3.10 Summary

This chapter provides a detailed description of each of the sub-models that comprise the one-dimensional and two-dimensional model of blood pressure shown in equations (3.2) and (3.4) respectively. In Sections 3.2.1 and 3.2.2 we can see how the one-dimensional and two-dimensional models are formed, respectively. In Section 3.2.1 we present the parameters that the likelihood consists of, having the SBP as the dependent variable and then we present analytically each of the likelihood's parameters whereas in Section 3.2.2 we add two more variables, DBP and INT. Sections 3.3.2 and 3.3.1 present the structure of the linear sub-models in the one- and two-dimensional models showing the priors and hyperpriors of their parameters. Following, in Sections 3.4.1 and 3.4.2 the non-linear sub-models in both the one- and two-dimensional models are explained by describing prior and hyperprior distributions as well as the constraints that are applied for managing identifiability. Moreover, additional constraints for the non-linear sub-model which vary depending on the data/likelihood are presented. Next, we have the description of the covariate effects in Sections 3.5.1 and 3.5.2 and the analysis of the age sub-model described in Sections 3.6.1 and 3.6.2. In sections 3.7.1 and 3.7.2 the priors and hyperpriors of the σ^2 hyperparameter are shown in one and two dimensions. In Sections 3.8.1 and 3.8.2 the hyperparameter ν is presented whereas in Sections 3.9.1, 3.9.2 the hyperparameter τ^2 is presented in one and two dimensions. We can observe the analysis that was used for the hyperparameters of the sub-models. In summary, this chapter is a detailed guide to the hierarchical structure and specifications of the one- and two-dimensional models of blood pressure.

Chapter 4

Implementation of a Two-Dimensional Blood Pressure Model

4.1 Introduction

This chapter describes the details of the implementation of the model discussed in chapter 3. It provides the prior, posterior and proposal distributions that are used for each of the sub-models in the one- and two-dimensional models. We explain in detail the posterior distributions and we describe the techniques that are used for updating each of the parameters. Although the dimensionality in the one- and two-dimensional models for each of the parameters has changed, the posterior, proposal and the prior distributions are defined in the same way. Important differences are presented in the two dimensional non-linear model which is described in the following section. Each of the models use either Metropolis-Hastings (M-H) or Gibbs Sampler algorithms. We carry out block updating using the M-H algorithm in the linear and the non-linear sub-models which is highly important for our final results [Rue and Held, 2005, chapter 4]. Finally, we present the application of the two-dimensional model using simulated and real data.

4.1.1 Reparameterisation

In order to simplify the MCMC implementation we re-write the likelihood, combining terms as follows, Danaei et al. [2011], Finucane et al. [2011]:

$$\mathbf{y}_{h,i} \sim \mathcal{N}(\mathbf{F}_i\boldsymbol{\theta} + \mathbf{u}_{j[i],t_i} + \boldsymbol{\gamma}_i(z_h), \boldsymbol{\Sigma}) \tag{4.1}$$

where the vector $\boldsymbol{\theta}$ contains the a_j's, b_j's, $\boldsymbol{\beta}$'s and e_i's. We use I and J to indicate the number of studies and the number of countries respectively. In that way, $\mathbf{F}$ is a mapping matrix with I rows that contains the covariates $\mathbf{X}$ and that assigns each a_j, b_j, $\boldsymbol{\beta}$, e_i to the appropriate studies. Analytically, $\mathbf{F}$ assigns the countries that comprise a specific study each time. We use $\boldsymbol{\Sigma}$ to denote the $I \times I$ diagonal matrix of $\mathbf{SD}^2/n + \boldsymbol{\tau}^2$ values which expresses the residuals' variance within and between the age groups in a specific study i. For ease of notation, we denote, Danaei et al. [2011], Finucane et al. [2011]:

$$1)\ \boldsymbol{\psi} = (\psi_1, \ldots, \psi_5)^T, \quad 2)\ \boldsymbol{\phi} = (\phi_1, \ldots, \phi_5)^T$$
$$3)\ \mathbf{c} = [\mathbf{c}_1, \ldots, \mathbf{c}_5] \quad 4)\ \mathbf{Z} = [N_{0,3}(t), N_{1,3}(t), N_{2,3}(t), N_{3,3}(t), N_{4,3}(t)]$$

The dimensions of $\boldsymbol{\psi}$ and $\boldsymbol{\phi}$ are 5×1, $\mathbf{c}$ is of dimension $J \times 5$ and $\mathbf{Z}$ is of dimension $I \times 5$. In addition, we define two more mapping matrices: $\mathbf{C}$ (of dimension $I \times J$) and $\mathbf{M}$ (of dimension $I \times JT$) that assign the elements of $\mathbf{c}$ and of $\mathbf{u}$ respectively to the appropriate studies, Danaei et al. [2011], Finucane et al. [2011]. The sparse matrices, C and M assign the countries to a specific study that comprise each time. Letting "$*$" indicate component-wise matrix with vector multiplication, the modified version of the likelihood mean will be, Danaei et al. [2011], Finucane et al. [2011]:

$$
\begin{aligned}
E[\mathbf{y}|.] &= \mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u} + \psi_1 N_{0,3}(t) + \cdots + \psi_5 N_{4,3}(t) + (\mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u}) * \phi_1 N_{0,3}(t) \\
&\quad + \cdots + ((\mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u}) * \phi_5 N_{4,3}(t) + (\mathbf{C}_{c_1}) * N_{0,3}(t) + \cdots + (\mathbf{C}_{c_5}) * N_{4,3}(t) \\
E[\mathbf{y}|.] &= \mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u} + \mathbf{Z}\boldsymbol{\psi} + (\mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u}) * (\mathbf{Z}\boldsymbol{\phi}) + ((\mathbf{C}_c) * \mathbf{Z})1_5 \\
&= \mathbf{F}\boldsymbol{\theta} * (\mathbf{Z}\boldsymbol{\phi} + 1) + (\mathbf{M}\mathbf{u}) * (\mathbf{Z}\boldsymbol{\phi} + 1) + \mathbf{Z}\boldsymbol{\psi} + ((\mathbf{C}_c) * \mathbf{Z})1_5 \\
E[\mathbf{y}|.] &= (\mathbf{F} * (\mathbf{Z}\boldsymbol{\phi} + 1))\boldsymbol{\theta} + (\mathbf{M} * (\mathbf{Z}\boldsymbol{\phi} + 1))\mathbf{u} + \mathbf{Z}\boldsymbol{\psi} + ((\mathbf{C}_c) * \mathbf{Z})1_5
\end{aligned}
\tag{4.2}
$$

where $\mathbf{y}$ and $\mathbf{u}$ are vectors.

Having written the mean in a more intuitive form in (4.2), the distribution of the blood pressure model or of the likelihood can be expressed as:

$$
\begin{aligned}
\log P(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\gamma}, .) &\propto (\mathbf{y} - E[\mathbf{y}|.])^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - E[\mathbf{y}|.]) \\
\log P(\mathbf{y}|.) &\propto (\mathbf{y} - \mathbf{F} * (\mathbf{Z}\phi + 1)\boldsymbol{\theta} - \mathbf{M} * (\mathbf{Z}\phi + 1)\mathbf{u} - \mathbf{Z}\psi - ((\mathbf{C}_c) * \mathbf{Z})1_5)^T \\
&\quad \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{F} * (\mathbf{Z}\phi + 1)\boldsymbol{\theta} - \mathbf{M} * (\mathbf{Z}\phi + 1)\mathbf{u} - \mathbf{Z}\psi - ((\mathbf{C}_c) * \mathbf{Z})1_5)
\end{aligned}
\tag{4.3}
$$

4.2 Linear sub-model: posterior distribution

4.2.1 1D Linear sub-model: posterior distribution

For the posterior distribution of the linear sub-model we will take the joint distribution of the parameter $\boldsymbol{\theta}$ and its hyperparameter $\boldsymbol{\kappa}$ as was explained by Danaei et al and Finucane et al Danaei et al. [2011], Finucane et al. [2011]. Following Bayes' Theorem we have:

$$
Posterior = \frac{Likelihood \times Prior \times Hyperprior}{Likelihood}
$$

$$
P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y}) = \frac{P(\mathbf{y}|.)P(\boldsymbol{\theta}|\mathbf{V}_\theta)\prod P(\sqrt{\kappa})}{P(\mathbf{y}|.)} \tag{4.4}
$$

The likelihood is defined by (4.1) while the prior is the joint distribution of the linear sub-model's priors, explained in (3.10) equation and both are defined in Danaei et al. [2011], Finucane et al. [2011]. $\mathbf{V}_\theta$ is a diagonal matrix with the κ variances in the diagonal for each a^c, a^r, a^s, b^c, b^r, b^s parameters repeated three times for DBP, SBP and INT. Finally, we define a non-informative hyperprior for the standard deviation parameters, $\sqrt(\kappa)$. For simplicity, we apply the log function to each of our distributions, Danaei et al. [2011], Finucane et al. [2011].

The detailed computation of the posterior is presented in (4.5).

$$
\begin{aligned}
\log(\text{posterior}) &\propto \log(\text{likelihood}) + \log(\text{prior}) + \log(\text{hyper-prior}) \\
\log P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y}) &\propto \log P(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\gamma}, .) + \log P(\boldsymbol{\theta}|\mathbf{V}_\theta) + \sum \log P(\sqrt{\kappa}) \\
\log P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y}) &\propto (\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} - \mathbf{M}(\mathbf{Z}\phi + 1)\mathbf{u} - \mathbf{Z}\psi - (\mathbf{C}_c Z)1_5)^T \\
&\quad \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} - \mathbf{M}(\mathbf{Z}\phi + 1)\mathbf{u} - \mathbf{Z}\psi - (\mathbf{C}_c\mathbf{Z})1_5) \\
&\quad + \log P(\boldsymbol{\theta}|\mathbf{V}_\theta) + \sum \log P(\sqrt{\kappa})
\end{aligned} \tag{4.5}
$$

Keeping only the values that depend on $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ the posterior distribution is as shown in equation (4.6).

$$
\begin{aligned}
\log P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y}) \propto\ & (\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta}) \\
& - (\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1}\mathbf{y} + (\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{Z}\phi + 1)\mathbf{u} \\
& + (\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1}(\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5) \\
& - \mathbf{y}^T \boldsymbol{\Sigma}^{-1}\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} + (\mathbf{M} * (\mathbf{Z}\phi + 1)\mathbf{u})^T \boldsymbol{\Sigma}^{-1}\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} \\
& + (\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5)^T \boldsymbol{\Sigma}^{-1}(\mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta}) \\
& + \boldsymbol{\theta}^T \mathbf{V}_\theta^{-1}\boldsymbol{\theta} + \sum \log P(\sqrt{\kappa})
\end{aligned} \tag{4.6}
$$

Let us consider now that we want to find the posterior distribution solely of the $\boldsymbol{\theta}$ parameter. We rewrite in (4.7) the equation (4.6) in a way that is focusing only on the $\boldsymbol{\theta}$ parameter.

$$\begin{aligned}
\log P(\boldsymbol{\theta}|\mathbf{y}) \propto\ & \boldsymbol{\theta}^T \underline{(\mathbf{F}(\mathbf{Z}\phi+1))^T \boldsymbol{\Sigma}^{-1} \mathbf{F}(\mathbf{Z}\phi+1)} \boldsymbol{\theta} \\
& - \boldsymbol{\theta}^T \underline{(\mathbf{F}(\mathbf{Z}\phi+1))^T \boldsymbol{\Sigma}^{-1} [\mathbf{y} - \mathbf{M}(\mathbf{Z}\phi+1)\mathbf{u} - \mathbf{Z}\psi - ((\mathbf{C}_c)\mathbf{Z})1_5]} \\
& - \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta} + (\mathbf{M} * (\mathbf{Z}\phi+1)\mathbf{u})^T \boldsymbol{\Sigma}^{-1} \mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta} \\
& + (\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5)^T \boldsymbol{\Sigma}^{-1} (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta}) + \boldsymbol{\theta}^T \underline{\mathbf{V}_\theta^{-1}} \boldsymbol{\theta} \\
& + \sum \log P(\sqrt{\kappa})
\end{aligned} \tag{4.7}$$

Let us assume that in (4.7) we need to find the posterior distribution of $\boldsymbol{\theta}$. By showing an example of a summation of two bivariate normal distributions we can easily identify the posterior distribution in (4.8).

$$\begin{aligned}
\log(f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)) &\propto (\boldsymbol{\mu} - \mathbf{x})^T \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\mu} - \mathbf{x}) + (\mathbf{x} - \mathbf{0})^T \boldsymbol{\Sigma}_2^{-1} (\mathbf{x} - \mathbf{0}) \\
&\propto x^T \boldsymbol{\Sigma}_1^{-1} \mathbf{x} - \mathbf{x}^T \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{x} + \boldsymbol{\mu}^T \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu} + \\
&\mathbf{x}^T \boldsymbol{\Sigma}_2^{-1} \mathbf{x} \\
\log(f(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2)) &\propto \mathbf{x}^T \underline{(\boldsymbol{\Sigma}_1^{-1} + \boldsymbol{\Sigma}_2^{-1})} \mathbf{x} - \mathbf{x}^T \underline{\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}} - \boldsymbol{\mu}^T \boldsymbol{\Sigma}_1^{-1} \mathbf{x} + \\
&\boldsymbol{\mu}^T \boldsymbol{\Sigma}_1^{-1} \boldsymbol{\mu}
\end{aligned} \tag{4.8}$$

The underlined terms in (4.7) and (4.8) help to identify the mean vector and the precision matrix of the distribution. As our parameter is $\boldsymbol{\theta}$ we will have:

$$\begin{aligned}
&\boldsymbol{\theta}|\boldsymbol{\mu}_\theta, \mathbf{Q}_\theta \boldsymbol{\mu}_\theta \sim \mathcal{N}_c(\mathbf{Q}_\theta \boldsymbol{\mu}_\theta, \mathbf{Q}_\theta) \\
&\qquad \mathbf{Q}_\theta = (\mathbf{F}(\mathbf{Z}\phi+1))^T \boldsymbol{\Sigma}^{-1} \mathbf{F}(\mathbf{Z}\phi+1) + \mathbf{V}_\theta^{-1} \\
&\qquad \mathbf{Q}_\theta \boldsymbol{\mu}_\theta = (\mathbf{F}(\mathbf{Z}\phi+1))^T \boldsymbol{\Sigma}^{-1} [\mathbf{y} - \mathbf{M}(\mathbf{Z}\phi+1)\mathbf{u} - \mathbf{Z}\psi - ((\mathbf{C}_c)\mathbf{Z})1_5]
\end{aligned} \tag{4.9}$$

We use the canonical parametrization defined in Section 2.4 and also by Rue and Held [Rue and Held, 2005, chapter 2]. Thus, the mean and the precision matrix for the posterior distribution are as in (4.9). All this information comes from the likelihood, $P(\mathbf{y}|\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\gamma})$ and the prior, $P(\boldsymbol{\theta}|\mathbf{V}_\theta)$. We can see that the term $\mathbf{Q}_\theta \boldsymbol{\mu}_\theta$ does not include the prior's variance, $\mathbf{V}_\theta$ as the prior's mean is zero and therefore it cancels out. Finally, we can observe that the precision matrix of the prior distribution, $\mathbf{V}_\theta^{-1}$ contributes only to the precision matrix of the posterior and for the $\mathbf{Q}_\theta \boldsymbol{\mu}_\theta$ term only the likelihood's precision matrix is present.

In the whole analysis we use the theory of canonical parametrisation and Cholesky factorisation, which was presented in Section 2.4. Therefore, instead of estimating the mean, $\boldsymbol{\mu}$ and the precision matrix, $\mathbf{Q}$ of the $\boldsymbol{\theta}$ parameter we will estimate the product, $\mathbf{Q}\boldsymbol{\mu}$ as the mean and $\mathbf{Q}$ as the precision matrix for the canonical parametrisation of $\boldsymbol{\theta}$. The reason for using canonical parametrisation is that often it is cumbersome to compute a distribution's mean by itself. However, we can easily find it as a product with the precision matrix in the analytical form of the distribution.

4.2.1.1 Acceptance Ratio

In this linear sub-model we could simply use a Gibbs sampler since we could sample from the distributions of $P(\boldsymbol{\theta}|\boldsymbol{\kappa}, \mathbf{y})$ and $P(\boldsymbol{\kappa}|\boldsymbol{\theta}, \mathbf{y})$, which is known as a two-block Gibbs sampler. However, it is observed that the $\boldsymbol{\theta}$ parameter and the hyper-parameter $\boldsymbol{\kappa}$ are strongly dependent which has as a result slow mixing and convergence, [Rue and Held, 2005, chapter 4]. The two-block Gibbs sampler only moves either horizontally or vertically, [Rue and Held, 2005, chapter 4]. Therefore, a joint update of both $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ seems more appropriate [Rue and Held, 2005, chapter 4], Danaei et al. [2011], Finucane et al. [2011]. Hence, the posterior is the joint distribution of $\boldsymbol{\theta}$ and $\boldsymbol{\kappa}$ which can be expressed as the product of the likelihood, $\boldsymbol{\theta}$'s prior and the hyperprior of $\boldsymbol{\kappa}$. The priors of the intercept and slope (3.10) are expressed as a vector and hence a vector of six values will be shown for the $\boldsymbol{\kappa}$ hyperparameter. The proposal distribution of $\boldsymbol{\theta}$ is presented in Section 4.2 whereas the proposal distribution of the $\boldsymbol{\kappa}$ parameter is a symmetrical distribution which is explained below. Since $\boldsymbol{\kappa}$ is six-variate we can use a simple scheme for updating all $\boldsymbol{\kappa}$'s if we delay the accept/reject step until $\boldsymbol{\theta}$ is updated. The joint proposal distribution for each of the hyper-parameters and the parameter $\boldsymbol{\theta}$ is generated as follows, [Rue and Held, 2005, chapter 4] :

$$
\begin{aligned}
&\kappa_a^{c*} \sim q(\kappa_a^{c*}|\kappa_a^c), \kappa_b^{c*} \sim q(\kappa_b^{c*}|\kappa_b^c)\\
&\kappa_a^{r*} \sim q(\kappa_a^{r*}|\kappa_a^r), \kappa_b^{r*} \sim q(\kappa_b^{r*}|\kappa_b^r)\\
&\kappa_a^{s*} \sim q(\kappa_a^{s*}|\kappa_a^s), \kappa_b^{s*} \sim q(\kappa_b^{s*}|\kappa_b^s)\\
&\boldsymbol{\theta}^* \sim P(\boldsymbol{\theta}|\kappa^*, y)
\end{aligned}
\tag{4.10}
$$

Here q can be a simple random walk proposal or other suitable proposal distribution. A proposal from $\boldsymbol{\kappa}$ to $\boldsymbol{\kappa}^*$ may take $\boldsymbol{\kappa}^*$ out of the diagonal, while sampling $\boldsymbol{\theta}^*$ from $P(\boldsymbol{\theta}^*|\boldsymbol{\kappa}, \mathbf{y})$ will take it back into the diagonal again. Therefore, we have more options for the $\boldsymbol{\kappa}$ parameter keeping also its dependence since we are going back to the diagonal again, [Rue and Held, 2005, chapter 4]. Thus, the mixing and the convergence can be substantially improved, [Rue and Held, 2005, chapter 4].

The Metropolis-Hastings algorithm needs a target and a proposal distribution. Hence we will use the likelihood, prior and hyperprior presented in (4.5) which comprise the target distribution and (4.10) as a proposal distribution. We will need two functions for the proposal density; as stated in (4.10), one will account for the parameter, $\boldsymbol{\theta}$ and the other one will account for the hyperparameter $\boldsymbol{\kappa}$. If we consider that the q proposal is a symmetric distribution that accounts only for the hyperparameters then it will be cancelled out in the r ratio shown in (4.11). Hence, we will have as a proposal the density solely for the $\boldsymbol{\theta}$ parameter which is the distribution in (4.9).

$$
\begin{aligned}
r &= \frac{Posterior^* \times Proposal}{Posterior \times Proposal^*}\\
r &= \frac{(P(\boldsymbol{\theta}^*, \boldsymbol{\kappa}^*|\mathbf{y}))(P(\boldsymbol{\theta}|\boldsymbol{\kappa}, \mathbf{y})q(\boldsymbol{\kappa}|\boldsymbol{\kappa}^*))}{(P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y}))(P(\boldsymbol{\theta}^*|\boldsymbol{\kappa}^*, \mathbf{y})q(\boldsymbol{\kappa}^*|\boldsymbol{\kappa}))}\\
r &= \frac{P(\boldsymbol{\theta}^*, \boldsymbol{\kappa}^*|\mathbf{y})P(\boldsymbol{\theta}|\boldsymbol{\kappa}, \mathbf{y})}{P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y})P(\boldsymbol{\theta}^*|\boldsymbol{\kappa}^*, \mathbf{y})}\\
r &= \frac{P(\boldsymbol{\theta}^*, \boldsymbol{\kappa}^*|\mathbf{y})}{P(\boldsymbol{\theta}^*|\boldsymbol{\kappa}^*, \mathbf{y})}\frac{P(\boldsymbol{\theta}|\boldsymbol{\kappa}, \mathbf{y})}{P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y})}\\
r &= P(\boldsymbol{\kappa}^*|\mathbf{y})\frac{1}{P(\boldsymbol{\kappa}|\mathbf{y})}
\end{aligned}
\tag{4.11}
$$

In equation (4.11) we notice that only the density of $\boldsymbol{\kappa}$ is needed. Since we sample $\boldsymbol{\theta}$

from its full conditional, we effectively integrate $\boldsymbol{\theta}$ out of the joint density $P(\boldsymbol{\theta}, \boldsymbol{\kappa})$, [Rue and Held, 2005, chapter 4]. The minor modification to delay the accept/reject step until $\boldsymbol{\theta}$ is updated as well can give a large improvement to convergence [Rue and Held, 2005, chapter 4].

Finally, we accept $\boldsymbol{\theta}^*$ and the $\boldsymbol{\kappa}^*$'s with probability r, Finucane et al. [2011], Danaei et al. [2011]:

$$\begin{aligned} r &= \frac{P(\boldsymbol{\theta}^*, \boldsymbol{\kappa}^*|\mathbf{y})P(\boldsymbol{\theta}|\boldsymbol{\kappa}, \mathbf{y})}{P(\boldsymbol{\theta}, \boldsymbol{\kappa}|\mathbf{y})P(\boldsymbol{\theta}^*|\boldsymbol{\kappa}^*, \mathbf{y})} \\ r &= \frac{P(\mathbf{y}|\boldsymbol{\theta}^*, .)P(\boldsymbol{\theta}^*|\mathbf{V}_{\boldsymbol{\theta}}^*)\prod P(\sqrt{\boldsymbol{\kappa}^*})P(\boldsymbol{\theta}^{i-1}|\mathbf{V}_{\boldsymbol{\theta}}^{i-1}, \mathbf{y}, .)}{P(\mathbf{y}|\boldsymbol{\theta}^{i-1}, .)P(\boldsymbol{\theta}^{i-1}|\mathbf{V}_{\boldsymbol{\theta}}^{i-1})\prod P(\sqrt{\boldsymbol{\kappa}^{i-1}})P(\boldsymbol{\theta}^*|\mathbf{V}_{\boldsymbol{\theta}}^*, \mathbf{y}, .)} \end{aligned} \tag{4.12}$$

- Likelihood, target: $\log P(\mathbf{y}|\theta, .) \propto \frac{1}{2}(\sum \log diag(\boldsymbol{\Sigma}^{-1}) - (\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{Mu} - \mathbf{R}\boldsymbol{\gamma})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{Mu} - \mathbf{R}\boldsymbol{\gamma}))$
- Prior, target: $\log P(\boldsymbol{\theta}|\mathbf{V}_{\boldsymbol{\theta}}) \propto \frac{1}{2}(\sum \log diag(\mathbf{V}_{\boldsymbol{\theta}}^{-1}) - \boldsymbol{\theta}^T\mathbf{V}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\theta})$
- Hyper-prior, target: $\log P(\sqrt{\boldsymbol{\kappa}}) \propto 1$,
- Proposal: $\log P(\boldsymbol{\theta}|\mathbf{V}_{\boldsymbol{\theta}}, \mathbf{y}, .) \propto \dfrac{1}{2}(\sum \log diag(\mathbf{Q}_{\boldsymbol{\theta}}) - (\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}})^T\mathbf{Q}_{\boldsymbol{\theta}}(\boldsymbol{\theta} - \boldsymbol{\mu}_{\boldsymbol{\theta}}))$

In conclusion, we use a joint distribution of $(\boldsymbol{\theta}, \boldsymbol{\kappa})$ as a target distribution to tackle the dependence problem between the hyper-parameter and parameter of the linear model. However, the proposal distribution will solely depend on the parameter $\boldsymbol{\theta}$ and the reason is that the hyperparameter $\boldsymbol{\kappa}$ has a symmetric distribution which is cancelled out in the r ratio. This algorithm is known as a block update Metropolis Hastings, [Rue and Held, 2005, chapter 4].

The two-dimensional equivalent linear sub-model, defined in (3.12) will still have the $\mathbf{F}$ mapping matrix and the $\boldsymbol{\theta}$ parameter. However, in this case we need to have linear changes in time for all three variables. Therefore, we will have intercepts and slopes for the globe, all the super-regions, regions and countries for each of the variables under study, DBP, SBP, INT. The $\mathbf{F}$ matrix also includes the covariate effects presented in Section 3.5.2 which are tripled because of the three variables under study. Therefore, the $\mathbf{F}$ matrix and the $\boldsymbol{\theta}$ parameter will be defined with three times more dimensions than for the one-dimensional linear sub-model.

4.3 Non-linear: Posterior Distribution

4.3.1 1D Non-linear: posterior distribution

The case of the non-linear model is similar to the linear model. We analyse again the likelihood and we keep the terms that are related to **Mu** and λ this time. For convenience we will only see the procedure about M_c and λ_c which are the terms for the country's level. The same applies to region, super-region and globe level. In (4.13) we present the posterior distribution using the log function in likelihood, prior and hyperprior distribution.

$$
\begin{aligned}
\log(\text{posterior}) &\propto \log(\text{likelihood}) + \log(\text{prior}) + \log\,(\text{hyper-prior}) \\
\log P(\mathbf{u}^c, \lambda_c|\mathbf{y}) &\propto \log P(\mathbf{y}|\mathbf{u}^c, \lambda_c) + \log P(\mathbf{u}^c|\lambda_c) + \sum \log P(\lambda_c) \\
\log P(\mathbf{u}^c, \lambda_c|\mathbf{y}) &\propto (\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} - \mathbf{M}^c(\mathbf{Z}\phi + 1)\mathbf{u}^c - \mathbf{Z}\psi - (\mathbf{C}_c\mathbf{Z})1_5)^T \qquad (4.13) \\
&\quad \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} - \mathbf{M}^c(\mathbf{Z}\phi + 1)\mathbf{u}^c - \mathbf{Z}\psi - (\mathbf{C}_c\mathbf{Z})1_5) \\
&\quad + \log P(\mathbf{u}^c|\lambda_c) + \sum \log P(\lambda_c)
\end{aligned}
$$

Having the joint posterior distribution of $\mathbf{u}^c, \lambda_c|\mathbf{y}$ in (4.13) we also want to find the posterior distribution of the $\mathbf{u}^c|y$ parameter. Equation (4.14) presents a detailed version of (4.13) emphasizing to the u^c parameter.

$$
\begin{aligned}
\log P(\mathbf{u}^c|\mathbf{y}) \propto\ & \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y} - \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta} - \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{M}(\mathbf{Z}\phi+1)\mathbf{u}^c - \mathbf{y}^T\boldsymbol{\Sigma}^{-1} \\
& (\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5) - (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}\mathbf{y} + (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1} \\
& (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta}) + (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}\mathbf{M}^c(\mathbf{Z}\phi+1)\mathbf{u}^c + \\
& (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}(\mathbf{Z}\psi + ((C_c)\mathbf{Z})1_5) - (\mathbf{u}^c)^T\underline{(\mathbf{M}^c(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}\mathbf{y}} \\
& + (\mathbf{u}^c)^T\underline{(\mathbf{M}^c(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta}} + \\
& (\mathbf{u}^c)^T\underline{(\mathbf{M}^c*(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}\mathbf{M}^c(\mathbf{Z}\phi+1)}\mathbf{u}^c + (\mathbf{u}^c)^T\underline{(\mathbf{M}^c(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}} \\
& \underline{(\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5)} - (\mathbf{Z}\psi + ((\mathbf{C}_c)Z)1_5)^T\boldsymbol{\Sigma}^{-1}y + \\
& (\mathbf{Z}\psi + ((C_c)\mathbf{Z})1_5)^T\boldsymbol{\Sigma}^{-1}(\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta}) + \\
& (\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5)^T\boldsymbol{\Sigma}^{-1}(\mathbf{M}^c(\mathbf{Z}\phi+1)\mathbf{u}) + \\
& (\mathbf{Z}\psi + ((C_c)\mathbf{Z})1_5)^T\boldsymbol{\Sigma}^{-1}(\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5) + (\mathbf{u}^c)^T\underline{I_j \otimes \lambda_c\mathbf{P}}\mathbf{u}^c + \sum \log P(\lambda_c)
\end{aligned}
\tag{4.14}
$$

Bearing in mind the bivariate normal distribution in (4.8) and that our parameter now is $\mathbf{u}^c$, from (4.14) we can derive the following:

$$
\begin{aligned}
\log P(\mathbf{u}^c|\lambda_c, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\gamma}) \propto\ & (\mathbf{u}^c)^T\underline{(\mathbf{M}^c*(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}(\mathbf{M}^c*(\mathbf{Z}\phi+1))}\mathbf{u}^c + (\mathbf{u}^c)^T\underline{I_j \otimes \lambda_c\mathbf{P}}\mathbf{u}^c \\
& - (\mathbf{u}^c)^T(\mathbf{M}^c*(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta} - (\mathbf{Z}\psi - ((\mathbf{C}_c)\mathbf{Z})1_5)] \\
& - \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{M}^c(\mathbf{Z}\phi+1)\mathbf{u}^c + (\mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta})^T\boldsymbol{\Sigma}^{-1}\mathbf{M}^c(\mathbf{Z}\phi+1)\mathbf{u}^c \\
& + (\mathbf{Z}\psi + ((\mathbf{C}_c)\mathbf{Z})1_5)^T\boldsymbol{\Sigma}^{-1}(\mathbf{M}^c*(\mathbf{Z}\phi+1)\mathbf{u}^c)
\end{aligned}
\tag{4.15}
$$

Therefore, we can see the distribution of $\mathbf{u}^c|.$:

$$
\begin{aligned}
& \mathbf{u}^c|\boldsymbol{\mu}_{u^c}, \mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c} \sim \mathcal{N}_c(\mathbf{Q}_{u_c}\boldsymbol{\mu}_{u^c}, \mathbf{Q}_{u^c}) \\
& \qquad \mathbf{Q}_{u^c} = \mathbf{M}^c*(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}(\mathbf{M}^c*(\mathbf{Z}\phi+1)) + I_j \otimes \lambda_c\mathbf{P} \\
& \qquad \mathbf{Q}_{u_c}\boldsymbol{\mu}_{u^c} = (\mathbf{M}^c*(\mathbf{Z}\phi+1))^T\boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi+1)\boldsymbol{\theta} - (\mathbf{Z}\psi - ((\mathbf{C}_c)\mathbf{Z})1_5)]
\end{aligned}
\tag{4.16}
$$

The first part of the precision in (4.16) is the likelihood's variance including only the variables related to $\mathbf{u}^c$ and the second part is the hyper-prior's variance which is dependent on λ_c. The mean and the precision matrix of the non-linear sub-model are also stated in Danaei et al. [2011], Finucane et al. [2011].

4.3.1.1 Proposal distribution

i) Identifiability constraints $\sum_i u_{[i]} = 0$

We need the linear and the non-linear terms to be identifiable. To do so, we introduce a dependence between the intercept and the linear trend of the linear sub-model and the intercept and slope of the non-linear sub-model Danaei et al. [2011], Finucane et al. [2011]. In order to achieve that we will implement a restriction; the summation of the intercept of the non-linear model should be equal to zero, $\sum_i u_{[i]} = 0$ and also the summation of the linear trend should be equal to zero, $\sum_i iu_{[i]} = 0$ where i express the years. A way to apply this restriction is by using the linear constraints of a GMRF. One approach is to compute the mean and the covariance from the joint distribution of $\mathbf{u}$ (variable) and $\mathbf{Au}$ (linear constraint). We condition on $\mathbf{Au} = \mathbf{e}$ which leads to the conditional moments where $E(\mathbf{u}|\mathbf{Au})$, $Cov(\mathbf{u}|\mathbf{Au})$ are

$$E(\mathbf{u}|\mathbf{Au}) = \boldsymbol{\mu}^*_{\mathbf{u}^c_j} = \boldsymbol{\mu} - \mathbf{Q}^{-1}\mathbf{A}^T(\mathbf{AQ}^{-1}\mathbf{A}^T)^{-1}(\mathbf{A}\boldsymbol{\mu} - \mathbf{e}) \tag{4.17}$$

$$Cov(\mathbf{u}|\mathbf{Au}) = \mathbf{V}^*_{\mathbf{u}^c_j} = \mathbf{Q}^{-1} - \mathbf{Q}^{-1}\mathbf{A}^T(\mathbf{AQ}^{-1}\mathbf{A}^T)^{-1}\mathbf{AQ}^{-1} \tag{4.18}$$

ii) Linear constraints in IGMRFs

The linear constraint ensures that the conditional distribution is normal but singular as the rank of the constrained covariance matrix is $T-2$. As the rank is $T-2$ this indicates that for the second-order IGMRF we will have two linear constraints. In order to produce a proper joint distribution we impose the constraints $\sum_i u_{[i]} = 0$ and $\sum_i iu_{[i]} = 0$ which also give a finite marginal standard deviation. In the same way, for the first-order IGMRF the rank is $T-1$ and there is only one restriction that we impose, $\sum_i u_{[i]} = 0$. In general, depending on the type of

IGMRF that we use we need to impose the corresponding constraints, Lehmann and Casella [2006].

In this section we explore further the proposal distribution, $P(\mathbf{u}^{c(i-1)}|\mathbf{A}\mathbf{u}_j^{c(i-1)} = 0\forall j, \lambda_c^{i-1}, \mathbf{y})$ stressing the linear restrictions, $Au_j^{c(i-1)} = 0$. The difference between the posterior/target distribution and the proposal is that the first is referring to a joint distribution of $(\mathbf{u}^c, \lambda_c)$ whereas the latter refers to the marginal of the $\mathbf{u}^c$ parameter imposing restrictions on it. In the $\mathbf{Q}_{u^c}$ matrix we can see that for each country we add all the studies that have been conducted for a specific year. $\mathbf{Q}_{u^c}$ is the precision matrix which is a quadratic and a block diagonal matrix. Each block accounts for a specific country and thus, in total we have J blocks, one for each country in the diagonal. In $\mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c}$ is presented again the number of the studies in each country and year but not in a quadratic form. The procedure that we follow divides the countries (also regions, super-regions) into three scenarios: 1) when we have no data 2) data for 1 year and 3) data for more years for a specific country. For each of these scenarios the proposal distribution will have some differences. Firstly, we start with the λ's proposal which is a symmetric one and therefore it cancels out in the r ratio. Propose λ_c from:

$$\lambda_c^* \sim \mathcal{N}(a_5, b_5)$$

Then, we need to explore the analysis that each scenario needs:

1. **No Data were observed in country** j

   There are no data so the likelihood will not contribute and $\mathbf{Q}_{u^c}$ depends solely on the prior. We will take into account only the second part of the precision, Equation (4.16), $\mathbf{I}j \otimes \lambda\mathbf{P}$ which means dependence through the $\mathbf{P}$ matrix. Hence, the rank of $\mathbf{P}$ stays constant, $T-2$ corresponding to infinite prior variance on the mean (intercept) and the linear trend of $\mathbf{u}_j^c$. We constrain these two linear combinations of $\mathbf{u}_j^c$ to zero by taking the generalized inverse of $\mathbf{P}$, settings $\mathbf{P}$'s last two eigenvalues to infinity. Only one type of the constraints is applicable, the two restrictions that the second order IGMRF imposes, [Rue and Held, 2005, chapter 2], Damien et al. [2013], Finucane et al. [2011]. The steps to follow are presented below:

(a) $\mathbf{u}_j^{c*} \sim \mathcal{N}(\mathbf{0}, (\lambda_c \mathbf{P})^{-1})$

(b) Two restrictions of using the second order IGMRF (RW2). The restrictions apply by eigendecomposing the $\mathbf{P}$ matrix. Therefore, we will use the $\mathbf{P}$ matrix through the eigenvalues and eigenvectors. $\mathbf{P} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$ or $\mathbf{P}^{-1} = \mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^T$ such that $\Lambda_{ii}^{-1} = \frac{1}{\Lambda_{ii}}$ for $i = 1, \ldots, T-2$ and $\Lambda_{ii}^{-1} = 0$ for $i = T-1$, T. Again, these two restrictions occur because of the undefined intercept and linear trend.

(c) Generate $\mathbf{u}_j^c$.
$\mathbf{u}_j^{c*} \sim \mathcal{N}(\mathbf{0}, (\lambda_c \mathbf{P})^{-1})$, where $\lambda_c \mathbf{P}$ is the precision matrix. Therefore, instead of dividing elements we multiply them.
$\mathbf{u}_j^{c*} \sim \frac{1}{\lambda_c^{1/2}} \mathcal{N}(\mathbf{0}, \mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^T) \Rightarrow \mathbf{u}_j^{c*} \sim \frac{1}{\lambda_c^{1/2}} \mathcal{N}(\mathbf{0}, \Gamma\Lambda^{-1/2}\Lambda^{-1/2}\Gamma^T)$
$\mathbf{u}_j^{c*} \sim \frac{\Gamma\Lambda^{-1/2}}{\lambda_c^{1/2}} \mathcal{N}(0, 1)$
So, we generate the $\mathbf{u}_j^c$ parameter from: $\frac{\Gamma\Lambda^{-1/2}}{\lambda_c^{1/2}} z$ where $z \sim \mathcal{N}(0, 1)$

(d) Calculate:

$$\begin{aligned}
\log P(\mathbf{u}_j^{c*} | \lambda_c^*, \mathbf{y}, \mathbf{A}\mathbf{u}_j^{c*} = 0, \forall j.) &= -\frac{T-2}{2} \log 2\pi + \frac{T-2}{2} \log(\det(\lambda_c^* P)) \\
&\quad - \frac{\lambda_c^*}{2} \mathbf{u}_j^{c*T} \mathbf{P} \mathbf{u}_j^{c*} \\
&\propto \frac{T-2}{2} \log \lambda_c^* - \frac{1}{2} \mathbf{u}_j^{c*T} \lambda_c^* \mathbf{P} \mathbf{u}_j^{c*}
\end{aligned}$$

Summing up, with no information from the data the estimation is solely based on the prior. The prior is a second order IGMRF and thus the constraints will be $k = 2$ based on the linear constraints of the IGMRF as a prior, $T-2$. Again, $u_j^c \sim \mathcal{N}(0, (\lambda_c \mathbf{P})^{-1})$, where $\lambda_c \mathbf{P}$ is the precision matrix.

2. **Data for one year were observed in country** j
$\mathbf{Q}_{u^c}$ has rank $T-1$ because the mean (intercept) but not the linear trend of $\mathbf{u}_j^c$ is defined by the data. Hence, we constrain the linear trend $\mathbf{u}_j^c$ to zero by taking the generalized inverse of $\mathbf{Q}_{u^c}$, setting $\mathbf{Q}_{u^c}$'s last eigenvalue to ∞. It is necessary to apply the restriction of an infinite slope that the second order

IGMRF imposes, Danaei et al. [2011], Finucane et al. [2011]. In addition, we will use the identifiability constraint on the summation of the intercept which is defined by the data. To do that, we use the "Rue corrections", [Rue and Held, 2005, chapter 2] to constrain the mean of $\mathbf{u}_j^c$'s intercept to zero. Note that we apply two different constraints in this case. Steps:

(a) Eigendecompose $\mathbf{Q}_{u_j^c} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$

(b) Calculate the generalized inverse of $\mathbf{Q}_{u_j^c}$: $\mathbf{Q}_{u_j^c}^{-1} = \mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^T$ such that $\Lambda_{ii}^{-1} = 1/\Lambda_{ii}$ for $i = 1, ..., T-1$ and $\Lambda_{ii}^{-1} = 0$ for $i = T$.
The restriction is referring only to the undefined linear trend.

(c) Calculate:
$\mathbf{V}_{u_j^c}^* = \mathbf{Q}_{u_j^c}^{-1} - \mathbf{Q}_{u_j^c}^{-1}1^T(1\mathbf{Q}_{u_j^c}^{-1}1^T)^{-1}1\mathbf{Q}_{u_j^c}^{-1}$
$\boldsymbol{\mu}_{u_j^c}^* = \mathbf{Q}_{u_j^c}^{-1}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c} - \mathbf{Q}_{u_j^c}^{-1}1^T(1\mathbf{Q}_{u_j^c}^{-1}1^T)^{-1}1\mathbf{Q}_{u_j^c}^{-1}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c}$
A new variance and mean are calculated. Since only the intercept can be identifiable from the data there is one identifiability restriction and it can be adjusted as: $A = 1^T$ and $e = 0$. The restriction is defined as: $1^T\mathbf{u}_j^c = 0 \Rightarrow u_{j[1]}^c + \cdots + u_{j[T]}^c = 0$.

(d) Eigendecompose $\mathbf{V}_{u_j^c}^* = \mathbf{\Gamma}\mathbf{\Lambda}^{-}\mathbf{\Gamma}^T$ such that:
$\Lambda_{ii}^{-1} = 1/\Lambda_{ii}$ for $i = 1, \ldots, T-2$ and $\Lambda_{ii}^{-} = 0$ for $i = T-1, T$.
The two restrictions of the RW2 have been added to the new variance.

(e) Generate $\mathbf{u}_j^{c*} = \boldsymbol{\mu}_{u_j^c}^* + \mathbf{\Gamma}\mathbf{\Lambda}^{-1/2}z$ with $z \sim \mathcal{N}(0,1)$

(f) Calculate:

$$\begin{aligned}\log P(\mathbf{u}_j^{c*}|\lambda_c^*, \mathbf{A}\mathbf{u}_j^{c*} = 0, \mathbf{y}) &= -\frac{T-2}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{T-2}\log\Lambda_{ii}^{-1} \\ &\quad - \frac{1}{2}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}_{u_j^{c*}})^T\mathbf{V}_{u_j^c}^{*-1}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}_{u_j^{c*}}) \\ &\propto -\frac{1}{2}\sum_{i=1}^{T-2}\log\Lambda_{ii}^{-1} - \frac{1}{2}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}_{u_j^{c*}})^T\mathbf{V}_{u_j^c}^{*-1}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}_{u_j^{c*}})\end{aligned}$$

3. **Data for multiple years were observed in country j**

$\mathbf{Q}_{u^c}$ has rank T because the mean (intercept) and the linear trend of $\mathbf{u}_j^c$ are identified by the data. However, we use the "Rue corrections" to constrain the mean of $\mathbf{u}_j^c$'s intercept and $\mathbf{u}_j^c$'s linear trend to zero (identifiability constraints) [Rue and Held, 2005, chapter 2]. Afterwards, to the modified precision matrix we apply the restrictions of the second order IGMRF. In this case only one constraint is applied. Steps:

(a) Calculate $\mathbf{V}_{u_j^c} = \mathbf{Q}_{u_j^c}^{-1}$

(b) Calculate $\boldsymbol{\mu}^*_{u_j^c} = \mathbf{V}_{u_j^c}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c} - \mathbf{V}_{u_j^c}\mathbf{A}^T(\mathbf{A}\mathbf{V}_{u_j^c}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{V}_{u_j^c}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c}$
Calculate $\mathbf{V}^*_{u_j^c} = \mathbf{V}_{u_j^c} - \mathbf{V}_{u_j^c}\mathbf{A}^T(\mathbf{A}\mathbf{V}_{u_j^c}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{V}_{u_j^c}$
A new variance and mean are calculated. Since both intercept and slope are identifiable, $\mathbf{A}$ will be a $2 \times T$ matrix which will include $\sum_i u_{[i]} = 0$, the summation of intercepts and also $\sum_i iu_{[i]} = 0$, the summation of slopes, which are referring to the identifiability constraints.

(c) Eigendecompose $\mathbf{V}^*_{u_j^c} = \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^T$ such that:
$\Lambda_{ii}^{-1} = 1/\Lambda_{ii}$ for $i = 1, \dots, T-2$ and $\Lambda_{ii}^{-1} = 0$ for $i = T-1, T$
The two restrictions of the RW2 have been added to the new variance.

(d) Generate $\mathbf{u}_j^c = \boldsymbol{\mu}_{u_j^c} + \boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1/2}z$ with $z \sim \mathcal{N}(0, 1)$

(e) Defining $\mathbf{V}^*_{u_j^c}$ as above, calculate:

$$\begin{aligned}
\log P(\mathbf{u}_j^{c*}|\lambda_c^*, \mathbf{A}\mathbf{u}_j^{c*} = 0, \mathbf{y}) &= -\frac{T-2}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{T-2}\log\Lambda_{ii}^{-1} \\
&\quad - \frac{1}{2}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}^*_{u_j^c})^T\mathbf{V}^{*-1}_{u_j^c}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}^*_{u_j^c}) \\
&\propto -\frac{1}{2}\sum_{i=1}^{T-2}\log\Lambda_{ii}^{-1} - \frac{1}{2}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}^*_{u_j^c})^T\mathbf{V}^{*-1}_{u_j^c}(\mathbf{u}_j^{c*} - \boldsymbol{\mu}^*_{u_j^c})
\end{aligned}$$

In summary, we need to highlight that by using the second order or any other IGMRF there are restrictions to be considered. In the case of the second-order IGMRF there are two restrictions to apply. However, the restrictions that the IGMRF imposes vary depending on the dataset we have for our parameter.

Moreover, due to the construction of the model between the linear sub-model and the non-linear sub-model we need to apply also identifiability constraints which are based on the intercept and slope of the non-linear model. These are known as linear constraints or "Rue constraints", [Rue and Held, 2005, chapter2].

Finally, after considering all the aforementioned we need to impose for the last time the two restrictions of the second order IGMRF to the new structure matrix.

4.3.1.2 Acceptance Ratio

Again, in the non-linear model we could use a Gibbs sampler since we could sample from the distributions of $P(\mathbf{u}^c|\lambda_c, \mathbf{y})$, $P(\mathbf{u}^r|\lambda_r, \mathbf{y})$, $P(\mathbf{u}^s|\lambda_s, \mathbf{y})$, $P(\mathbf{u}^g|\lambda_g, \mathbf{y})$ and $P(\lambda_c|\mathbf{u}^c, \mathbf{y})$, $P(\lambda_r|\mathbf{u}^r, \mathbf{y})$, $P(\lambda_s|\mathbf{u}^s, \mathbf{y})$, $P(\lambda_g|\mathbf{u}^g, \mathbf{y})$ respectively. However, due to the high dependence of the $\mathbf{u}$ parameter and the hyper-parameter λ, in each of the four cases, this results in slow mixing in the chain, Danaei et al. [2011], Finucane et al. [2011]. Therefore, instead of using a two-block Gibbs sampler we will update the $(\mathbf{u}, \lambda)$ jointly as we first saw in Section 4.2.1.1 .
Since λ is univariate for each of the country's, region's, super-region's and globe's hyper-parameter, we can use a simple scheme for updating all the λ's as long as we delay the accept/reject step until $\mathbf{u}$ is updated. The joint proposal is generated as follows, [Rue and Held, 2005, chapter 4]:

$$\begin{aligned}
\lambda_c^* &\sim q(\lambda_c^*|\lambda_c), \quad \mathbf{u}^{c*} \sim P(\mathbf{u}^c|\lambda_c^*, \mathbf{y}) \\
\lambda_r^* &\sim q(\lambda_r^*|\lambda_r), \quad \mathbf{u}^{r*} \sim P(\mathbf{u}^r|\lambda_r^*, \mathbf{y}) \\
\lambda_s^* &\sim q(\lambda_s^*|\lambda_s), \quad \mathbf{u}^{s*} \sim P(\mathbf{u}^s|\lambda_s^*, \mathbf{y}) \\
\lambda_g^* &\sim q(\lambda_g^*|\lambda_g), \quad \mathbf{u}^{g*} \sim P(\mathbf{u}^g|\lambda_g^*, \mathbf{y})
\end{aligned} \tag{4.19}$$

Note that in equations (4.19) each line is updated separately. Specifically, we will update the following four pairs: $(\mathbf{u}^c, \lambda_c)$, $(\mathbf{u}^r, \lambda_r)$, $(\mathbf{u}^s, \lambda_s)$ and $(\mathbf{u}^g, \lambda_g)$. We will describe the acceptance ratio for the group $(\mathbf{u}_c, \lambda_c)$. However, the same methodology is used for the other three pairs as well. We have

$$\begin{aligned}
r &= \frac{\text{Posterior}^* \times \text{Proposal}}{\text{Posterior} \times \text{Proposal}^*} \\
r &= \frac{\text{Posterior}^* \times \text{Proposal}(1) \times \text{Proposal}(2)}{\text{Posterior} \times \text{Proposal}(1)^* \times \text{Proposal}(2)^*} \\
r &= \frac{P(\mathbf{u}_c^*, \lambda_c^*|\mathbf{y}) P(\mathbf{u}_c|\lambda_c) q(\lambda_c|\lambda_c^*)}{P(\mathbf{u}_c, \lambda_c|\mathbf{y}) P(\mathbf{u}_c^*|\lambda_c^*) q(\lambda_c^*|\lambda_c)}
\end{aligned} \tag{4.20}$$

and if we consider that the q proposal in (4.19) is a symmetric distribution then it will be cancelled out in the r ratio.

$$\begin{aligned} r &= \frac{P(u_c^*, \lambda_c^*|y)P(u_c|\lambda_c, y)}{P(u_c, \lambda_c|y)P(u_c^*|\lambda_c^*, y)} \\ r &= \frac{P(u_c^*, \lambda_c^*|y)}{P(u_c^*|\lambda_c^*, y)} \frac{P(u_c|\lambda_c, y)}{P(u_c, \lambda_c|y)} \\ r &= P(\lambda_c^*|y)\frac{1}{P(\lambda_c|y)} \end{aligned} \tag{4.21}$$

Note that only the density of λ_c is needed in (4.21), [Rue and Held, 2005, chapter 4]. Since we sample $\mathbf{u}_c$ from its full conditional, we effectively integrate $\mathbf{u}_c$ out of the joint density $P(\mathbf{u}_c, \lambda_c|.)$. The minor modification to delay the accept/reject step until $\mathbf{u}_c$ is updated as well can improve substantially the time needed for convergence and the dependence between the parameter and the hyperparameter, [Rue and Held, 2005, chapter 4]. In other words, we update the λ_c hyperparameter first which practically affects the update of u_c which gives time for appropriate updating.

We accept $\mathbf{u}_c^*$ and λ_c^*'s with probability r:

$$r = \frac{P(\mathbf{y}|\mathbf{u}_c^*)P(\mathbf{u}_c^*|\lambda_c^*)P(\log\lambda_c^*)P(\mathbf{u}_c^{(i-1)}|\mathbf{A}\mathbf{u}_{jc}^{(i-1)} = 0\forall j, \lambda_c^{i-1}, \mathbf{y})}{P(\mathbf{y}|\mathbf{u}_c^{(i-1)})P(\mathbf{u}_c^{(i-1)}|\lambda_c^{i-1})P(\log\lambda_c^{i-1})P(\mathbf{u}_c^*|\mathbf{A}\mathbf{u}_j^{c*} = 0\forall j, \lambda_c^*, \mathbf{y})} \tag{4.22}$$

- Likelihood, target: $\log P(\mathbf{y}|\mathbf{u}_c^*) \propto \frac{1}{2}(\sum \log \operatorname{diag}(\boldsymbol{\Sigma}^{-1}) - (\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{M}\mathbf{u} - \mathbf{R}\boldsymbol{\gamma})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{M}\mathbf{u} - \mathbf{R}\boldsymbol{\gamma}))$
- Prior, target: $\log P(\mathbf{u}_c^*|\lambda_c^*) \propto \frac{1}{2}(\sum \log \operatorname{diag}(\lambda_c^*) - (\mathbf{u}_c^*)^T(\lambda_c^*)^{-1}\mathbf{u}_c^*)$
- Hyper-prior, target: $P(\log\lambda_c^*) \propto 1$
- Proposal: $P(\mathbf{u}_c^{(i-1)}|\mathbf{A}\mathbf{u}_{c[j]}^{(i-1)} = 0\forall j, \lambda_c^{i-1}, \mathbf{y}) \propto \dfrac{1}{2}(\sum \log \operatorname{diag}(\mathbf{Q}_{\mathbf{u}_c}) - (\mathbf{u}_c - \boldsymbol{\mu}_{\mathbf{u}_c})^T\mathbf{Q}_{\mathbf{u}_c}(\mathbf{u}_c - \boldsymbol{\mu}_{\mathbf{u}_c}))$

The proposal distribution depends each time on the data of each country (whether there is no data, 1 year's data or more years' data). Since the q density of (4.19) is a symmetric distribution we have as a proposal solely the $\mathbf{u}_c$ density.

4.3.2 2D: Non-linear sub-model

In this section we present the two-dimensional non-linear sub-model which has been a key part of our methodological development. The variables of interest are DBP, SBP and their interaction, INT. In the same way as for the one-dimensional non-linear sub-model in Section 4.3.1, for the posterior distribution of the non-linear model we use a block Metropolis-Hastings algorithm which is a joint distribution of the parameter $\mathbf{u}^i_{(D,S)}$ and its hyper-parameter λ_i where $i \in \{c, r, s, g\}$. The reason for using a joint distribution is to speed up convergence and because of the high correlation between the parameter and the hyper-parameter. By having the likelihood and taking only the terms that are related to $\mathbf{u}^i_{(D,S)}$, λ_i we specify the full conditional normal distribution for each component of $\mathbf{u}^i$ separately. Here, we describe the posterior at the country level of the hierarchy, but the posteriors for the other levels are analogous. For $\mathbf{u}^c_{(D,S)}$, we have:

$$\begin{aligned}
\log P(\mathbf{u}^c_{(D,S)}, \lambda_c|\mathbf{y}) &\propto \log P(y|(\mathbf{u}^c_{(D,S)}, \lambda_c, .) + \log(\mathbf{u}^c_{(D,S)}|\lambda_c) + \log P(\lambda_c) \\
\log P(\mathbf{u}^c_{(D,S)}, \lambda_c|\mathbf{y}) &\propto (\mathbf{y} - \mathbf{F}(\mathbf{Z}\boldsymbol{\phi} + 1)\boldsymbol{\theta} - \mathbf{M}_c(\mathbf{Z}\boldsymbol{\phi} + 1)\mathbf{u}^c_{(D,S)} - \mathbf{Z}\boldsymbol{\psi} \\
&\quad - (\mathbf{C}_c\mathbf{Z})1_5)^T\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}(\mathbf{Z}\boldsymbol{\phi} + 1)\boldsymbol{\theta} - \mathbf{M}_c(\mathbf{Z}\boldsymbol{\phi} + 1)\mathbf{u}^c_{(D,S)} \\
&\quad - \mathbf{Z}\boldsymbol{\psi} - (\mathbf{C}_c\mathbf{Z})1_5) + \log(P(\mathbf{u}^c_{(D,S)}|\lambda_c)) + \log P(\lambda_c)
\end{aligned} \tag{4.23}$$

For convenience, instead of writing $\mathbf{u}^c_{(D,S)}$ we will write $\mathbf{u}^c$ instead throughout this section. For the posterior of the non-linear model, we will keep only the $\mathbf{u}^c$ and $\mathbf{M}_c$ terms that comprise the non-linear model, and we can state the posterior distribution using canonical parameterisation, $\mathbf{u}^c, \lambda_{u^c}|\boldsymbol{\mu}_{u^c}, \mathbf{Q}_{u^c} \sim N_c(\mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c}, \mathbf{Q}_{u^c})$, which we have seen in (4.15) in Section 4.3.1. Therefore, the distribution of $\mathbf{u}^c$ is:

$$\begin{aligned}
\mathbf{Q}_{u^c} &= (\mathbf{M}_c * (\mathbf{Z}\phi + 1))^T\boldsymbol{\Sigma}^{-1}(\mathbf{M}_c * (\mathbf{Z}\phi + 1)) + I_j \otimes \lambda_c\mathbf{P} \\
\mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c} &= (\mathbf{M}_c * (\mathbf{Z}\phi + 1))^T\boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} - (\mathbf{Z}\psi - ((\mathbf{C}c)\mathbf{Z})1_5)]
\end{aligned} \tag{4.24}$$

where $\mathbf{Z}$ expresses the age data, $\boldsymbol{\psi}$ expresses the constant term, $\boldsymbol{\phi}$ is the mean term and $\mathbf{C}$ are country-specific random effects in the age model.

Table 4.1: Mapping matrix which assigns each country × year to the appropriate studies

	$A'01$	$A'02$	$A'03$	$A'04$	$A'05$
study1	1				
study2	1				
study3					1

At this stage it is important to stress the differences from the one-dimensional non-linear sub-model, 4.3.1. Since we have three variables, DBP, SBP and their interaction, the precision matrix $\lambda_c\mathbf{P}$ will be the one presented in (2.31). The dimensionality of $\mathbf{u}^c$ and the $\mathbf{P}$ matrix will be changed. The $\mathbf{u}^c$ will be a vector of T^2 values and $\mathbf{P}$ a $T^2 \times T^2$ matrix whereas in the one-dimensional equivalent we have a vector of size T and a $T \times T$ matrix respectively.

4.3.2.1 Matrix intuition

Each of the $\mathbf{M}$ matrices are large matrices but a useful property of them is that they are sparse matrices which means that is not necessary to store the whole matrix. We can take only the part required for our computations. In the following example, we show that $\mathbf{Q}_{u^c}$ and $\mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c}$ can be computed in a way that does not involve manipulation of these large matrices Danaei et al. [2011], Finucane et al. [2011].

Let us assume that we have studies in 5 years for one country as depicted in Table 4.1, Danaei et al. [2011], Finucane et al. [2011]. We compute $\mathbf{Q}_{u_c}$ separately for DBP, SBP and INT. First we set the 3-vector $(\mathbf{Z}\phi + 1)^T = (v_1, v_2, v_3)$ and we can compute first component of $\mathbf{Q}_{u_c}$ as:

$$\mathbf{M}^c * (\mathbf{Z}\phi + 1))^T \boldsymbol{\Sigma}^{-1} (\mathbf{M}^c * (\mathbf{Z}\phi + 1)) = \begin{pmatrix} v_1\Sigma_{11}^{-1} + v_2\Sigma_{22}^{-1} & & & & \\ & & & & \\ & & & & \\ & & & & \\ & & & & v_3\Sigma_{33}^{-1} \end{pmatrix}$$

The above result shows that all the information we need appears in the diagonal and only to the years with data. Following, the above can be easily generalised using DBP, SBP and INT variables since all these three measurements are referring to the same study, year and country. Therefore, for a specific year a specific blood pressure measurement will give us all these three values. Hence, we could sum the information that the variables gives us for each year for a specific country which is the information we elicit through the data (likelihood).

Finally, taking into account all the three variables DBP, SBP and INT together in a matrix form we can define the matrix multiplication of $\mathbf{M}^c * (\mathbf{Z}\phi + 1))^T \mathbf{\Sigma}^{-1} (\mathbf{M}^c * (\mathbf{Z}\phi + 1))$ as expressed below:

$$= \begin{pmatrix} \upsilon_1^2 \Sigma_{1D}^{-1} + \upsilon_2^2 \Sigma_{2D}^{-1} + \upsilon_1^2 \Sigma_{1S}^{-1} + \upsilon_2^2 \Sigma_{2S}^{-1} + \upsilon_1^2 \Sigma_{1I}^{-1} + \upsilon_2^2 \Sigma_{2I}^{-1} & & \\ & \dots & \\ & & \upsilon_3^2 \Sigma_{3D}^{-1} + \upsilon_3^2 \Sigma_{3S}^{-1} + \upsilon_3^2 \Sigma_{3I}^{-1} \end{pmatrix}$$

In the posterior distribution we want to estimate the information for 25 possible combinations of years. However, the maximum number of year combinations with information per country will be the square root of 25, 5. The number is referring to the pair of years that are exactly the same, the year that is repeated twice. By having the prior defined in Section 2.6.5, as a precision matrix we will include all the possible combination of years each time in the diagonal. The possible combinations of years are presented below: $\text{diag}(01.01, 01.02, 01.03, \dots, 02.02, \dots, 03.05, \dots, 05.05)$. Hence, if we say that country 1 has information in the first and the second year it means that in the diagonal the first and the seventh element will have values. In summary, although the precision matrix, $\mathbf{Q}_{u_c}$ has values for the years $\text{diag}([1,1],[1,2],[1,3],[1,4],[1,5],[2,1],[2,2],\dots)$ only the matching years will have information from the data (likelihood) which are $[1,1]$ and $[2,2]$ in our example. Five values out of 25 possible combinations of years in the diagonal of the precision matrix $\mathbf{Q}_{u_c}$ are non-zero and a prior with a matrix of 25×25 giving information to those combinations of $years \times country$ which do not have adequate information. The 25 possible combinations of years are going to be used in the posterior distribution but for the predictive probability density we will solely need these five values when both DBP and SBP have the same year. For the non-linear model we need to find the posterior distribution separately for the country, region, super-region and globe.

4.3.2.2 Defining the constraints

For the two-dimensional non-linear sub-model we will use a two-dimensional second order IGMRF as a prior, Yue and Speckman [2010]. Strictly speaking the precision matrices that Rue & Held (Rue and Held [2005]) and Yue (Yue and Speckman [2010]) define are of order one since the rank of the precision matrices is one whereas in Terzopoulos Terzopoulos [1988] the precision matrix is of order 3. However locally, they can retain their desirable properties.
As we have seen in Section 4.3.1.1 we need again to impose the identifiability constraints. Those constraints express the need for the summation of the intercept and the linear trend of the non-linear sub-model to be equal to zero, the so-called identifiability constraints. Also, restrictions exist because of the use of the two-dimensional IGMRF. The two-dimensional IGMRF imposes the restriction of three terms being equal to zero. These are referring to the intercept, the linear trend of the first variable and the linear trend of the second variable and are expressed as $\sum_i \sum_j u_{[i,j]} = \sum_i iu_{[i,j]} = \sum_j ju_{[i,j]} = 0$. Since, the constraints are referring to three variables we have the $T^2 - 3$ term in the distribution. In our analysis it means that we need one more year of data than the one-dimensional model in Section 3.1 in order to define all the parameters. Hence, the possible scenarios for the two dimensional non-linear sub-model will be four whereas in the one-dimensional equivalent we had three (see Section 4.3.1.1). The four possible scenarios are: 1) no data were observed 2) data of one year observed 3) data of two years observed and 4) data of three or more years observed in a specific country. For each of these scenarios the proposal distribution changes. Firstly, we start with the λ's proposal which is a symmetric one and therefore it cancels out in the

r ratio. We propose λ_c from:

$$\lambda_c^* \sim \mathcal{N}(a_5, b_5)$$

Then, we need to explore the analysis that each scenario needs:

1. **No data were observed in country j**

   There are no data so the likelihood will not contribute and $\mathbf{Q}_{u^c}$ depends solely on the prior. We will take into account only the second part of the precision, equation (4.24), $\mathbf{I}j \otimes \lambda\mathbf{P}$ which means dependence through the $\mathbf{P}$ matrix. Hence, the rank of $\mathbf{P}$ stays as it is, $T^2 - 3$ corresponding to infinite prior variance on the mean (intercept) and the two linear trends of $\mathbf{u}_j^c$. We constrain these three linear combinations of $\mathbf{u}_j^c$ to zero by taking the generalized inverse of $\mathbf{P}$, settings $\mathbf{P}$'s last three eigenvalues to ∞. We will apply the restriction that the two-dimensional second order IGMRF (RW2D) imposes, (see Section 2.6.5). In this scenario only one type of the two constraints is used; the identifiability constraint does not apply as we do not have defined mean and linear trends. The steps we follow are:

   (a) $\mathbf{u}_j^{c*} \sim \mathcal{N}(\mathbf{0}, (\lambda_c\mathbf{P})^{-1})$

   (b) Restrictions of using the RW2D by eigendecomposing the $\mathbf{P}$ to eigenvalues and eigenvectors. So, $\mathbf{P} = \mathbf{\Gamma\Lambda\Gamma}^T$ or $\mathbf{P}^{-1} = \mathbf{\Gamma\Lambda}^{-1}\mathbf{\Gamma}^T$ such that $\Lambda_{ii}^{-1} = \frac{1}{\Lambda_{ii}}$ for $i = 1, \ldots, T^2 - 3$ and $\Lambda_{ii}^{-1} = 0$ for $i = T^2 - 2, T^2 - 1, T^2$.

   (c) Generate $\mathbf{u}_j^c$.

   $\mathbf{u}_j^{c*} \sim \mathcal{N}(\mathbf{0}, (\lambda_c\mathbf{P})^{-1})$, where $\lambda_c\mathbf{P}$ is the precision matrix. Therefore, instead of dividing elements we multiply them.

   $\mathbf{u}_j^{c*} \sim \frac{1}{\lambda_c^{1/2}}\mathcal{N}(\mathbf{0}, \mathbf{\Gamma\Lambda}^{-1}\mathbf{\Gamma}^T)$ $\quad$ $\mathbf{u}_j^{c*} \sim \frac{1}{\lambda_c^{1/2}}\mathcal{N}(\mathbf{0}, \mathbf{\Gamma\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2}\mathbf{\Gamma}^T)$

   $\mathbf{u}_j^{c*} \sim \frac{\mathbf{\Gamma\Lambda}^{-1/2}}{\boldsymbol{\lambda}_c^{1/2}}\mathcal{N}(0, 1)$

   So, we generate a $\mathbf{u}_j^c$ value from: $\frac{\mathbf{\Gamma\Lambda}^{-1/2}}{\boldsymbol{\lambda}_c^{1/2}} z$ where $z \sim \mathcal{N}(0, 1)$

(d) Calculate:

$$\begin{aligned}\log P(\mathbf{u}_j^{c*}|\lambda_c^*, \mathbf{y}, \mathbf{A}\mathbf{u}_j^{c*} = 0, \forall j.) &= -\frac{T^2-3}{2}\log 2\pi + \frac{T^2-3}{2} log\lambda_c^* \\ &- \frac{\lambda_c^*}{2}\mathbf{u}_j^{c*T}\mathbf{P}\mathbf{u}_j^{c*} \\ &\propto \frac{T^2-3}{2}\log \lambda_c^* - \frac{1}{2}\mathbf{u}_j^{c*T}\lambda_c^*\mathbf{P}\mathbf{u}_j^{c*}\end{aligned}$$

2. **Data for one year were observed in country j**

$\mathbf{Q}_{u^c}$ has rank $T^2 - 2$ because the mean (intercept) but neither of the linear trends of $\mathbf{u}_j$ are defined by the data. Hence, we constrain the linear trends $\mathbf{u}_j$ to zero by taking the generalized inverse of $\mathbf{Q}_{u^c}$, setting $\mathbf{Q}_{u^c}$'s last two eigenvalues to ∞. Thus, we apply the restriction of the infinite linear trends that the two-dimensional second order IGMRFs imposes. In addition, we need to restrict the summation of the intercept in order to have identifiability with the linear sub-model. To do that, we use the "Rue corrections" to constrain the mean of $\mathbf{u}_j^c$'s intercept to zero. Two different types of constraints are applied. Steps:

(a) Eigendecompose $\mathbf{Q}_{u_j^c} = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$

(b) Calculate the generalised inverse of $\mathbf{Q}_{u_j^c}$: $\mathbf{Q}_{u_j^c}^{-1} = \mathbf{\Gamma}\mathbf{\Lambda}^{-1}\mathbf{\Gamma}^T$ such that $\Lambda_{ii}^{-1} = \frac{1}{\Lambda_{ii}}$ for $i = 1, \ldots, T^2 - 2$ and $\Lambda_{ii}^{-1} = 0$ for $i = T^2 - 1, T^2$.
The restriction is referring to the two undefined linear trends.

(c) Calculate:
$\mathbf{V}_{u_j^c}^* = \mathbf{Q}_{u_j^c}^{-1} - \mathbf{Q}_{u_j^c}^{-1}1^T(1\mathbf{Q}_{u_j^c}^{-1}1^T)^{-1}1\mathbf{Q}_{u_j^c}^{-1}$
$\boldsymbol{\mu}_{u_j^c}^* = \mathbf{Q}_{u_j^c}^{-1}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c} - \mathbf{Q}_{u_j^c}^{-1}1^T(1\mathbf{Q}_{u_j^c}^{-1}1^T)^{-1}1\mathbf{Q}_{u_j^c}^{-1}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c}$
We calculate a new variance and mean. Since only the intercept can be identifiable from the data there is one identifiability restriction for it and it can be adjusted as: $A = 1^T$ and $e = 0$. The restriction is defined as: $1^T\mathbf{u}_j^c = 0 \Rightarrow u_{j[1]}^c + \cdots + u_{j[T^2]}^c = 0$

(d) Eigendecompose $\mathbf{V}_{u_j^c}^* = \mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T$ such that $\Lambda_{ii}^{-1} = 1/\Lambda_{ii}$ for $i = 1, \ldots,$

T^2-3 and $\Lambda_{ii}^{-1}=0$ for $i=T^2-2, T^2-1, T^2$.

In the new variance we apply the three restrictions of the RW2D.

(e) Generate $\mathbf{u}_j^{c*}=\boldsymbol{\mu}_{u_j^c}^{*}+\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1/2}z$, with $z\sim\mathcal{N}(0,1)$

(f) Calculate:

$$\begin{aligned}\log P(\mathbf{u}_j^{c*}|\lambda_c^*, \mathbf{A}\mathbf{u}_j^{c*}=0,\mathbf{y}) &= -\frac{T^2-3}{2}\log 2\pi-\frac{1}{2}\sum_{i=1}^{T^2-3}\log\Lambda_{ii}^{-1}\\ &-\frac{1}{2}(\mathbf{u}_j^{c*}-\boldsymbol{\mu}_{u_j^{c*}})^T\mathbf{V}_{u_j^c}^{*-1}(\mathbf{u}_j^{c*}-\boldsymbol{\mu}_{u_j^{c*}})\\ &\propto -\frac{1}{2}\sum_{i=1}^{T^2-3}\log\Lambda_{ii}^{-1}-\frac{1}{2}(\mathbf{u}_j^{c*}-\boldsymbol{\mu}_{u_j^{c*}})^T\mathbf{V}_{u_j^c}^{*-1}\\ &(\mathbf{u}_j^{c*}-\boldsymbol{\mu}_{u_j^{c*}})\end{aligned}$$

3. **Data for two years were observed in country j**

$\mathbf{Q}_{u^c}$ has rank T^2-1 because the mean and one of the linear trends of $\mathbf{u}_j$ are defined by the data. Hence, we constrain the second linear trend $\mathbf{u}_j$ to zero by taking the generalized inverse of $\mathbf{Q}_{u^c}$, setting $\mathbf{Q}_{u^c}$'s last eigenvalue to ∞ which means that we apply the restriction for the infinite linear trend that the two-dimensional second order IGMRF imposes. In addition, we need to restrict the summation of the intercept and the first variable's slope to zero in order to have identifiability with the linear sub-model. To do that we use the "Rue corrections" to constrain the mean of $\mathbf{u}_j^c$'s intercept and the mean of one of the $\mathbf{u}_j^c$'s linear trends to zero. In summary we apply two different types of constraints. The steps we follow are:

(a) Eigendecompose $\mathbf{Q}_{u_j^c}=\boldsymbol{\Gamma}\boldsymbol{\Lambda}\boldsymbol{\Gamma}^T$

(b) Calculate the generalised inverse of $\mathbf{Q}_{u_j^c}$: $\mathbf{Q}_{u_j^c}^{-1}=\boldsymbol{\Gamma}\boldsymbol{\Lambda}^{-1}\boldsymbol{\Gamma}^T$ such that$\Lambda_{ii}^{-1}=\frac{1}{\Lambda_{ii}}$ for $i=1,\ldots,T^2-1$ and $\Lambda_{ii}^{-1}=0$ for $i=T^2$.

The constraint is referring to the one undefined linear trend.

(c) Calculate:

$\mathbf{V}_{u_j^c}^{*}=\mathbf{Q}_{u_j^c}^{-1}-\mathbf{Q}_{u_j^c}^{-1}\mathbf{A}[1:2,]^T(\mathbf{A}[1:2,]\mathbf{Q}_{u_j^c}^{-1}\mathbf{A}[1:2,]^T)^{-1}\mathbf{A}[1:2,]\mathbf{Q}_{u_j^c}^{-1}$

$\boldsymbol{\mu}_{u_j^c}^{*}=\mathbf{Q}_{u_j^c}^{-1}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c}-\mathbf{Q}_{u_j^c}^{-1}\mathbf{A}[1:2,]^T(\mathbf{A}[1:2,]\mathbf{Q}_{u_j^c}^{-1}\mathbf{A}[1:2,]^T)^{-1}\mathbf{A}[1:2,]\mathbf{Q}_{u_j^c}^{-1}\mathbf{Q}_{u_j^c}\boldsymbol{\mu}_{u_j^c}$.

We calculate a new variance and mean. Since the intercept and the first variable's slope are identifiable, $\mathbf{A}$ will be a $2 \times T^2$ matrix which will restrict the summation of intercepts and the summation of slopes to be equal to zero, the identifiability constraint.

(d) Eigendecompose $\mathbf{V}^*_{u^c_j} = \mathbf{\Gamma\Lambda\Gamma}^T$ such that $\Lambda^-_{ii} = 1/\Lambda_{ii}$ for $i = 1, \ldots, T^2 - 3$ and $\Lambda^{-1}_{ii} = 0$ for $i = T^2-2, T^2-1, T^2$.

In the new variance we apply the three restrictions of the RW2D.

(e) Generate $\mathbf{u}^{c*}_j = \boldsymbol{\mu}^*_{u^c_j} + \mathbf{\Gamma\Lambda}^{-1/2} z$ where $z \sim N(0,1)$

(f) Calculate:

$$\log P(\mathbf{u}^{c*}_j|\lambda^*_c, \mathbf{A}\mathbf{u}^{c*}_j = 0, \mathbf{y}) = -\frac{T^2-3}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{T^2-3}\log\Lambda^{-1}_{ii}$$
$$- \frac{1}{2}(u^{c*}_j - \boldsymbol{\mu}^*_{u^c_j})'\mathbf{V}^*_{\mathbf{u}^c_j}(\mathbf{u}^{c*}_j - \boldsymbol{\mu}^*_{\mathbf{u}^c_j})$$
$$\propto -\frac{1}{2}\sum_{i=1}^{T^2-3}\log\Lambda^{-1}_{ii} - \frac{1}{2}(\mathbf{u}^{c*}_j - \boldsymbol{\mu}_{u^{c*}_j})^T\mathbf{V}^{*-1}_{u^c_j}$$
$$(\mathbf{u}^{c*}_j - \boldsymbol{\mu}_{u^{c*}_j})$$

4. **Data for multiple years were observed in country j**

$\mathbf{Q}_{u^c}$ has rank T^2 because the mean and both of the linear trends of $\mathbf{u}^c_j$ are identified by the data. Since all the three parameters are identified by the data, we use the "Rue corrections" to constrain the mean of $\mathbf{u}^c_j$'s intercept and $\mathbf{u}^c_j$'s linear trends to zero (identifiability constraints). Afterwards, to the modified precision matrix we apply the restrictions of the two-dimensional second order IGMRF. Steps:

(a) Calculate $\mathbf{V}_{u^c_j} = \mathbf{Q}^{-1}_{u^c_j}$

(b) Calculate:

$\boldsymbol{\mu}^*_{u^c_j} = \mathbf{V}_{u^c_j}\mathbf{Q}_{u^c_j}\boldsymbol{\mu}_{u^c_j} - \mathbf{V}_{u^c_j}\mathbf{A}^T(\mathbf{A}\mathbf{V}_{u^c_j}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{V}_{u^c_j}\mathbf{Q}_{u^c_j}\boldsymbol{\mu}_{u^c_j}$

$\mathbf{V}^*_{u^c_j} = \mathbf{V}_{u^c_j} - \mathbf{V}_{u^c_j}\mathbf{A}^T(\mathbf{A}\mathbf{V}_{u^c_j}\mathbf{A}^T)^{-1}\mathbf{A}\mathbf{V}_{u^c_j}$

We calculate a new variance and mean. Since the intercept and both variables' linear trends are identifiable, $\mathbf{A}$ will be a $3 \times T^2$ matrix which will restrict the summation of the intercept and of the linear trends to be equal to zero.

(c) Eigendecompose $\mathbf{V}^*_{u^c_j} = \mathbf{\Gamma}\mathbf{\Lambda}^{-}\mathbf{\Gamma}^T$ such that $\Lambda^{-1}_{ii} = 1/\Lambda_{ii}$ for $i = 1, ..., T^2 - 3$ and $\Lambda^{-1}_{ii} = 0$ for $i = T^2-2, T^2-1, T^2$.

In the new variance we apply the three restrictions of the RW2D.

(d) Generate $\mathbf{u}^c_j = \boldsymbol{\mu}_{u^c_j} + \mathbf{\Gamma}\mathbf{\Lambda}^{-1/2} z$ with $z \sim \mathcal{N}(0, 1)$

(e) Defining $\mathbf{V}^*_{u^c_j}$ as above, calculate:

$$\begin{aligned}\log P(\mathbf{u}^{c*}_j|\lambda^*_c, \mathbf{y}, \mathbf{A}\mathbf{u}^{c*}_j = 0, .) &= -\frac{T^2-3}{2}\log 2\pi - \frac{1}{2}\sum_{i=1}^{T^2-3}\log \Lambda^{-}_{ii} \\ &- \frac{1}{2}(\mathbf{u}^{c*}_j - \boldsymbol{\mu}^*_{u^c_j})^T \mathbf{V}^{*-}_{u^c_j}(\mathbf{u}^{c*}_j - \boldsymbol{\mu}^*_{u^c_j}) \\ &\propto -\frac{1}{2}\sum_{i=1}^{T^2-3}\log \Lambda^{-}_{ii} - \frac{1}{2}(\mathbf{u}^{c*}_j - \boldsymbol{\mu}^*_{u^c_j})^T \mathbf{V}^{*-}_{u^c_j} \\ &(\mathbf{u}^{c*}_j - \boldsymbol{\mu}^*_{u^c_j})\end{aligned}$$

A is a $3 \times T^2$ constraint matrix whose first row is a vector of 1's, the second row is a vector of centred time values for the DBP variable and the third row is a vector of centred time values for the SBP variable. Since, we have information for the intercept and the linear trend for both variables, we take the summation of each of the three rows to be restricted to zero.

In summary, we need to stress that by using any type of IGMRF there are restrictions to be considered. In the case of the two-dimensional second order random walk there are three restrictions to apply, $\sum_i \sum_j u_{[ij]} = 0$, $\sum_i i u_{[ij]} = 0$ and $\sum_j j u_{[ij]} = 0$.

Finally, after accounting for the aforementioned restrictions we need to impose for the last time the three restrictions of the two-dimensional second order IGMRFs to the new (modified) final precision matrix.

4.3.3 Age sub-model: posterior distribution

Having found the distributions of the parameters $\boldsymbol{\theta}$ and $\mathbf{u}$ of the linear and the non-linear sub-model respectively we can write the mean of the likelihood in a simplified way, Danaei et al. [2011], Finucane et al. [2011]:

$$E[\mathbf{y}|\boldsymbol{\theta}, \mathbf{u}, \boldsymbol{\gamma}] = \mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u} + \mathbf{R}\gamma \tag{4.25}$$

$\mathbf{R}$ is a mapping matrix with I rows which contain the age matrix using the B-splines $\mathbf{Z} = (N_{0,3}(t), N_{1,3}(t), N_{2,3}(t), N_{3,3}(t), N_{4,3}(t))$ and the $\mu = \mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u}$ term and assigns each c to the appropriate studies. The columns of the R matrix are defined below

$$\begin{aligned}\mathbf{R}_{h,i} = [&N_{0,3}(t) \quad N_{1,3}(t) \quad N_{2,3}(t) \quad N_{3,3}(t) \quad N_{4,3}(t)\\ &\mu_i N_{0,3}(t) \quad \mu_i N_{1,3}(t) \quad \mu_i N_{2,3}(t) \quad \mu_i N_{3,3}(t) \quad \mu_i N_{4,3}(t)\\ &C_i * N_{0,3}(t) \quad C_i * N_{1,3}(t) \quad C_i * N_{2,3}(t) \quad C_i * N_{3,3}(t) \quad C_i * N_{4,3}(t)]\end{aligned} \tag{4.26}$$

A more compressed way of writing the $\mathbf{R}$ matrix is:

$$R = [\mathbf{Z} \quad \mathrm{diag}(\mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u})\mathbf{Z} \quad \mathrm{diag}(N_{0,3}(t))\mathbf{C} \quad \mathrm{diag}(N_{1,3}(t))\mathbf{C} \ldots \mathrm{diag}(N_{4,3}(t))\mathbf{C}] \tag{4.27}$$

Using the simplified version of the likelihood of (4.25) we present in (4.28) the posterior distribution of the age sub-model

$$\begin{aligned}
&\log P(\boldsymbol{\gamma}|\mathbf{y}) \propto \log P(\mathbf{y}|\gamma) + \log P(\boldsymbol{\gamma})\\
&\log P(\boldsymbol{\gamma}|\mathbf{y}) \propto (\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{M}\mathbf{u} - \mathbf{R}\boldsymbol{\gamma})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{M}\mathbf{u} - \mathbf{R}\boldsymbol{\gamma}) + \boldsymbol{\gamma}^T V_\gamma^{-1} \boldsymbol{\gamma}\\
&\log(P(\boldsymbol{\gamma}|\mathbf{y}) \propto -\mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma} + (\mathbf{F}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma} + (\mathbf{M}\mathbf{u})^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma}\\
&\qquad - (\mathbf{R}\boldsymbol{\gamma})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{M}\mathbf{u}) + (\mathbf{R}\boldsymbol{\gamma})^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma} + \boldsymbol{\gamma}^T V_\gamma^{-1} \boldsymbol{\gamma}\\
&\log(P(\boldsymbol{\gamma}|\mathbf{y}) \propto \boldsymbol{\gamma}^T \underline{\mathbf{R}^T \boldsymbol{\Sigma}^{-1} \mathbf{R}} \boldsymbol{\gamma} - \boldsymbol{\gamma}^T \underline{\mathbf{R}^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{M}\mathbf{u})} + \boldsymbol{\gamma}^T \underline{V_\gamma^{-1}} \boldsymbol{\gamma}\\
&\qquad - \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma} + (\mathbf{F}\boldsymbol{\theta})^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma} + (\mathbf{M}\mathbf{u})^T \boldsymbol{\Sigma}^{-1} \mathbf{R}\boldsymbol{\gamma}
\end{aligned} \tag{4.28}$$

Again, bearing in mind the equations in (4.8), and by using the underlined terms in (4.28) the $\boldsymbol{\gamma}$ parameter can have a Normal distribution shown in (4.29):

$$\begin{aligned} \boldsymbol{\gamma}|\mathbf{y} &\sim N(\mathbf{Q}\boldsymbol{\mu}_{\gamma}, \mathbf{Q}_{\gamma}) \\ \mathbf{Q}\boldsymbol{\mu}_{\gamma} &= \mathbf{R}^{T}\boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{F}\boldsymbol{\theta} - \mathbf{Mu}) \\ \mathbf{Q}_{\gamma} &= \mathbf{R}^{T}\boldsymbol{\Sigma}^{-1}\mathbf{R} + \mathbf{V}_{\gamma}^{-1} \end{aligned} \tag{4.29}$$

In equation (4.29) we can see the parameters of the conditional distribution for each Gibbs sampler's step. Gibbs sampler algorithm is going to be implemented instead of M-H used in Sections 4.2 and 4.3. The reason that we used M-H in the two previous sub-models is the high dependence between the parameters and hyperparameters in each model and therefore, we updated parameters and hyperparameters in a joint distribution. In the age sub-model we do not observe any strong dependence therefore, we can use the Gibbs sampler.

For the two-dimensional equivalent we will have the same terms but with different dimensionality for each one of them. Considering that we have the two-dimensional linear and non-linear sub-model, the $\mathbf{R}$ matrix will be expressed as follows:

$$\begin{aligned} R = [&\mathbf{Z} \quad \mathbf{Z} \quad \mathbf{Z} \quad \mathrm{diag}(\mathbf{F}\boldsymbol{\theta} + \mathbf{Mu})\mathbf{Z} \quad \mathrm{diag}(\mathbf{F}\boldsymbol{\theta} + \mathbf{Mu})\mathbf{Z} \quad \mathrm{diag}(\mathbf{F}\boldsymbol{\theta} + \mathbf{Mu})\mathbf{Z} \\ &\mathrm{diag}(N_{0,3}(t))\mathbf{C} \quad \mathrm{diag}(N_{1,3}(t))\mathbf{C} \ldots \quad \mathrm{diag}(N_{4,3}(t))\mathbf{C} \\ &\mathrm{diag}(N_{0,3}(t))\mathbf{C} \quad \mathrm{diag}(N_{1,3}(t))\mathbf{C} \ldots \quad \mathrm{diag}(N_{4,3}(t))\mathbf{C} \\ &\mathrm{diag}(N_{0,3}(t))\mathbf{C} \quad \mathrm{diag}(N_{1,3}(t))\mathbf{C} \ldots \quad \mathrm{diag}(N_{4,3}(t))\mathbf{C}] \end{aligned} \tag{4.30}$$

In (4.30) we can see that in the $\mathbf{R}$ matrix the $\mathbf{Z}$ matrix of the B-splines is repeated three times as well as $\mu\mathbf{Z}$ and the random effects, $\mathbf{C}$ since we have three variables under consideration, DBP, SBP, INT. Therefore, the γ parameter will have information for three variables and its dimensionality will coincide with the columns of the $\mathbf{R}$ matrix. It needs to be stressed that the C matrix is referring to all the J countries multiplied each time with one of the terms from the B-splines. Therefore, due to the random effects we generate $5 \times J$ columns for each variable.

4.3.4 Age sub-model's hyperparameter: posterior distribution

We introduce the hyperprior for the age sub-model. The σ^2 parameter is the variance of the random effect, c_i. In equation (4.32) we define the posterior distribution of the $\boldsymbol{\gamma}$ parameter from which we extract information for σ^2.

$$P(\sigma_1^2|\mathbf{c}_1) \propto P(\mathbf{c}_1|\sigma_1^2)P(\sigma_1^2) \tag{4.31}$$

$$\begin{aligned} P(\boldsymbol{\gamma}|\mathbf{y}) \propto P(\mathbf{y}|\boldsymbol{\gamma},\boldsymbol{\sigma}^2)P(\boldsymbol{\gamma}|\boldsymbol{\sigma}^2) \\ P(\boldsymbol{\sigma}^2|\mathbf{y}) \propto P(\mathbf{y}|c,\boldsymbol{\sigma}^2)P(c|\boldsymbol{\sigma}^2) \end{aligned} \tag{4.32}$$

We need to highlight that what we want to calculate is not the distribution of the random effects, $c_{1_j}, c_{2_j}, c_{3_j}, c_{4_j}, c_{5_j}$ but the distribution of their variances, σ_1^2, σ_2^2, σ_3^2, σ_4^2, σ_5^2. We know that the random effects, c_i follow a Normal distribution. We give an example for c_{1j} in (4.33) but the remaining terms, $c_{2_j}, \ldots, c_{5_j}$ are defined analogously.

$$c_{1j} \sim \mathcal{N}(0, \sigma_1^2) \tag{4.33}$$

Since we use c_{1j} $(c_{2_j}, \ldots, c_{5_j})$ for all the J countries we can write it in a matrix form.

$$\mathbf{c}_1 \sim \mathcal{N}(\mathbf{0}, \sigma_1^2\mathbf{I}_J) \tag{4.34}$$

In (4.34) $\mathbf{c}_1$ is a vector of J values with a diagonal covariance matrix with the σ_1^2 value in the diagonal. We will make use of $\log \sigma_1^2$ for computational convenience.

$$\begin{aligned}
\log P(\sigma_1^2|\mathbf{c}_1,\gamma,\mathbf{y}.) &\propto \log P(\mathbf{c}_1|\sigma_1^2) + \log P(\sigma_1^2) \\
\log P(\log\sigma_1^2|\mathbf{c}_1,\gamma,\mathbf{y}.) &\propto \log(2\pi\sigma_1^2)^{-J/2} - \frac{1}{2}\mathbf{c}_1^T(\sigma_1^2\mathbf{I}_J)^{-1}\mathbf{c}_1 + \log(\sigma_1^2)/2 \\
\log P(\log\sigma_1^2|\mathbf{c}_1,\gamma,\mathbf{y}.) &\propto -\frac{1}{2}J\log(\sigma_1^2) - \frac{1}{2}\mathbf{c}_1^T(\sigma_1^2\mathbf{I}_J)^{-1}\mathbf{c}_1 + \log(\sigma_1^2)/2 \\
\log P(\log\sigma_1^2|\mathbf{c}_1,\gamma,\mathbf{y}.) &\propto -\frac{1}{2}(J\log(\sigma_1^2) + \mathbf{c}_1^T(\sigma_1^2\mathbf{I}_J)^{-1}\mathbf{c}_1) + \log(\sigma_1^2)/2 \\
\log P(\log\sigma_1^2|\mathbf{c}_1,\gamma,\mathbf{y}.) &\propto -\frac{1}{2}(J\log(\sigma_1^2) + \frac{1}{\sigma_1^2}\mathbf{c}_1^T\mathbf{c}_1) + \underline{\log(\sigma_1^2)/2}
\end{aligned} \tag{4.35}$$

Each of the σ_x^2 values are random effects in each of the countries which means we will use each one of the σ^2 terms J times. In (4.35) the first term is coming from the likelihood and the underlined one expresses the prior of σ_1^2.
We use the M-H algorithm for the update of σ_i^2. The difference from the updating of the previous sub-models is that each σ_i^2 will be updated separately. However, the posterior distribution for each of the σ_i^2 terms is a joint distribution since it includes J values from J countries.
We will update the $\log\sigma_1^2$ instead of σ_1^2 and hence accept $\log\sigma_1^2$ with probability r:

$$\begin{aligned}
r &= \frac{Posterior^* \times Proposal}{Posterior \times Proposal^*} \\
r &= \frac{P(\sigma_1^{2*}|\mathbf{c}_1,\gamma,\mathbf{y}.) \times q(\sigma_1^2|\sigma_1^{2*})}{P(\sigma_1^2|\mathbf{c}_1,\gamma,\mathbf{y}.) \times q(\sigma_1^{2*}|\sigma_1^2)} \\
r &= \frac{P(\log\sigma_1^{2*}|\mathbf{c}_1,\gamma,.)}{P(\log\sigma_1^{2(i-1)}|\mathbf{c}_1,\gamma,.)} \\
r &= \exp\left(\log\left(\frac{P(\log\sigma_1^{2*}|\mathbf{c}_1,\gamma,.)}{P(\log\sigma_1^{2(i-1)}|\mathbf{c}_1,\gamma,.)}\right)\right) \\
r &= \exp\left(\log P(\log\sigma_1^{2*}|\mathbf{c}_1,\gamma,.) - \log P(\log\sigma_1^{2(i-1)}|\mathbf{c}_1,\gamma,.)\right)
\end{aligned} \tag{4.36}$$

As we have seen in the linear and the non-linear sub-models in Sections 4.2 and 4.3, we use M-H and not the Gibbs Sampler to tackle the dependence that exists between parameters and hyper-parameters. Here, the dependence exists between the random effects of the J countries using the σ_x^2 parameter. Therefore, we want to include joint distributions for each of the σ_x^2 parameters for all the J countries where $x \in [1, 5]$. Equation (4.36) is used in the same way for the rest of the σ_x^2 terms.

For the two-dimensional sub-model the whole analysis is exactly the same, except for the change in dimensionality. The $\mathbf{c}$ parameter will no longer be a vector of length 5 but of 15 since we will have 5 random effects for each of the variables, DBP, SBP, INT. In other words, we will have 15 parameters of σ_x^2 that is used for each of the J countries and again there are 5 for each of the DBP, SBP and INT variables.

4.3.5 ν, e_i sub-model's hyperparameter: posterior distribution

Since we have a hierarchical model, we introduce priors for the parameters of the sub-models and hyperpriors for the priors' parameters. The ν parameter is the variance of the random effects, e_i. In the previous subsections we compute the posterior distribution of $\boldsymbol{\theta}$ which as we indicate in Section 3.2.1, equation (4.1) includes also the e parameter:

$$P(\boldsymbol{\theta}|\mathbf{y}) \propto P(\mathbf{y}|\boldsymbol{\theta}, e, \nu)P(\boldsymbol{\theta}|e, \nu)P(e|\nu) \tag{4.37}$$

The variance of e, ν depends on the coverage so we need each time to update it as well:

$$P(\nu|e) \propto P(e|\nu)P(\nu) \tag{4.38}$$

We already have defined a Normal distribution for $e|\nu$ and we define a Beta distribution for the prior of ν.

$$\begin{aligned} e|\nu &\sim \mathcal{N}(0,\nu) \\ \log \nu &\sim \mathcal{B}eta(3/2,1) \end{aligned} \tag{4.39}$$

It is necessary to highlight that we do not want to compute the distribution of the random effect, e_i but the distribution of the variance of the random effect, ν_{nat}, ν_{subn}, ν_{com}, the so-called hyperprior of the random effect e_i. The distribution of $e_{nat}, e_{subn}, e_{com}$ is inside the $\boldsymbol{\theta}$ variable and we know that the random variable follows a Gaussian distribution for each of the national, subnational and community coverage. Taking as an example the national coverage, but with the same application to subnational and community coverage, we have the following:

$$\mathbf{e}_{nat}|\nu_{nat} \sim \mathcal{N}(\mathbf{0}, \nu_{nat}\mathbf{I}_{N_{nat}}) \tag{4.40}$$

where $\mathbf{I}_{N_{nat}}$ is a diagonal matrix of $N_{nat} \times N_{nat}$ dimensions and N_{nat} is the number of national studies. Lastly, instead of calculating the distribution of ν_{nat} we will compute the distribution of $\log(\nu_{nat})$, Danaei et al. [2011], Finucane et al. [2011].

$$\begin{aligned} \log P(\nu_{nat}|\mathbf{y}) &\propto \log P(\mathbf{e}_{nat}|\nu_{nat}) + \log P(\nu_{nat}) \\ \log P(\nu_{nat}|\mathbf{y}) &\propto \log(2\pi\nu_{nat})^{-N_{nat}/2} - \frac{1}{2}\mathbf{e}_{nat}^T\nu_{nat}^{-1}\mathbf{e}_{nat} + \log(\nu_{nat}^{1/2}) \\ \log P(\nu_{nat}|\mathbf{y}) &\propto -\frac{N_{nat}}{2}\log(\nu_{nat}) - \frac{1}{2}\mathbf{e}_{nat}^T\nu_{nat}^{-1}\mathbf{e}_{nat} + \log(\nu_{nat})/2 \\ \log P(\log\nu_{nat}|\mathbf{y}) &\propto -\frac{1}{2}(N_{nat}(\log(\nu_{nat}) + \mathbf{e}_{nat}^T\nu_{nat}^{-1}\mathbf{e}_{nat}) + \underline{\log(\nu_{nat})/2} \end{aligned} \tag{4.41}$$

In equation (4.41) the first term indicates the terms in the likelihood related to $\mathbf{e}_{nat}$ and ν_{nat} whereas the underlined one indicates the prior of ν_{nat}. In the distribution of $P(\mathbf{e}_{nat}|\nu_{nat})$ we have the term N_{nat} as we take from all the studies the national studies together in a joint Gaussian distribution with zero correlation between them.

Instead of using a Gibbs step to sample from the inverse Gamma full conditional distribution for each ν we will use a random walk M-H update for all the three types of coverage (ν_{nat}, ν_{subn}, ν_{com}) jointly to tackle issues of strong dependence, Danaei et al. [2011], Finucane et al. [2011]. We now propose a new configuration (ν_{nat}, ν_{subn}, ν_{com}) using one-block update in (4.42), Danaei et al. [2011], Finucane et al. [2011].

$$
\begin{aligned}
P(\nu_{nat}, \nu_{subn}, \nu_{com}|\mathbf{y}) &\propto P(\mathbf{e}_n|\nu_{nat})P(\nu_{nat})P(\mathbf{e}_n|\nu_{subn})P(\nu_{subn}) \\
&\quad P(\mathbf{e}_n|\nu_{com})P(\nu_{com}) \\
\log P(\nu_{nat}, \nu_{subn}, \nu_{com}|\mathbf{y}) &\propto -\frac{1}{2}(N_{nat}(log(\nu_{nat}) + \mathbf{e}_n^T \nu_{nat}^{-1}\mathbf{e}_n) + log(\nu_{nat})/2 \\
&\quad - \frac{1}{2}(N_{subn}(log(\nu_{subn}) + \mathbf{e}_n^T \nu_{subn}^{-1}\mathbf{e}_n) + log(\nu_{subn})/2 \\
&\quad - \frac{1}{2}(N_{com}(log(\nu_{com}) + \mathbf{e}_n^T \nu_{com}^{-1}\mathbf{e}_n) + log(\nu_{com})/2
\end{aligned}
\tag{4.42}
$$

Since we use a random walk M-H the proposal distribution is a symmetric distribution. The proposal distribution will refer to the marginal distributions of each ν_{nat}, ν_{subn} and ν_{com}. Therefore, in the fraction of the r acceptance ratio the proposals will be cancelled out.

$$
\begin{aligned}
r &= \frac{Posterior^* \times Proposal}{Posterior \times Proposal^*} \\
r &= \frac{P(\nu_{nat}^*, \nu_{subn}^*, \nu_{com}^*|\mathbf{y})q(\nu_{nat}|\nu_{nat}^*)q(\nu_{subn}|\nu_{subn}^*)q(\nu_{com}|\nu_{com}^*)}{P(\nu_{nat}, \nu_{subn}, \nu_{com}|\mathbf{y})q(\nu_{nat}^*|\nu_{nat})q(\nu_{subn}^*|\nu_{subn})q(\nu_{com}^*|\nu_{com})} \\
r &= \frac{P(\nu_{nat}^*, \nu_{subn}^*, \nu_{com}^*|\mathbf{y})}{P(\nu_{nat}, \nu_{subn}, \nu_{com}|\mathbf{y})} \\
&= \log P(\nu_{nat}^*, \nu_{subn}^*, \nu_{com}^*|\mathbf{y}) - \log P(\nu_{nat}, \nu_{subn}, \nu_{com}|\mathbf{y})
\end{aligned}
\tag{4.43}
$$

For the r acceptance rate we propose the three $\log \nu^*$'s from normal proposal distributions centred at the previous $\log \nu$ values, each with their own proposal variance. If $\log \nu_{nat}^* > \log \nu_{subn}^*$ or $\log \nu_{subn}^* > \log \nu_{com}^*$ we reject all three proposals. Otherwise, we accept all the three proposals. The posterior distribution that is used in (4.43) is the one presented in (4.42).

For the two-dimensional equivalent we need to repeat the procedure three times for the three variables, DBP, SBP and INT.

4.3.6 τ^2, likelihood's variance's hyperparameter: posterior distribution

What comes next is to find the posterior of τ^2 variable is inside the $\boldsymbol{\Sigma}$ parameter. Therefore, all the information that we want to extract from the likelihood is within the $\boldsymbol{\Sigma}$ term. We apply the appropriate methodology for τ^2_{nat} which is analogous for τ^2_{subn} and τ^2_{com}.

$$
\begin{aligned}
\log P(\tau^2_{nat}|\mathbf{y}) &\propto \log P(\mathbf{y}|\tau^2_{nat}) + \log P(\tau^2_{nat}) \\
\log P(\tau^2_{nat}|\mathbf{y}) &\propto -\frac{1}{2}(\log|\boldsymbol{\Sigma}| + (\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})^T\boldsymbol{\Sigma}^{-1}(\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})) \\
&\quad + \log(\tau^2_{nat})/2 \\
\log P(\tau^2_{nat}|\mathbf{y}) &\propto -\frac{1}{2}(\log|\boldsymbol{\Sigma}| + \mathrm{diag}(\boldsymbol{\Sigma}^{-1}) * (\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})^2) + \log(\tau^2_{nat})/2 \\
\log P(\log\tau^2_{nat}|\mathbf{y}) &\propto \frac{1}{2}1^T(\log(\mathrm{diag}(\boldsymbol{\Sigma}^{-1})) - diag(\boldsymbol{\Sigma}^{-1}) * (\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})^2) \\
&\quad + \underline{\log(\tau^2_{nat})/2}
\end{aligned}
\tag{4.44}
$$

In equation (4.44) is presented the posterior distribution of the τ^2 parameter solely for the national studies. In the following we are interested in presenting the posterior distribution for τ^2_{nat}, τ^2_{subn}, τ^2_{com} jointly. The terms are independent of each other.

$$
\begin{aligned}
P(\tau^2_{nat},\tau^2_{subn},\tau^2_{com}|\mathbf{y}) &\propto P(\boldsymbol{\Sigma}|\tau_{nat})P(\tau_{nat})P(\boldsymbol{\Sigma}|\tau_{subn})P(\tau_{subn}) \\
&\quad P(\boldsymbol{\Sigma}|\tau_{com})P(\tau_{com}) \\
\log P(\tau^2_{nat},\tau^2_{subn},\tau^2_{com}|\mathbf{y}) &\propto \frac{1}{2}1^T(\log(\mathrm{diag}(\boldsymbol{\Sigma}^{-1})) \\
&\quad - \mathrm{diag}(\boldsymbol{\Sigma}^{-1}) * (\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})^2) + \log(\tau^2_{nat})/2 \\
&\quad \frac{1}{2}1^T(\log(\mathrm{diag}(\boldsymbol{\Sigma}^{-1})) \\
&\quad - \mathrm{diag}(\boldsymbol{\Sigma}^{-1}) * (\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})^2) + \log(\tau^2_{subn})/2 \\
&\quad \frac{1}{2}1^T(\log(\mathrm{diag}(\boldsymbol{\Sigma}^{-1})) \\
&\quad - \mathrm{diag}(\boldsymbol{\Sigma}^{-1}) * (\mathbf{y}-\mathbf{F}\boldsymbol{\theta}-\mathbf{Mu}-\mathbf{R}\boldsymbol{\gamma})^2) + \log(\tau^2_{com})/2
\end{aligned}
\tag{4.45}
$$

As the covariance matrix is a diagonal matrix we can multiply the diagonal elements which we can see in the two last lines of (4.45), and were also applied previously Danaei et al. [2011], Finucane et al. [2011].
Again, we use random walk M-H for the joint update of τ^2_{nat}, τ^2_{subn}, τ^2_{com}. Due to symmetry of the proposal density it is cancelled out in the r ratio. Finally, we use only the posterior distribution in the r probability and we accept $\tau^2_{nat}, \tau^2_{subn}, \tau^2_{com}$ with probability r shown in (4.46).

$$
\begin{aligned}
r &= \frac{Posterior^* \times Proposal}{Posterior \times Proposal^*} \\
r &= \frac{P(\tau^{2*}_{nat}, \tau^{2*}_{subn}, \tau^{2*}_{com}|\mathbf{y})q(\tau^2_{nat}|\tau^{2*}_{nat})q(\tau^2_{subn}|\tau^{2*}_{subn})q(\tau^2_{com}|\tau^{2*}_{com})}{P(\tau^2_{nat}, \tau^2_{subn}, \tau^2_{com}|\mathbf{y})q(\tau^{2*}_{nat}|\tau^2_{nat})q(\tau^{2*}_{subn}|\tau^2_{subn})q(\tau^{2*}_{com}|\tau^2_{com})} \\
r &= \frac{P(\tau^{2*}_{nat}, \tau^{2*}_{subn}, \tau^{2*}_{com}|\mathbf{y})}{P(\tau^2_{nat}, \tau^2_{subn}, \tau^2_{com}|\mathbf{y})} \\
\log r &= \log P(\tau^{2*}_{nat}, \tau^{2*}_{subn}, \tau^{2*}_{com}|\mathbf{y}) - \log P(\tau^2_{nat}, \tau^2_{subn}, \tau^2_{com}|\mathbf{y})
\end{aligned}
\tag{4.46}
$$

For the two-dimensional equivalent we need to repeat the procedure three times for the three variables, DBP, SBP and INT.

4.4 Application of two-dimensional model: simulated data

To test the model and its implementation, we simulated data from 61 countries using the NCD Risk Factor Collaboration regional classification for blood pressure, Risk Factor Collaboration [2017]. We randomly generated data from a Normal distribution for each year in each country to cover each of the eight different age groups. We set as mean, values close to the average DBP and SBP respectively changing every time the value of variance for each country, year and gender. Moreover, we randomly chose some of the studies to have increased blood pressure measurements in the older age groups. For the INT variable we used the product of DBP and SBP.
We assume we have one study per year. This study contains eight mean values of

blood pressure measurements. These eight mean values are for the eight different age groups which are 18-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79 and > 80 years. The years we study are 2003-2007 which are five years in total, therefore, we have five different studies in five years for each country.

We separate the globe into 21 regions. Among the 21 regions is the region of high-income Asia Pacific which contains only one country, Japan, in the simulated data. The rest of the regions all include more than three countries. Therefore, we generated data from three countries in each of the 20 regions and one country for one of the regions hence, we have 61 countries with data in total. In summary, every country has information for five years and each year has information for the eight age groups.
Overall, the simulated data present ideal conditions since we have covered most of the uncertainty. That is because all the regions have three countries with data. All these countries have data for five years and each year contains data for each of the eight different age groups. Since, we simulated SBP and DBP data for five years for each country this is the fourth scenario in Section 4.3.2.2 of the restrictions that apply. Nonetheless, there is still some uncertainty that has not been covered. We still have substantial sparsity in our dataset since we do not have data for all the year combinations. We split the data into training and test sets, such that most countries would be represented in the training set. As Japan was the only country in its region, it was included in each training set.

We carried out five folds of cross-validation, with the average coverage and errors shown for super-regions in Table 4.2. The coverage of actual uncertainty by estimated credible intervals is generally more than 95% (i.e. conservative), although with cases where the coverage is somewhat less (i.e., anti-conservative). As shown in Table 4.2, the relative errors are small for both DBP, SBP and the interaction.

This can also be seen in Figure 4.1, which provides an example of model fitting for 50-year-old females in Belgium, where the patterns observed in the data are complex. The x-axis has values from 1 to 5 that encode the five years that are the same for both DBP and SBP. As explained before we have data for five years out

TABLE 4.2: Coverage and error in cross-validation

Regions	Coverage DBP	Coverage SBP	Coverage Int	Relative Error DBP	Relative Error SBP	Relative Error Int
Total	0.984	0.986	0.948	-0.002	-0.002	-0.045
Central and Eastern Europe	0.994	1	0.952	-0.003	-0.001	-0.017
C. Asia, Middle East & N. Africa	0.990	0.987	0.865	-0.002	-0.005	0.031
East and South East Asia	0.955	0.997	0.897	-0.003	-0.001	-0.03
High-income Western countries	0.998	0.994	0.996	0	-0.001	0.046
Latin America and Caribbean	0.978	0.982	1	-0.001	0	-0.151
Oceania	0.970	1	0.955	-0.002	-0.002	-0.05
South Asia	0.990	0.983	0.830	-0.004	-0.003	0.05
Sub-Saharan Africa	0.990	0.982	0.954	-0.002	-0.002	-0.095

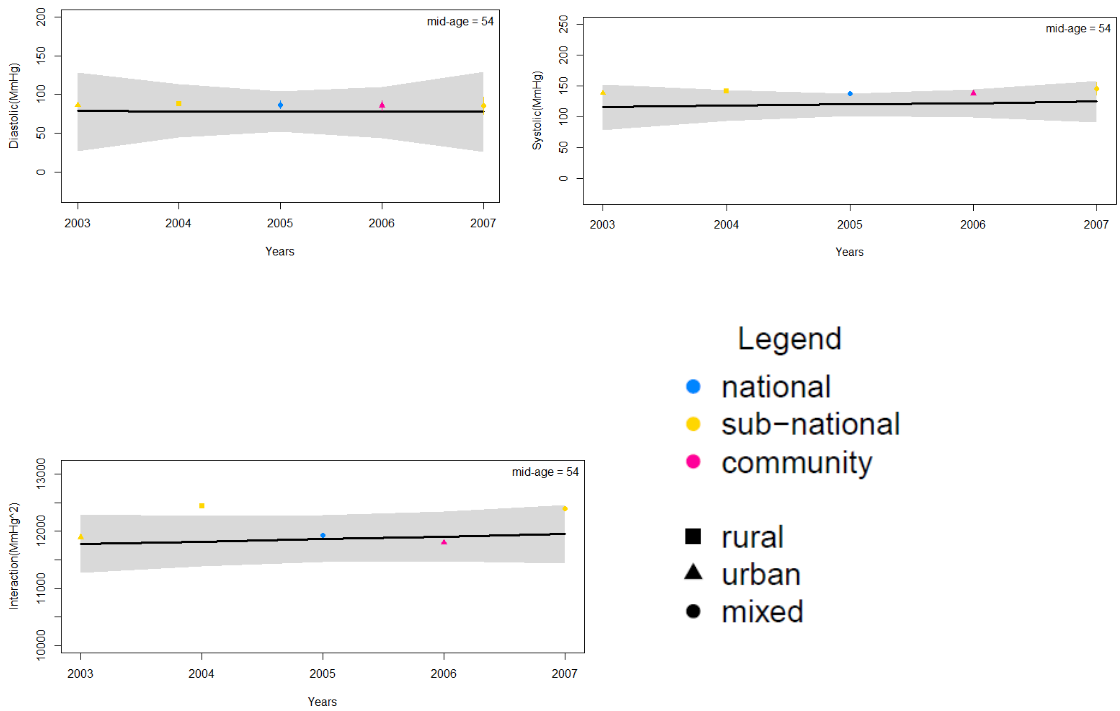

FIGURE 4.1: Fit plots for 50-year-old females in Belgium from the second set of cross validation. The solid line represents the posterior mean and the shaded area the 95% credible interval.

of the 25 possible combinations. These years are when both DBP and SBP have blood pressure measurements in the same year, i.e. $(2003, 2003), \ldots (2007, 2007)$. Although the simulated data are for five of these combinations, as described above, this allows us to make estimates for the other possible combinations. More specifically in Figure 4.1 the plots of DBP, SBP and interaction show that their datapoints for each variable are inside the credible interval. On the other hand, DBP's datapoints are inside the estimation's credible interval and quite close to the regression line; however, the datapoints for SBP appear to be somewhat further away. Finally, the third plot of Figure 4.1 depicts the estimations of DBP's and SBP's interaction. We can observe that the estimations are not as good a fit as the DBP and SBP since one of the datapoints is not in the credible interval. However, overall, the majority of the datapoints in the three variables DBP, SBP, INT are included in the credible intervals that our calculations provide.

4.5 Application of the two-dimensional blood pressure model: real data

Performance of the model using real blood pressure data was tested using the STEPwise Approach to NCD Risk Factor Surveillance (STEPS) studies, whose results are made publicly available by WHO
(https://www.who.int/teams/noncommunicablediseases/ surveillance/data). We downloaded data from 38 countries for the period 2004 to 2014 that comprised 478 country-year-sex-age combinations. Data were available from each year but were unevenly spread over time, with particularly large numbers of data points in 2007 and 2014. While a large proportion of the data comprised national-level estimates, 3% of the data points were at subnational level, and 6% were from rural-only or urban-only samples. Data were available for ages from 15 to 74 years, with the largest samples being for age groups in the range from 25 to 64 years. Sample sizes ranged from 46 to 3,036, while the ranges for DBP and SBP were 59.4-96.7 mmHg

and 93.5-155.2 mmHg, respectively. The data were from low-income and middle-income countries in central, south, southeast and east Asia, eastern Europe, Latin America and the Caribbean, the Middle East and north Africa, Oceania and sub-Saharan Africa. Although we had data for 38 countries, we made estimates for 48 countries in total from those regions, with cases where countries had data for zero, one, two or three or more years, i.e., each of the cases described in Section 4.3.2.2. The precision matrix analysed here included 11 years, therefore was 121 $\times$ 121 dimensional, and while this increased the time required for the MCMC algorithm to converge and mix, it was still tractable using a high-performance computing system. Figures 4.2, 4.3, 4.4 and 4.5 show examples of model fit for Myanmar, Nepal, North Korea and Kuwait respectively which exhibit realistically smooth but non-linear trends in the three blood pressure variables simultaneously. These effects, and others seen for different countries, would not have been captured fully by the models used in previous estimation of blood pressure, Finucane et al. [2014], Danaei et al. [2011].

Analytically, Figure 4.2 shows that the interaction, INT between SBP and DBP is greatest where SBP and DBP have the lowest and greatest value respectively. In more detail, for Myanmar, Figure 4.2, there is an increase in the INT through the years. We can observe there is a decrease in the values of SBP through the years while DBP does not change significantly, it is as if remains the same. On the contrary in Figure 4.3 and 4.5 we observe that Nepal and Kuwait exhibit a different behaviour. Specifically, for Nepal INT increases through time while DBP increases and SBP remains relatively the same. Same on Kuwait INT increases too. Although both DBP and SBP have increased values the change that is observed is not big. Both for Nepal and Kuwait the interaction between DBP and SBP is greatest where these values are both greatest in both figures. Lastly, in Figure 4.4 we observe North Korea which presents similar behaviour to Myanmar (Figure 4.2). For North Korea however, a decrease in interaction is noticed as there is a significant change between DBP and SBP values. DBP values increase a lot whereas SBP values decrease through the years. We can see that variability exists between countries for example, Myanmar and Nepal (Figures 4.2 and 4.3) which is not concerning as each country can follow a different lifestyle. On the other hand, we can also observe that there are countries with similar behaviour such as Nepal and Kuwait, Figures 4.3 and 4.5. Figures 4.2-4.5 illustrate the differences and the similarities that can exist between countries. That is a result of similar or different lifestyle, culture or urbanization. A priori a specific grouping among countries has been used and that can be reflected in the a posteriori results of the model.

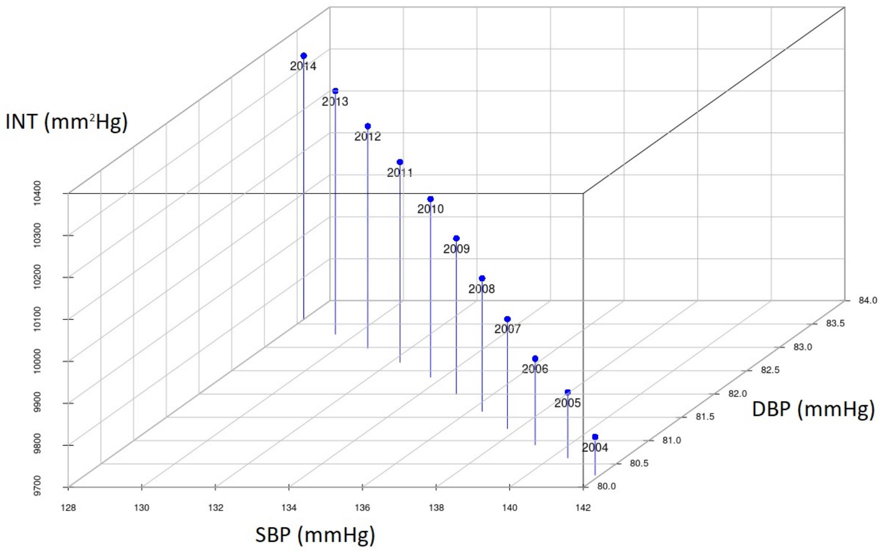

FIGURE 4.2: Estimates for 50-year old males in Myanmar from 2004 to 2014.

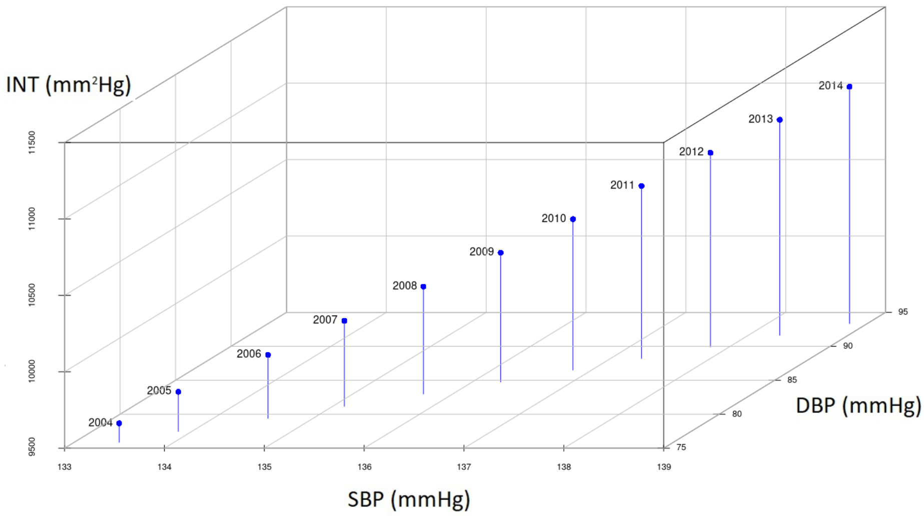

FIGURE 4.3: Estimates for 50-year old males in Nepal from 2004 to 2014.

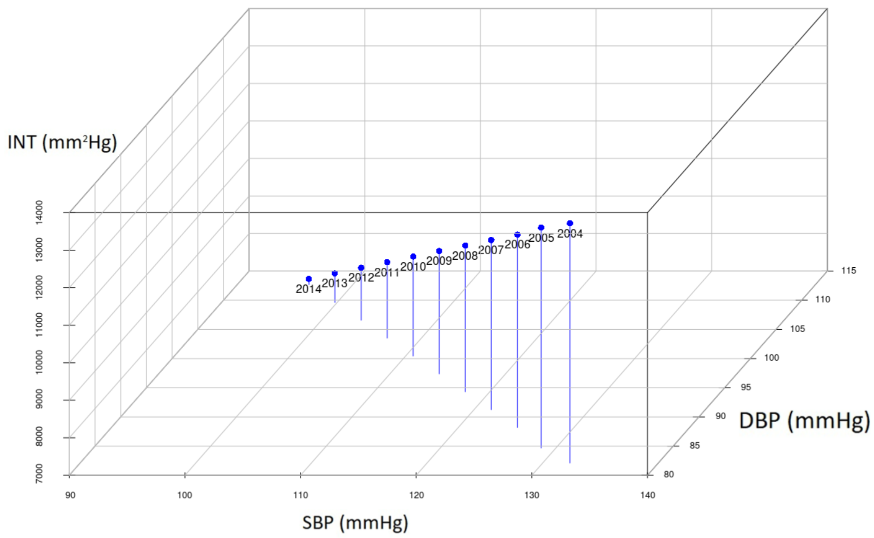

FIGURE 4.4: Estimates for 50-year old males in North Korea from 2004 to 2014.

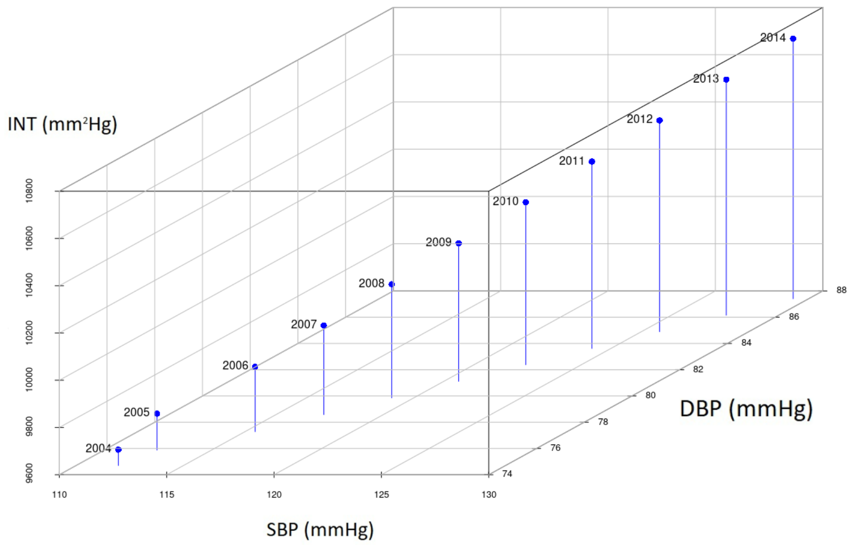

FIGURE 4.5: Estimates for 50-year old males in Kuwait from 2004 to 2014.

Overall, it is highlighted that the interaction is not the same for all the countries. Specifically, in Figures 4.2, 4.3 and 4.5 the interaction increases with time whereas in Figure 4.4 it diminishes through time.

In conclusion, interaction, INT expresses the relation between DBP and SBP. If both DBP and SBP increase or decrease then INT will increase as well whereas if they have a big difference of negative change, i.e. one increases and the other

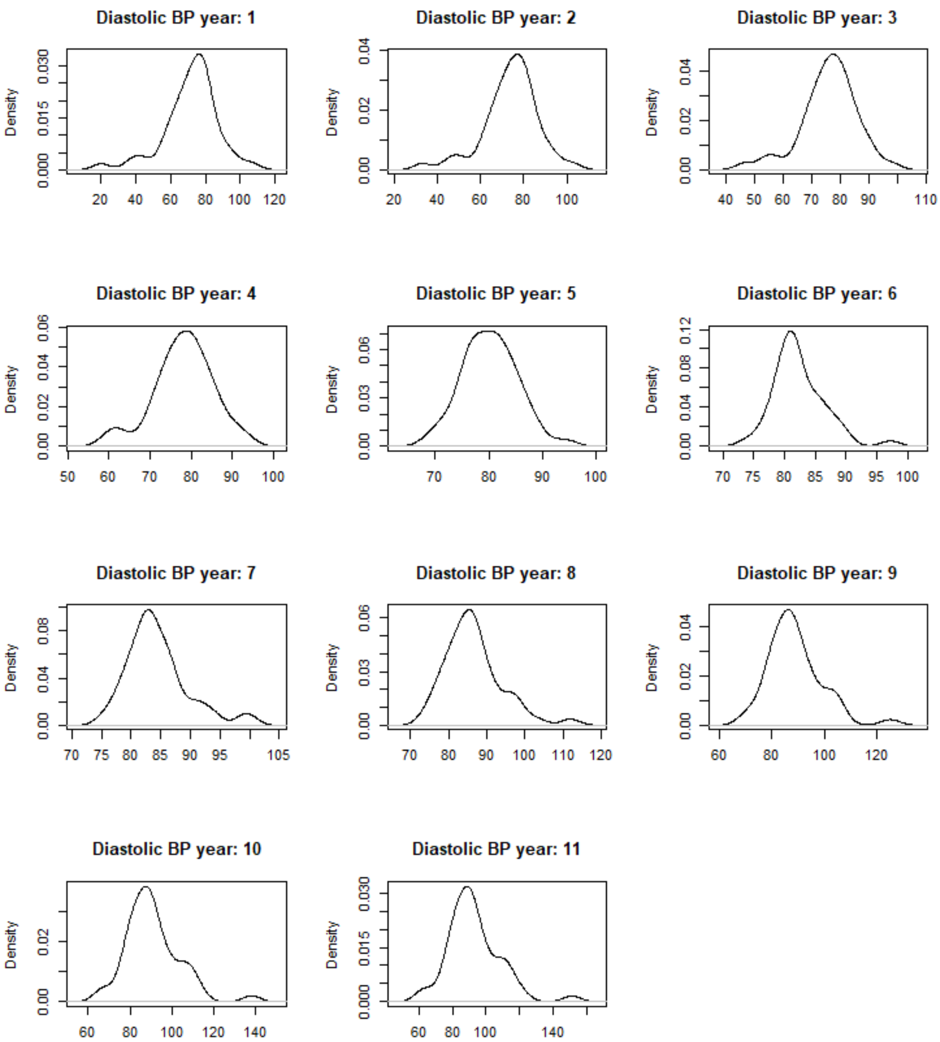

FIGURE 4.6: Posterior Distribution for DBP variable for all the countries together for each year.

decreases then INT will decrease through time. Moreover, in the Figures 4.6, 4.7 and 4.8 we can see the posterior distributions for each of DBP, SBP and INT taking all the countries together for the years 2004-2014. Posterior plots for all the countries together for each of the 11 years are added. Analytically, in Figure 4.6 we can observe that DBP overall has increased from year 1 to year 11. In years 1-6 the peak of DBP was around 80 mmHg, in years 7-8 was to 85 mmHg whereas in years 9-11 the peak of DBP was around 85 mmHg. A similar pattern is observed in SBP, Figure 4.7. In more detail, in the years 1-6 the peak of SBP is observed at the 130 value, in the 7-9 years is increased to 140mmHg whereas for the last two years, 10-11 the value has increased to 145-150 mmHg. Finally, for the interaction

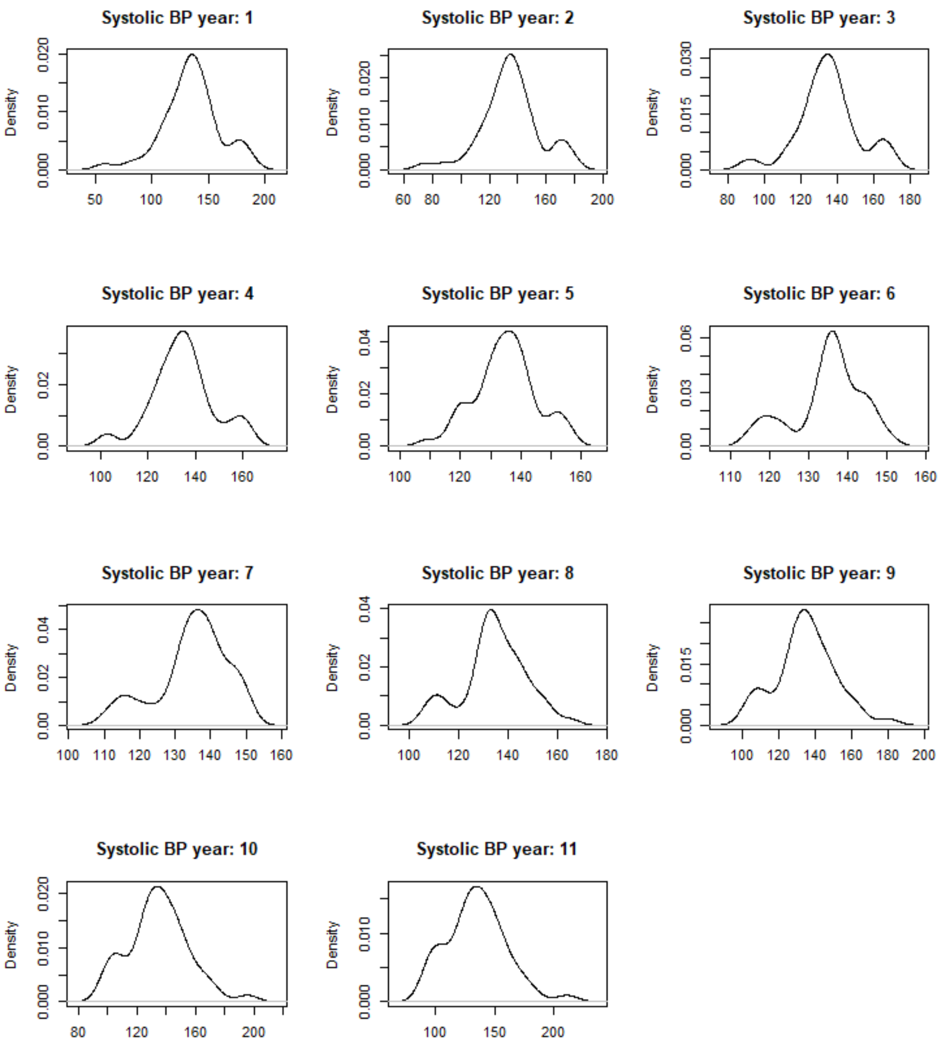

FIGURE 4.7: Posterior Distribution for SBP variable for all the countries together for each year.

plots, Figure 4.8 we can observe that interaction has increased as well throughout the years. Precisely, for the first five years the peak was observed at the value of 10000, for the years 6-9 the peak was at the 10200 value whereas at the final two years the value was at the 11000 value. Also, it is important to mention that the range of the values has increased as well in DBP, SBP and INT through time. Summing up, the contribution of this model is the ability to capture interaction effects that are present but cannot be captured using the existing one-dimensional model.

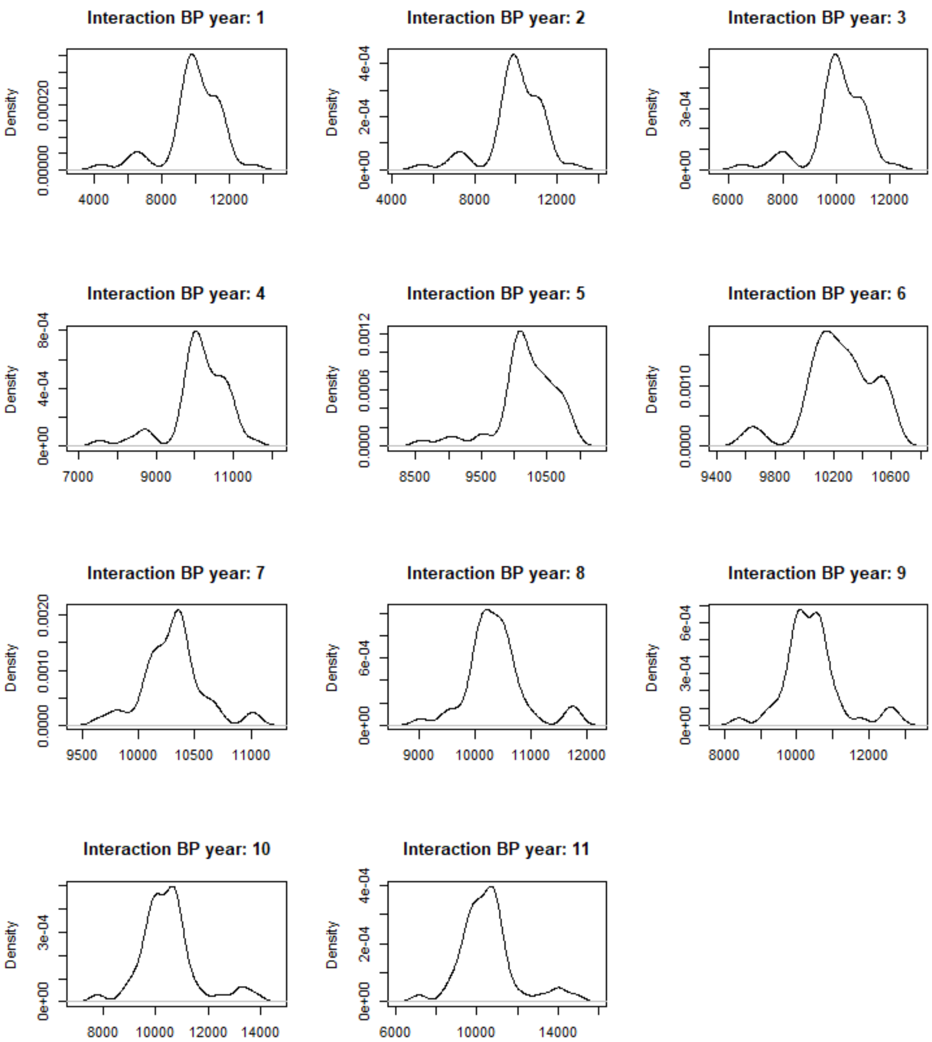

FIGURE 4.8: Posterior Distribution for INT variable for all the countries together for each year.

4.5.1 Diagnostics

In this section we present the verification of our results by applying suitable diagnostics for our analysis, Gelman et al. [1997]. Through these diagnostics we provide an indication that the MCMC chains have converged and they have mixed well. For the validity of the model we also implement cross-validation for the simulated data and calculate the deviance in each iteration for simulated and real data, Gelman et al. [2014]. For the real dataset we ran 10 different chains of which we keep three since, these three have converged and mixed much better than the rest and therefore are considered to be appropriate for our analysis. We present plots and diagnostic tests to observe the behaviour for each of the parameters that we estimated for the results of the real data. The diagnostics are presented below:

TABLE 4.3: Potential scale reduction factors for the model's parameters

Parameter	Point est.	Upper C.I
Deviance	1.02	1.06
$\theta_{Interc,DBP}$	1	1
$\theta_{Interc,SBP}$	1	1
$\theta_{Interc,INT}$	1	1
$\theta_{slope,DBP}$	1.01	1.01
$\theta_{slope,SBP}$	1.03	1.03
$\theta_{slope,INT}$	1	1
$\gamma_{1,DBP}$	1	1.01
$u_{[1]}$	1.01	1.01
$v_{[1]}$	1.015	1.015
$sv_{[1]}$	1.02	1.02
$w_{[1]}$	1.11	1.12

1. **Gelman-Rubin Diagnostic**

   The first is the Gelman-Rubin diagnostic, Gelman and Rubin [1992], Gelman et al. [2013]. It is known as the scale reduction factor diagnostic. In more detail, the Gelman-Rubin diagnostic makes a comparison of within-chain and between-chain variance, where a large deviation between these two variances indicates non-convergence. Therefore, desirable values are considered from 1 to 1.1 which indicate convergence. In Table 4.3 and 4.4 we can see the potential scale reduction in the model's parameters and hyper-parameters respectively. In Table 4.3 most of the parameters have a point estimate less than 1.1 except the non-linear parameter for the globe of the first year under study, $w_{[1]}$ which has a value slightly more than 1.1 which is 1.12. In Table 4.4 the hyper-parameters are observed to have less close to ideal behaviour than the parameters since the hyper-parameter of the slope in the linear sub-model, κ_b for DBP has a large value. However, values of these hyperparameters do not influence pearameters of interest strongly.

Table 4.4: Potential scale reduction factors for the model's hyperparameters

Hyperparameter	Point est.	Upper C.I
$\kappa_{a,DBP}$	1.11	1.33
$\kappa_{b,DBP}$	3.83	17.1
λ_c	1	1
λ_r	1	1.01
λ_s	1	1
λ_g	1	1
$\sigma^2_{1,DBP}$	1.22	1.72
τ^2_{DBP}	1	1.01
τ^2_{SBP}	1	1
τ^2_{INT}	1	1.05

Table 4.5: $\frac{N_{eff}}{N}$ in total and for each of the three chains

Parameter	$\frac{N_{eff}}{N}$:Total	Chain1	Chain2	Chain3
Deviance	0.46	0.64	0.64	0.1
θ	0.78	1	1	0.34
$\gamma_{1,DBP}$	0.41	0.47	0.42	0.34
$u_{[1]}$	1.08	0.92	1.1	1.23
$v_{[1]}$	1.08	0.92	1.1	1.23
$sv_{[1]}$	1.13	1.61	1	0.79
$w_{[1]}$	0.53	0.57	0.53	0.49

Table 4.6: $\frac{N_{eff}}{N}$ in total and for each of the three chains

Hyperparameter	$\frac{N_{eff}}{N}$:Total	chain1	chain2	chain3
$\kappa_{a,DBP}$	0.02	0.01	0.006	0.04
$\kappa_{b,DBP}$	0.002	0.003	0.0005	0.002
λ_c	0.2	0.22	0.23	0.16
λ_r	0.23	0.26	0.24	0.19
λ_s	0.27	0.3	0.29	0.21
λ_g	0.37	0.41	0.39	0.32
τ^2_{DBP}	0.51	0.49	0.48	0.54
τ^2_{SBP}	0.36	0.34	0.36	0.36
τ^2_{INT}	0.57	0.71	0.7	0.31

2. **Effective Sample Size (ESS)**

   We can use the fraction ESS which is the count of effective sample size divided by the total number of iterations, $\frac{N_{eff}}{N}$, Vehtari [2021]. The effective sample size expresses which of the iterations' values are considered not correlated out of the total number of the values. Values with a proportion close to 0.5 and above are considered desirable since it means that more than half of the values are uncorrelated. In Table 4.5 we can observe the ESS for each of the three chains. Overall it appears that three chains give an appropriate sample size for each of the parameters but those that perform better are the first and the second chains. For example the Deviance for the first chain has a 0.64 value which means the 64% of its sample is uncorrelated. In Table 4.6 we observe the ESS for the hyperparameters of the model. Similarly to the Gelman-Rubin Diagnostic we can observe that there is a smaller effective sample size in the hyperparameters than in parameters. The smaller effective sample size is observed in κ hyperparameter.

3. **Traceplots**

   Traceplots are used to identify whether the chains have mixed well together and have converged to a stationary distribution. In Figure 4.9 we can see how the deviance, the intercepts and the slope for each of DBP, SBP and INT in linear sub-model behave keeping 3000 iterations with thinning. In Figures 4.9, 4.10, 4.11 we can see that the three chains have mixed well together and converged.

4. **Autocorrelation**

   With the help of the autocorrelation plot we can see how correlated is the sample for each of the parameters. In Figure 4.12 we observe the autocorrelation plots for the deviance and the parameters of the intercept and slope for the three variables in the linear sub-model. We can observe that there is not a strong correlation within the sample in each of the parameters. Furthermore, Figure 4.13 shows the autocorrelation plots of the parameters λ_c, λ_r, λ_s and λ_g which are the precision parameters of the non-linear sub-model. We can notice a stronger autocorrelation in Figure 4.13 than in Figure 4.12. However, in both plots the autocorrelation does not remain high for long and it diminishes after the tenth lag.

4.6 Summary

This chapter provides a detailed description of the distributions for each of the sub-models of our Bayesian hierarchical model of blood pressure in two dimensions. In Section 4.2 the construction of the posterior and proposal distribution of the one-dimensional linear sub-model is implemented and the way in which they are included in the acceptance ratio of the M-H algorithm is described. Following, Section 4.3 presents the analytical calculations of the posterior, the proposal distribution and the acceptance ratio of the M-H algorithm for both, one- and two-dimensional non-linear sub-models. Moreover, Section 4.3.3 highlights the full conditional distribution of the age sub-model using a Gibbs sampler algorithm at this point. Next, Sections 4.3.4, 4.3.5 and 4.3.6 present the posterior and proposal distribution of the hyperparameters σ^2, ν and τ^2 respectively using the M-H algorithm. Lastly, Section 4.4 describes the application of the model to simulated data whereas Section 4.5 illustrates the application to real data showing also the diagnostics for each of the parameters. Overall, this chapter is a guide to the computational methods that were used for each sub-model in the MCMC algorithms, as well as showing that the model fits appropriately to the more complex case of two-dimensional data.

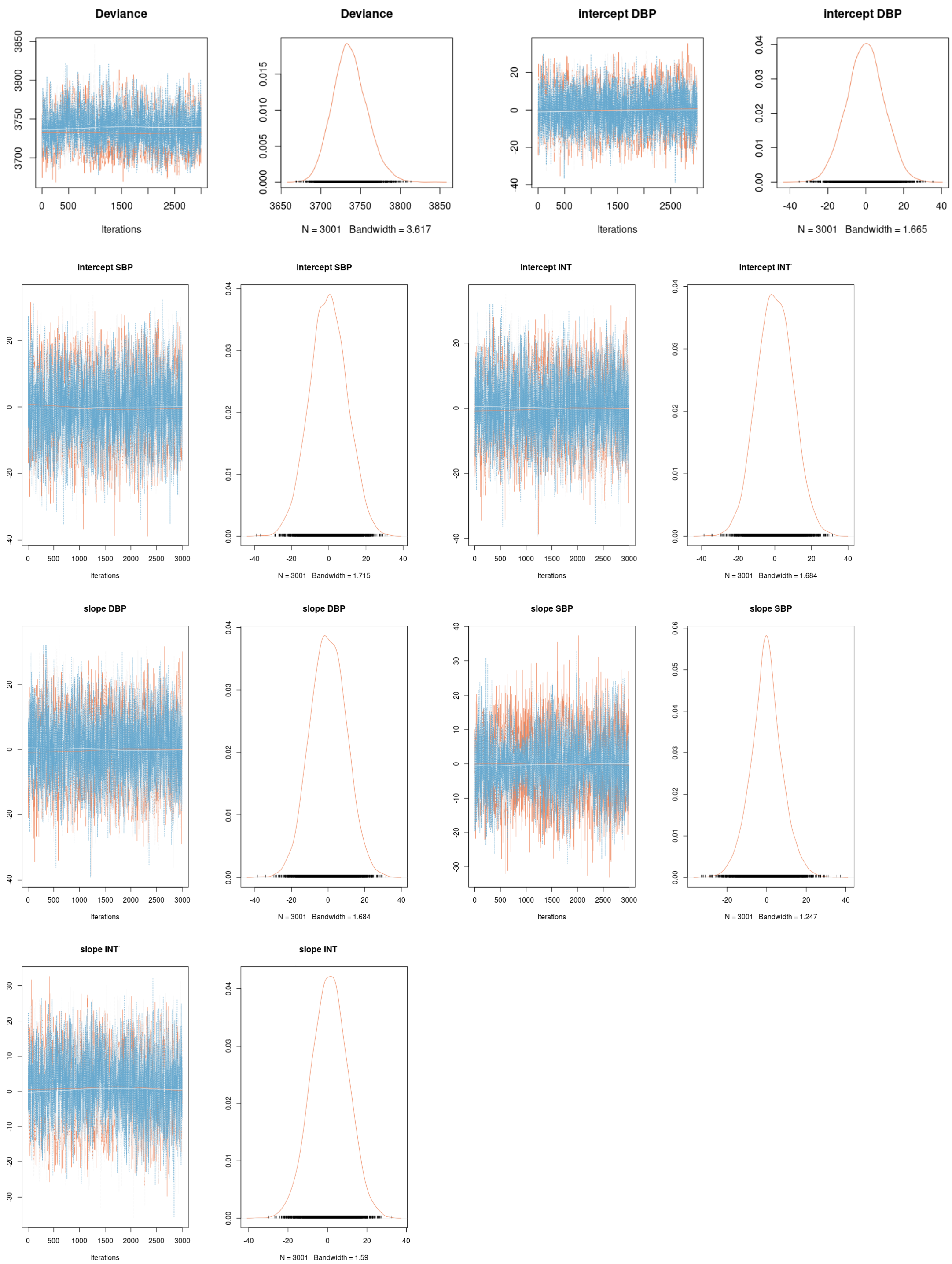

FIGURE 4.9: Traceplots for deviance, intercept and slope of the linear model for each of DBP, SBP, INT.

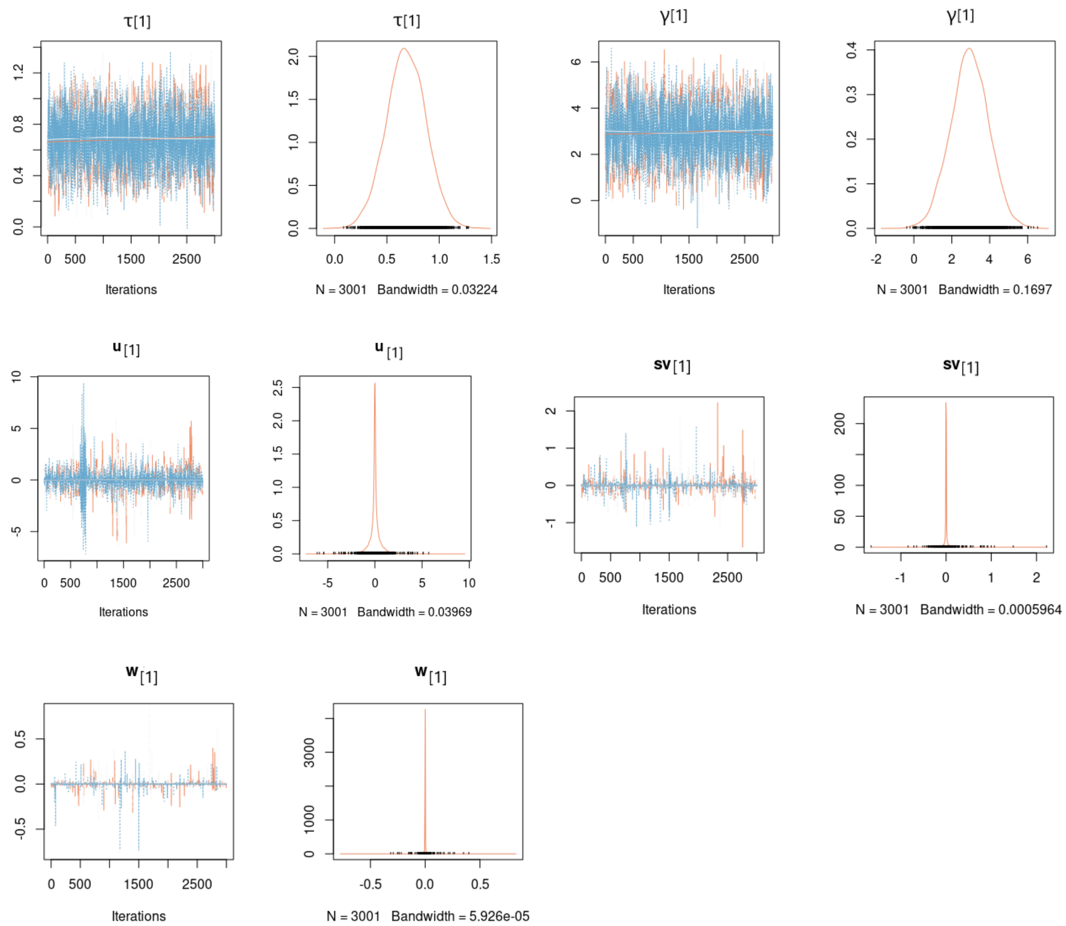

FIGURE 4.10: Traceplots for the first value of the vectors for the τ^2,γ, u, v and w parameters.

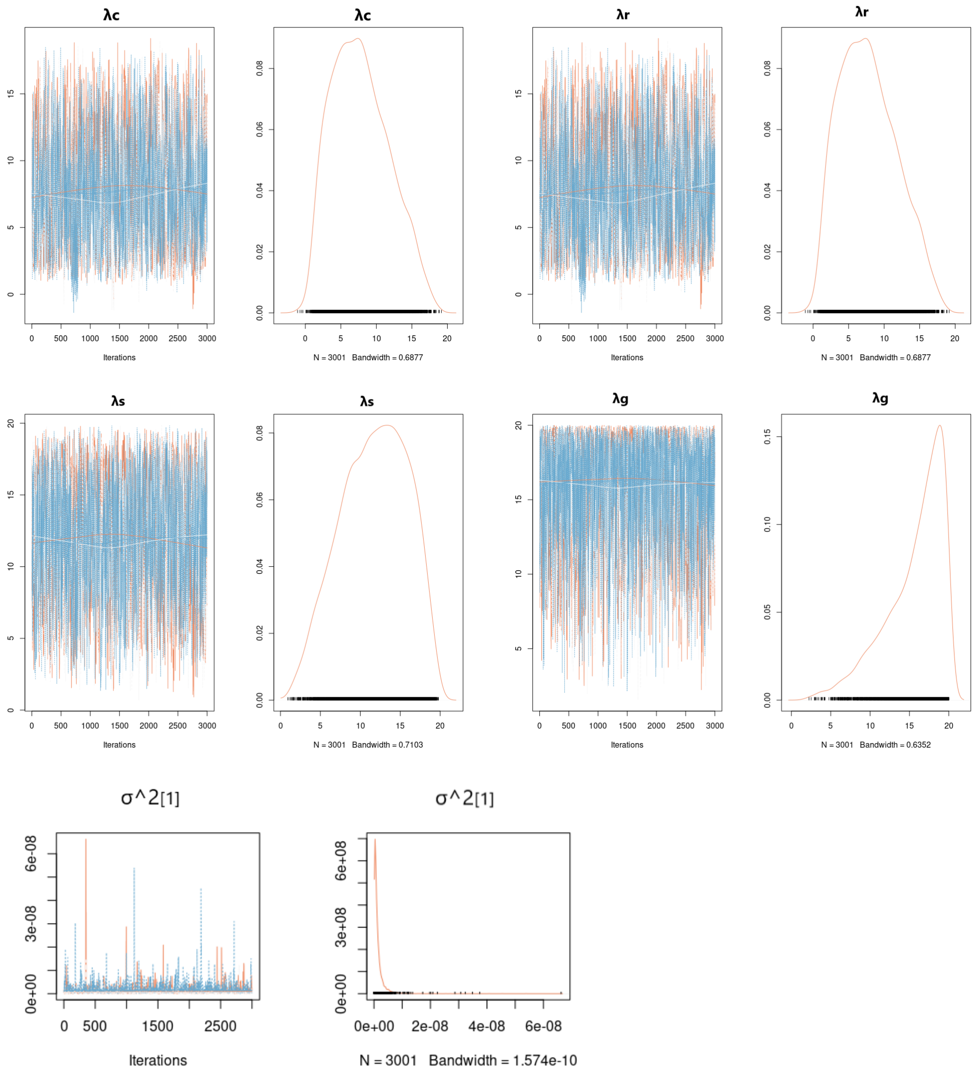

FIGURE 4.11: Traceplots for the scalar hyperparameters of the non-linear model, λ_c, λ_r, λ_s,λ_g and the first value of the vector for σ^2 parameter

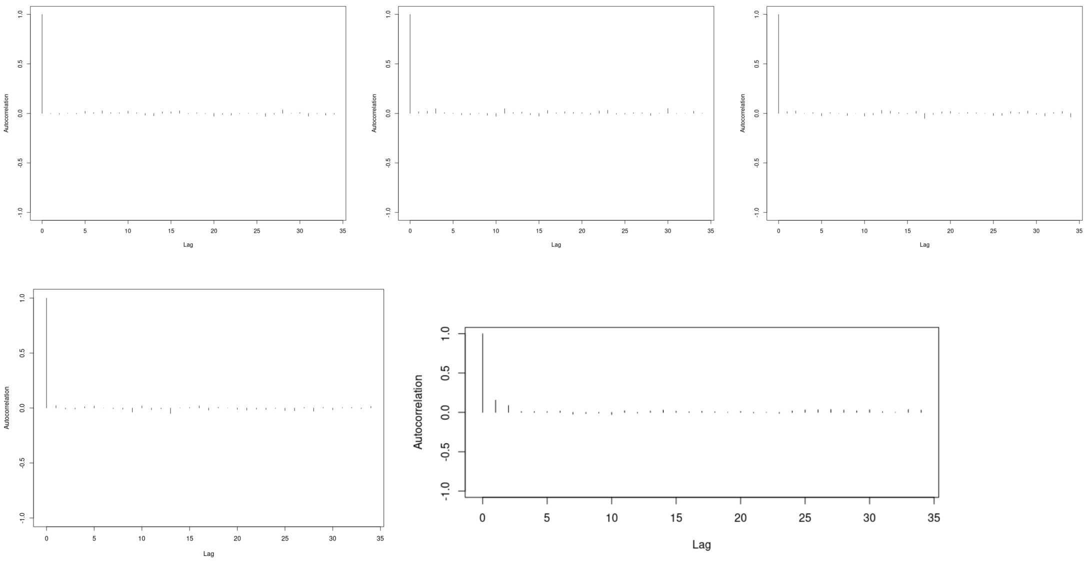

FIGURE 4.12: Autocorrelation plots for theta and deviance parameter.

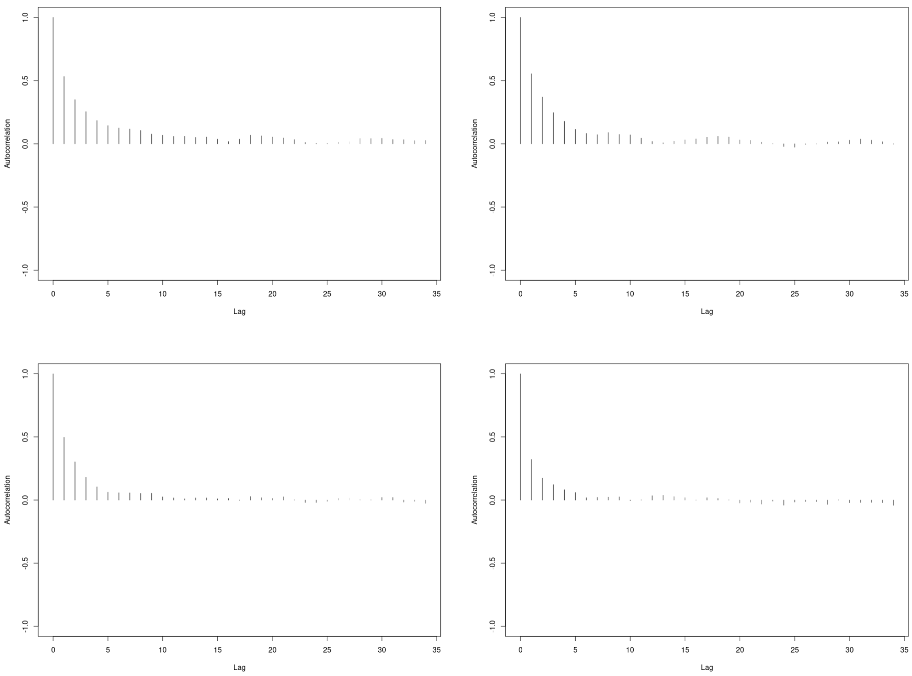

FIGURE 4.13: Autocorrelation plots for the scalar hyperparameter of the non-linear model, λ_c, λ_r, λ_s, λ_g.

Chapter 5

Defining priors for IGMRFs

5.1 Introduction

In the previous chapters we have seen how each of the sub-models which comprise the model are defined. Specifically, substantial attention has been given to the non-linear sub-model (see Sections 2.6.5, 3.4.1, 4.3). Through these sections we have discussed the one-dimensional first and second order IGMRF (or random walk) and a variety of types for a two-dimensional second order IGMRF (or random walk) for the non-linear sub-model (see Sections 2.5, 2.6). We have seen that given the number of years (i.e., nodes) included in the analysis, the size of the precision matrix changes accordingly. Therefore, as is stated in Sørbye and Rue [2014] the priors that we use should also be changed depending on the precision matrix that is used. In other words, since there are substantial differences from one structure matrix to the other it makes sense for the priors between these two precision matrices to be different. Additionally, these changes can be even larger when we change from a one- to two-dimensional IGMRF. In conclusion, as is highlighted in Sørbye and Rue [2014] and as we will observe later in this chapter the prior can vary depending on the number of nodes we include each time both in one- and two-dimensional IGMRFs. Therefore, we will discuss the priors that are referring to the precision (scaling) parameters which are used in the precision matrices of the non-linear sub-model and how they change from one to two dimensions, Spyropoulou and Bentham [2022a].

5.2 Scaling Hyper-priors

As we have mentioned previously, when the precision matrices change, the hyper-priors set for the precision parameters need to change as well. Let us say we have two different precision matrices and we are interested in constructing the hyper-prior distributions of the two precision parameters. In order to apply the same degree of (hyper-)prior information we need to implement a proper scaling. In other words, we need to apply a scaling to the hyper-prior of the first precision parameter in order to be used best for the new second precision parameter. In that way, there is a difference between the hyper-priors but without losing the properties that we desire them to follow.

5.2.1 Reference standard deviation for one-dimensional IGMRF

A prior distribution for an IGMRF can be described using the following definition:

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, (\lambda \mathbf{P})^{-1}) \rightarrow \sigma^2_\lambda(u_i) = \lambda^{-1}\Sigma^{2*}_{ii} \tag{5.1}$$

where $\mathbf{P}$ is the structure matrix of the precision matrix and Σ_{ii} is the diagonal element of the covariance matrix in position i such that $P = \Sigma^{-1}$. For the standardized normal distribution, we have

$$\mathbf{u}\sqrt{\lambda} \sim \mathcal{N}(\mathbf{0}, \mathbf{P}^{-1}), \quad \lambda\sigma^2_\lambda(u_i) = \Sigma^*_{ii} \tag{5.2}$$

Let us say that $\lambda = 1$ in (5.2) then we have:

$$\sigma^2_{\{\lambda=1\}}(u_i) = \Sigma^*_{ii} \tag{5.3}$$

Combining the results in (5.1) and (5.3) we have:

$$\sigma^2_\lambda(u_i) = \lambda^{-1}\Sigma^*_{ii}, \quad \sigma^2_\lambda(u_i) = \lambda^{-1}\sigma^2_{\{\lambda=1\}}(u_i) \tag{5.4}$$

This means that for any fixed precision parameter λ, the marginal standard deviation of the components of a Gaussian vector $\mathbf{u}$ can be expressed, as a function of λ, (Sørbye and Rue [2014]) by

$$\sigma_\lambda(u_i) = \frac{\sigma_{\{\lambda=1\}}(u_i)}{\sqrt{\lambda}}, \quad i = 1, ..., n \tag{5.5}$$

For a given IGMRF $\mathbf{u}$ with random precision λ, we can calculate a reference standard deviation for fixed $\lambda = 1$ and then approximate the marginal standard deviation for each component of $\mathbf{u}$ (Sørbye and Rue [2014]) by

$$\sigma_\lambda(u_i) \approx \frac{\sigma_{ref}(\mathbf{u})}{\sqrt{\lambda}}, \quad i = 1, ..., n \tag{5.6}$$

A suggestion for a reference standard deviation given by Sørbye and Rue [2014] is to calculate the geometric mean, which is an appropriate measure for a set of positive numbers. Hence, the reference standard deviation for $\mathbf{u}$ in the one-dimensional case is presented as

$$\sigma_{ref}(\mathbf{u}) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\sigma_{\{\lambda=1\}}(u_i)\right) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\Sigma^*_{ii}\right) \tag{5.7}$$

where the values Σ^*_{ii} denote the diagonal elements of the inverse matrix $\mathbf{\Sigma}^* = \mathbf{Q}^{-1}$ calculated for $\lambda = 1$. In particular, Σ^*_{ii} express the diagonal elements of the inverse of the precision matrix and this is calculated as $Q^- = \mathbf{\Gamma}^T\mathbf{\Lambda}^-\mathbf{\Gamma}$ in which $\mathbf{\Gamma}$ expresses the eigenvectors and $\mathbf{\Lambda}$ the eigenvalues of the $\mathbf{Q}$ matrix when $\lambda = 1$. Since $\mathbf{Q}$ and similarly $\mathbf{\Sigma}^*$ are $n \times n$ dimensional in the first and second order of the one-dimensional IGMRFs we will have n diagonal values to calculate the geometric mean from. A further consideration is that for any IGMRF, we must take into account linear restrictions when calculating reference standard deviations so that the latter are finite. Specifically, in the one-dimensional case for the first-order

IGMRF we need to set the last eigenvalue to infinity and for the second-order IGMRF we set the last two eigenvalues to infinity or we set to zero the last and the last two inverse of eigenvalues respectively.

5.2.2 Reference standard deviation for two-dimensional IGMRFs

For two-dimensional second-order IGMRFs, there is an extension for $\sigma_{ref}(\mathbf{u})$ that is presented in (5.7). The precision matrix is $(n_1 \times n_2) \times (n_1 \times n_2)$ dimensional, where n_1 and n_2 are the total number of nodes for the first and second variables respectively. Therefore, the scaling will no longer be for n_1 or n_2 values, but their product $n_1 \times n_2$ and we have

$$\sigma_{ref}(\mathbf{u}) = \exp\left(\frac{1}{n_1 \times n_2}\sum_{i=1}^{n_1 \times n_2} \log \sigma_{\{\lambda=1\}}(u_i)\right) = \exp\left(\frac{1}{n_1 \times n_2}\sum_{i=1}^{n_1 \times n_2} \log \Sigma^*_{ii}\right) \tag{5.8}$$

with a special case when $n = n_1 = n_2$

$$\sigma_{ref}(\mathbf{u}) = \exp\left(\frac{1}{n^2}\sum_{i=1}^{n^2} \log \sigma_{\{\lambda=1\}}(u_i)\right) = \exp\left(\frac{1}{n^2}\sum_{i=1}^{n^2} \log \mathbf{\Sigma}^*_{ii}\right) \tag{5.9}$$

Again, $\mathbf{\Sigma}^*_{ii}$ denotes the diagonal elements of the inverse matrix $(\mathbf{\Sigma}^*)^{1/2} = (\mathbf{Q}^{-1})^{1/2}$ whereas $\mathbf{\Sigma}^*$ is calculated as $\mathbf{\Sigma}^* = \mathbf{Q}^{-1} = \mathbf{\Gamma}^T \Lambda^{-1} \mathbf{\Gamma}$ for $\lambda = 1$. The precision matrix, $\mathbf{Q}$ and therefore, $\mathbf{\Sigma}^*$ is $n^2 \times n^2$ dimensional hence we will have n^2 elements in the diagonal.

Again we have linear restriction to take into account. For the two-dimensional second-order IGMRF we set the three last eigenvalues to infinity Rue and Held [2005] or another way of implementing the linear restrictions is setting to zero the inverse of the eigenvalues.

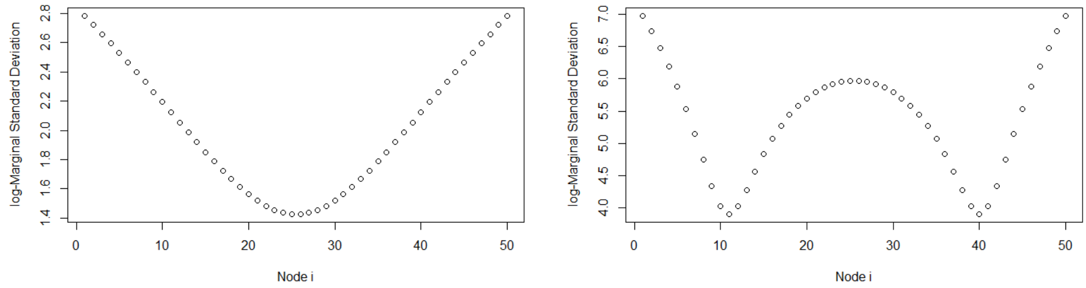

FIGURE 5.1: The marginal standard deviation of a one-dimensional first- and second-order IGMRFs, calculated using fixed precision λ=1.

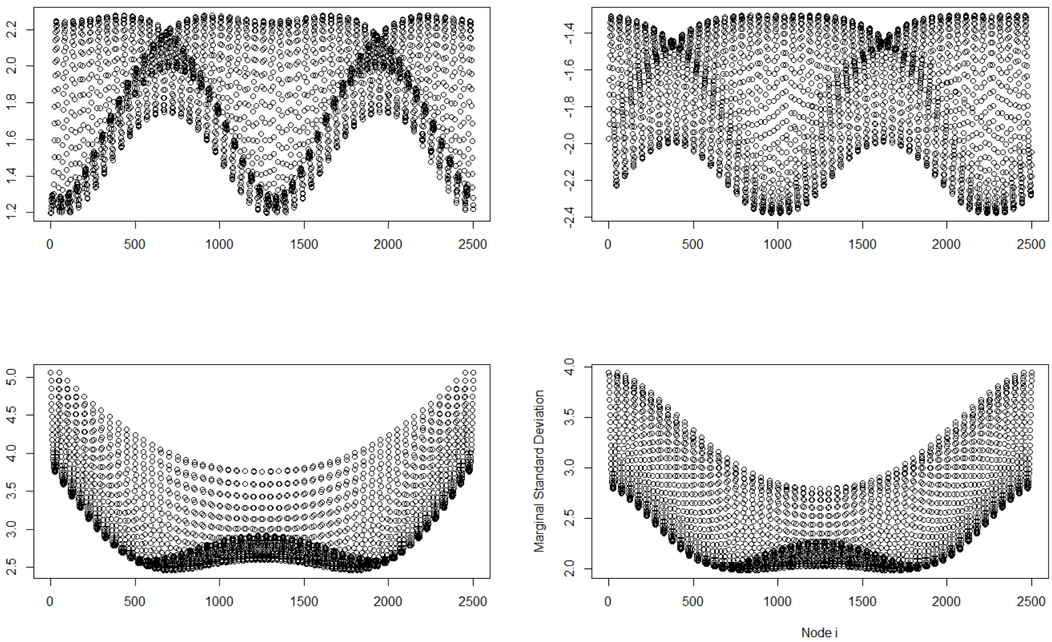

FIGURE 5.2: The marginal standard deviation of a two-dimensional second-order IGMRF random walk, calculated using fixed precision λ=1.

5.2.3 Interpretation of reference standard deviation

In this section we compare the behaviour of each of the IGMRF types. Figure 5.1 illustrates the diagonal values of the marginal standard deviations using fixed precision of $\lambda = 1$ for all the components of the one-dimensional first- and second-order IGMRFs on a line represented in Section 2.5.1 and 2.5.2. In Figure 5.2 we see four different types of a two-dimensional second-order IGMRF. These four types are known as Torus 1, Torus 2, Bound 1 and Bound 2 which are explained in

Section 2.6.2, 2.6.3, 2.6.5 and 2.6.4 respectively. Again, in Figure 5.2 the diagonal values of the marginal standard deviations using fixed precision $\lambda = 1$ are presented as nodes. By observing both Figures 5.1 and 5.2 we can conclude that each type of IGMRF has different behaviour and interpretation and thus we need different priors for each case to properly capture the IGMRFs' behaviour.

Both one-dimensional types of IGMRF are defined with $n = 50$ equidistant nodes in Figure 5.1. In the two-dimensional second-order IGMRF, that is equivalent to defining 50 nodes for each variable which means $n = n_1 = n_2 = 50$ equidistant nodes. Therefore, the precision matrix will be $n^2 \times n^2$ giving a $(50^2 \times 50^2)$-dimensional precision matrix. In Figure 5.2 we can see that the number of nodes is 2500 since we account for every possible combination between the nodes of the first variable and the nodes of the second variable.

Given the different structure of these six types, the shape and the level of these curves are quite different. As has been stated previously Sørbye and Rue [2014], the reference standard deviations for the one-dimensional types are equal to σ_{ref} $(\mathbf{u}_{rw1} = 2.75)$ and σ_{ref} $(\mathbf{u}_{rw2} = 14.64)$ for the first-order and the second-order respectively using 50 nodes. Applying our suggestion of (5.9) for the two-dimensional second-order IGMRF the reference standard deviation for each of the four types is: for Torus 1, (Section 2.6.2) σ_{ref} $(\mathbf{u}_{rw2D_1} = 2.51)$, for Torus 2 (Section 2.6.3) σ_{ref} $(\mathbf{u}_{rw2D_2} = 0.41)$, for Bound 1 (Section 2.6.5) σ_{ref} $(\mathbf{u}_{rw2D_{terz}} = 4.83)$ and for Bound 2 (Section 2.6.4) σ_{ref} $(\mathbf{u}_{rw2D} = 3.63)$. In each of the two-dimensional cases the number of nodes is still 50 for each of the two dimensions, i.e. $50^2 = 2500$. Therefore, if we choose the same hyperprior for each case, it allows larger variance in the one-dimensional second-order IGMRF than the two-dimensional equivalent, for each of the four types, and more still than the one-dimensional first-order case. These results are in agreement with other results of Lindgren and Rue [2008], Lindgren et al. [2011] which show that for a one-dimensional first-order IGMRF model, the precision using the new nodes is $(k\lambda)^{-1}$, for a one-dimensional second-order IGMRF model, the precision using the new nodes is $(k^3\lambda)^{-1}$ and for the two-dimensional second-order IGMRF model, the precision using the new nodes is $(k^2\lambda)^{-1}$. Here, k is the number of equally sized subintervals between the original nodes x_1 and x_2 giving us as a result new equidistant nodes: $x_1^{'}, x_2^{'}, ..., x_{k+1}^{'}$. In Figure 5.3 the standard deviation for the second-order two-dimensional IGMRFs with bounds, s_{RW2D} and $s_{RW2Dterz}$ are always between those for the first and the second order one-dimensional IGMRFs, s_{RW1} and s_{RW2} while the number of nodes is increasing. In contrast, the two-dimensional IGMRFs defined on torus have a completely different behaviour. Finally, we can observe that depending on the number of nodes that we have, the standard deviation increases and varies by first-order, second-order and two-dimensional second order IGMRF. As shown in Figure 5.3, as the number of nodes increases, the effect of these differences becomes more pronounced where each line represents the (reference) standard deviation of IGMRF(or random walk) in the one-dimensional first and second order and in the two-dimensional second order for four different cases. Lastly, the procedure that

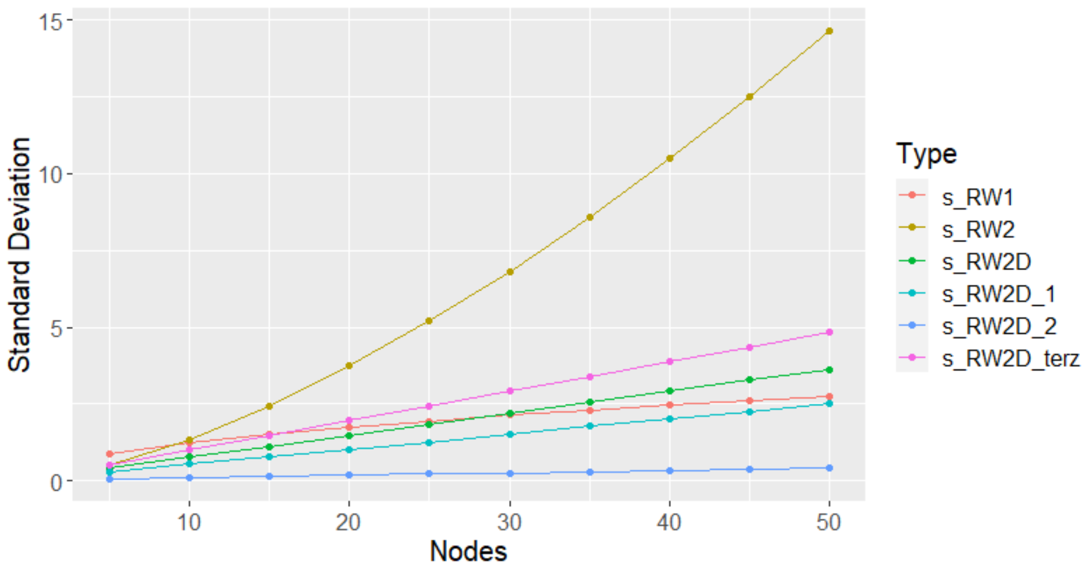

FIGURE 5.3: Change in marginal standard deviations depending on the number of nodes

we follow for creating the plot is summarised in the following steps:

1. Select the number of years under study

2. Compute the precision matrix for each type of IGMRF

3. Compute the reference standard deviation as indicated in (5.7), (5.9) for the first- and second- order one-dimensional IGMRFs and the four different choices of second-order two-dimensional IGMRFs.

We repeat these steps for a different number of years for each type of IGMRF.

5.2.4 Applying scaling using Gaussian hyperpriors

In order to account for the large differences that are seen in the calculated reference standard deviations as shown in the previous section and allow the same variance we need to impose an upper limit on the marginal standard deviation Sørbye and Rue [2014]. A priori, this indicates how large we allow the local deviation or the influence of the different random effects in a regression model to be. We denote the upper limit by U and define it as

$$P\left(\sigma(u_i) > U\right) \approx P\left(\frac{\lambda}{\sigma^2_{ref}(\mathbf{u})} < \frac{1}{U^2}\right) = \alpha \tag{5.10}$$

where α is a small probability Sørbye and Rue [2014]. By assigning a hyperprior to $\lambda\left(\sigma^2_{ref}(\mathbf{u})\right)^{-1}$ the interpretation of the hyperprior remains the same for different models. Having again the prior:

$$\mathbf{u} \sim \mathcal{N}(\mathbf{0}, (\lambda \mathbf{P})^{-1}) \tag{5.11}$$

Let us assume that the hyperprior of λ follows a Gaussian distribution

$$\lambda \sim \mathcal{N}(\mu, b^2) \tag{5.12}$$

then the upper limit in (5.10) using a probabilistic form can be written as

$$\begin{aligned} P\left(\frac{\sigma^2_{ref}(\mathbf{u})}{\lambda} > U^2\right) &= \alpha \Rightarrow P\left(\lambda < \frac{\sigma^2_{ref}(\mathbf{u})}{U^2}\right) = \alpha \\ F_\lambda\left(\frac{\sigma^2_{ref}(\mathbf{u})}{U^2}\right) &= \alpha \Rightarrow \frac{\sigma^2_{ref}(\mathbf{u})}{U^2} = F_\lambda^{-1}(\alpha) \\ U^2 &= \frac{\sigma^2_{ref}(\mathbf{u})}{F_\lambda^{-1}(\alpha,\mu,b)} = \frac{\sigma^2_{ref}(\mathbf{u})}{(1/b)F_\lambda^{-1}(\alpha,\mu,1)} \\ U &= \left(\frac{b\sigma^2_{ref}(\mathbf{u})}{F_\lambda^{-1}(\alpha,\mu,1)}\right)^{1/2} \end{aligned} \tag{5.13}$$

where $F^{-1}()$ denotes the quantiles of a normal distribution. For a given value of α, we can then interpret the mean and standard deviation parameters μ and b in terms of this upper limit. This is also useful to gain an intuition of which values of U seem reasonable for a specific application.

To recalculate hyperpriors for different IGMRFs, we can use the same mean parameter, μ, for each model and calculate a new standard deviation parameter, b. By using the upper limit provided in (5.13) the new standard deviation parameter denoted as b_{new} is expressed as:

$$b_{new} = \frac{U^2 F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw2})} \tag{5.14}$$

We see now that the new value of the standard deviation in (5.14) depends on the σ^2_{ref} terms which capture the precision matrix for a specific type of IGMRF. In conclusion, it is then only necessary to recalculate the standard deviation parameter, b, to account for the different shapes and sizes of the graph for a specific IGMRF (Sørbye and Rue [2014]).

This can be done for the three types of IGMRF considered using

$$\begin{aligned} \frac{b_{rw2}}{b_{rw1}} &= \frac{U^2 F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw2})} \Big/ \frac{U^2 F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw1})} \\ b_{rw2} &= b_{rw1}\frac{\sigma^2_{ref}(x_{rw1})}{\sigma^2_{ref}(x_{rw2})} \end{aligned} \tag{5.15}$$

and for the two-dimensional type

$$b_{rw2D} = b_{rw2} \frac{\sigma^2_{ref}(x_{rw2})}{\sigma^2_{ref}(x_{rw2D})} \qquad \text{or} \qquad b_{rw2D} = b_{rw1} \frac{\sigma^2_{ref}(x_{rw1})}{\sigma^2_{ref}(x_{rw2D})} \tag{5.16}$$

Here *rw1* is referring to the first-order IGMRF, *rw2* to the second-order IGMRF and *rw2D* is referring to the two-dimensional second-order IGMRF.

5.2.4.1 Types of two-dimensional second order IGMRFs

We can also compare IGMRFs with fixed order and dimensionality, but different numbers of nodes and boundary conditions. We do so for two-dimensional second order IGMRFs with four different structure matrices: Torus 1 and Torus 2 Rue and Held [2005], Thon et al. [2012] (see Section 2.6.2 and 2.6.3), Bound 1 Yue and Speckman [2010] and Bound 2 Terzopoulos [1988] (see Sections 2.6.5 and 2.6.4). Torus 1 has a structure matrix defined on a torus, while Torus 2 has a similar structure matrix but with boundaries at its four corners, $u_{1,1}$, $u_{n1,1}$, $u_{1,n2}$, $u_{n1,n2}$. Bound 1 and Bound 2 have boundary effects and induce the same neighbours in the structure matrix for each node, but give different weightings to these neighbours. In Table 5.1 we see that Torus 2 consistently has the lowest reference standard deviation, with the changes in each IGMRF being similar proportionally when the number of nodes is increased. The torus models are simpler as they do not consider different behaviour at the boundaries which explains the results. Comparing the Bound 1 and 2 precision matrices we expect Bound 2 to have a longer time of computing as it has a bigger order deficiency. Bound 2 has the largest reference standard deviation followed by Bound 1 which is used in our two-dimensional model of blood pressure Spyropoulou and Bentham [2022b,a]. Overall, the changes in each of the types are similar when the number of nodes increases. These findings show that it is clearly necessary to scale the hyperparameter each time the precision matrix or number of nodes is changed, especially when boundary conditions are introduced. Specifically, comparing the reference standard deviation horizontally we can notice differences from the type of the two-dimensional second order IGMRF we chose.

TABLE 5.1: Reference standard deviations σ_{ref} for second-order two-dimensional IGMRFs.

Nodes	Torus 1	Torus 2	Bound 1	Bound 2
5	0.30	0.05	0.41	0.53
6	0.35	0.06	0.48	0.63
8	0.44	0.08	0.62	0.81
10	0.54	0.09	0.76	1.00
11	0.58	0.10	0.83	1.10
12	0.63	0.11	0.90	1.19
14	0.73	0.12	1.04	1.39
16	0.82	0.14	1.19	1.58
18	0.93	0.16	1.33	1.77
20	1.02	0.17	1.47	1.96
25	1.25	0.21	1.83	2.44
30	1.51	0.25	2.19	2.92
40	2.00	0.33	2.91	3.88
50	2.50	0.42	3.63	4.84
100	5.00	0.83	7.24	9.64

5.2.5 Blood pressure data application

In this section we compare hyperprior scaling for one- and two-dimensional second order IGMRFs using blood pressure data Spyropoulou and Bentham [2022b,a]. In more detail, for the one-dimensional case, Section 3.1, we use a published model Finucane et al. [2014], Danaei et al. [2011] that employs a one-dimensional second order IGMRF as the prior for the non-linear sub-model for the SBP variable. For the two-dimensional case, Section 4, we use the model that we developed and is presented in Spyropoulou and Bentham [2022a,b] that employs a two-dimensional second order IGMRF as the prior for the non-linear sub-model for both SBP and DBP variables (and implicitly, their interaction).

Although the models are of different dimensionality, we can compare the hyperprior scaling for the precision parameter of the non-linear trends. This varies both by dimensionality and number of nodes, which in our data correspond to the number of years considered. The hyperpriors must be set for each of the precision parameters, λ_c, λ_r, λ_s and λ_g that are used at different levels of the hierarchical model: countries nested in regions, super-regions and globe both in one and two dimensions. In conclusion, we want to observe which variance suggestion is optimal for a specific number of years (nodes) to set for the hyperprior's distribution. Knowing the number of years we can easily find the σ_{ref} for both one- and two-dimensional second order IGMRFs.

5.2.5.1 Scaling using 5 years

Having the total number of years as 5 we can find the σ_{ref} using the formula on (5.9) as shown below:

$$\sigma_{ref}(u_{rw2}) = 0.52, \quad \sigma_{ref}(u_{rw2D}) = 0.41 \tag{5.17}$$

The distribution under consideration for a country's precision parameter is $\lambda_c \sim \mathcal{N}(\mu, b)$ with μ and b parameters that will be assigned as:

$$\mu = 7, \quad b = 2, \quad \alpha = 0.001 \tag{5.18}$$

Subsequently, the upper levels for both the one-dimensional and two-dimensional second-order IGMRF are:

$$\begin{aligned} U &= \left(\frac{b\sigma^2_{ref}(\mathbf{u})}{F^{-1}(\alpha, \mu, 1)}\right)^{1/2} \\ U_{rw2} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2})}{F^{-1}(\alpha, \mu, 1)}\right)^{1/2} = 0.377 \\ U_{rw2D} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2D})}{F^{-1}(\alpha, \mu, 1)}\right)^{1/2} = 0.293 \end{aligned} \tag{5.19}$$

By taking the median of these two upper levels we have:

$$U = \text{median}(U_{rw2}, U_{rw2D}) = 0.335$$

Hence, by using the aforementioned we can calculate the scaled new standard deviation, b, for the one- and two-dimensional second order IGMRF respectively.

$$\begin{aligned} b_{rw2} &= \frac{U^2 F^{-1}(\alpha, \mu, 1)}{\sigma^2_{ref}(\mathbf{u}_{rw2})} = 1.58 \\ b_{rw2D} &= \frac{U^2 F^{-1}(\alpha, \mu, 1)}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 2.61 \end{aligned} \tag{5.20}$$

An alternative method is to set directly $b_{rw2} = 1.58$ and use it for the calculation of b_{rw2D}:

$$\begin{aligned} b_{rw2} &= 1.58 \\ b_{rw2D} &= b_{rw2}\frac{\sigma^2_{ref}(\mathbf{u}_{rw2})}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 2.61 \end{aligned} \tag{5.21}$$

In our problem the variables of interest are the precision parameters for each of country, region, super-region and globe. However, we are not expecting all these four parameters to have the same prior distribution. If the prior's standard deviation in λ_g is equal to 3 this does not mean that λ_c, λ_r and λ_s will have the same value in the standard deviation.

Setting different values for the b_{rw2} we can see how the b_{rw2D} changes as well:

$$\begin{aligned} b_{rw2} &= 2 \\ b_{rw2D} &= b_{rw2}\frac{\sigma^2_{ref}(\mathbf{u}_{rw2})}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 3.3 \end{aligned} \tag{5.22}$$

$$\begin{aligned} b_{rw2} &= 1.5 \\ b_{rw2D} &= b_{rw2}\frac{\sigma^2_{ref}(\mathbf{u}_{rw2})}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 2.47 \end{aligned} \tag{5.23}$$

Given that we use the reference standard deviation we can make the two methods comparable even when we decide to change the number of nodes.

Attention is needed because if we increase the number of nodes the suggestion for the prior's variances changes as well. In the next subsection we observe how the standard deviations of the priors change when the years are equal to 10.

5.2.5.2 Scaling using 10 years

Considering that we have 10 years under study we first need to calculate the reference standard deviation for the one- and two-dimensional second order IGMRF:

$$\sigma_{ref}(\mathbf{u}_{rw2}) = 1.35, \quad \sigma_{ref}(\mathbf{u}_{rw2D}) = 1.76 \tag{5.24}$$

Following this, we find the upper levels for the one- and two-dimensional second order IGMRF:

$$\begin{aligned} U &= \left(\frac{b\sigma^2_{ref}(\mathbf{u})}{F^{-1}(\alpha,\mu,1)}\right)^{1/2} \\ U_{rw2} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2})}{F^{-1}(\alpha,\mu,1)}\right)^{1/2} = 0.96 \\ U_{rw2D} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2D})}{F^{-1}(\alpha,\mu,1)}\right)^{1/2} = 0.54 \end{aligned} \tag{5.25}$$

By taking the median of these two upper levels we have:

$$U = \text{median}(U_{rw2}, U_{rw2D}) = 0.754$$

$$\begin{aligned} b_{rw2} &= \frac{U^2F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw2})} = 1.22 \\ b_{rw2D} &= \frac{U^2F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 3.84 \end{aligned} \tag{5.26}$$

An alternative method is setting directly the standard deviation of the two dimensional second order random walk equal to 3.

$$\begin{aligned} b_{rw2D} &= 3 \\ b_{rw2} &= b_{rw2D}\frac{\sigma^2_{ref}(\mathbf{u}_{rw2D})}{\sigma^2_{ref}(\mathbf{u}_{rw2})} = 0.95 \end{aligned} \tag{5.27}$$

In conclusion observing the above results we can notice that the reference standard deviation in a total of five years presented in (5.17) is greater for the one-dimensional second order IGMRF but when the number of years is changing to 10 presented in (5.24) the reference standard deviation of the two-dimensional second order IGMRF becomes greater which states the variability of standard deviations for different types of IGMRFs when the number of years change.

5.2.5.3 Scaling using 20 years

Lastly, we present the results with a number of years equal to 20.

$$\sigma_{ref}(\mathbf{u}_{rw2}) = 3.72, \quad \sigma_{ref}(\mathbf{u}_{rw2D}) = 1.47 \tag{5.28}$$

Following, we find the upper levels for both of them:

$$\begin{aligned} U &= \left(\frac{b\sigma^2_{ref}(\mathbf{u})}{F^{-1}(\alpha,\mu,1)}\right)^{1/2} \\ U_{rw2} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2})}{F^{-1}(\alpha,\mu,1)}\right)^{1/2} = 2.66 \\ U_{rw2D} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2D})}{F^{-1}(\alpha,\mu,1)}\right)^{1/2} = 1.05 \end{aligned} \tag{5.29}$$

By taking the median of these two upper levels we have:

$$U = \text{median}(U_{rw2}, U_{rw2D}) = 1.859$$

$$\begin{aligned} b_{rw2} &= \frac{U^2F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw2})} = 0.97 \\ b_{rw2D} &= \frac{U^2F^{-1}(\alpha,\mu,1)}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 6.24 \end{aligned} \tag{5.30}$$

Again, setting directly the standard deviation of the two dimensional second order IGMRF equal to 4, we have:

$$\begin{aligned} b_{rw2D} &= 4 \\ b_{rw2} &= b_{rw2D}\frac{\sigma^2_{ref}(\mathbf{u}_{rw2D})}{\sigma^2_{ref}(\mathbf{u}_{rw2})} = 0.624 \end{aligned} \tag{5.31}$$

In the work presented previously Danaei et al. [2011], a prior is set for the precision parameter of the one-dimensional second-order IGMRF, λ_g with a value of 2 as the standard deviation. With the suggested scaling a suitable standard deviation for the two-dimensional second-order IGMRF equivalent is 3.30 for 5 years shown also in Table 5.3. However, if the number of years under consideration changes the scaling needs to change again.

In Table 5.2, we can observe the differences in the reference standard deviations when the number of years is 11 and 20. We compare the reference standard deviations between the first-order and the second-order IGMRFs and the two-dimensional second-order IGMRFs. We can see that the one-dimensional second-order IGMRF has the biggest changes as we can also observe in Figure 5.3. The more years the bigger the differences in the one-dimensional second-order IGMRF.

Table 5.2: Reference standard deviations, σ_{ref}, for models with multiple nodes.

Nodes	$\sigma_{ref}(\mathbf{u}_{rw1})$	$\sigma_{ref}(\mathbf{u}_{rw2})$	$\sigma_{ref}(\mathbf{u}_{rw2D})$
5	0.85	0.53	0.41
6	0.94	0.65	0.48
8	1.09	0.96	0.62
10	1.22	1.35	0.76
11	1.28	1.54	0.83
12	1.34	1.75	0.90
14	1.45	2.20	1.04
16	1.55	2.68	1.19
18	1.65	3.19	1.33
20	1.74	3.73	1.47
25	1.94	5.20	1.83
30	2.13	6.82	2.19
40	2.46	10.49	2.91
50	2.75	14.65	3.63
100	3.89	41.39	7.24

Lastly, in Table 5.3 we gather all the information for the standard deviation for each type of IGMRF to make the comparison more straightforward and to see how the standard deviation of the hyperprior, b_{rw1}, b_{rw2} and b_{rw2D} is tuned when the number of years is increased.

TABLE 5.3: Scaling the standard deviation parameters (hyperparameters), b_{rw1}, b_{rw2} and b_{rw2D} as the adjusted parameter b and number of nodes are varied.

Nodes	b=1			b=2			b=3		
	b_{rw1}	b_{rw2}	b_{rw2D}	b_{rw1}	b_{rw2}	b_{rw2D}	b_{rw1}	b_{rw2}	b_{rw2D}
5	0.39	1.00	1.67	0.78	2.00	3.34	1.17	3.00	5.01
6	0.48	1.00	1.84	0.96	2.00	3.67	1.43	3.00	5.49
8	0.78	1.00	2.40	1.55	2.00	4.80	2.33	3.00	7.19
10	1.00	0.82	2.58	2.00	1.63	5.16	3.00	2.45	7.74
11	1.00	0.69	2.39	2.00	1.39	4.78	3.00	2.08	7.17
12	1.00	0.59	2.22	2.00	1.17	4.43	3.00	1.76	6.65
14	1.00	0.43	1.94	2.00	0.87	3.89	3.00	1.30	5.83
16	1.00	0.33	1.70	2.00	0.67	3.40	3.00	1.00	5.09
18	1.00	0.27	1.54	2.00	0.53	3.08	3.00	0.80	4.62
20	1.00	0.22	1.40	2.00	0.43	2.80	3.00	0.65	4.20
25	1.00	0.14	1.12	2.00	0.28	2.25	3.00	0.42	3.37
30	1.06	0.10	1.00	2.11	0.21	2.00	3.17	0.31	3.00
40	1.40	0.08	1.00	2.80	0.15	2.00	4.20	0.23	3.00
50	1.74	0.06	1.00	3.48	0.12	2.00	5.23	0.18	3.00
100	3.46	0.03	1.00	6.93	0.06	2.00	10.39	0.09	3.00

In Table 5.3 we notice that the adjusted parameter, $b = 1$, $b = 2$ or $b = 3$ can be set as a final suggestion for one of the cases depending on the number of nodes selected each time. For example when the number of years are 5, 6, 8 and the adjusted standard deviation parameter is $b = 1$ then the scaled standard deviation parameter b_{rw1} is equal to 1 as well. However, when the number of nodes is increased to 10, 11 or 12 the b_{rw2} coincides with the adjusted scaled parameter $b = 1$ and so on. A similar behaviour can be observed when $b = 2$ or $b = 3$. In summary, the scaling is necessary when moving from the one-dimensional second-order IGMRFs to the two-dimensional second-order IGMRFs, Martino et al. [2011], Sørbye and Rue [2011] as we can see that there are cases where the scaled standard deviation parameters have values further away from the adjusted parameter. For instance, when the adjusted parameter is $b = 3$ for 10 nodes (or years) the scaled standard deviation for the one-dimensional second order is $b_{rw2} = 2.45$ whereas for the two-dimensional equivalent is $b_{rw2D} = 7.74$ which shows the big differences.

Table 5.4 shows the scaling that was applied to the standard deviation of λ_c,

Table 5.4: Scaling the standard deviation parameters (hyperparameters) b_{rw2} and b_{rw2D} for one and two dimensional IGMRFs for four different precision parameters λ_c, λ_r, λ_s and λ_g respectively having a different fixed value of the adjusted parameter b in each case with number of nodes equal to 11.

λ	b	b_{rw2}	b_{rw2D}
λ_c	0.9	0.53	1.83
λ_r	1.2	0.71	2.44
λ_s	1.59	0.94	3.24
λ_g	3.55	2.10	7.23

λ_r, λ_s and λ_g, the precision parameters of the non-linear sub-model in the one-dimensional second order (rw2) and the two-dimensional second order (rw2D) IGMRF models. Specifically, for the λ_c parameter having computed the upper levels of rw2 and rw2D using as adjusted parameter $b = 0.9$ the scaling for the standard deviation of λ_c parameter will be 0.53 in the rw2 model whereas in the rw2D model it will be 1.83. In that way we achieve the desirable acceptance ratio to be around 0.44 in MCMC as Gelman and Rubin [1992], Gelman et al. [1996] state for the one-dimensional parameters.

Analytically, in Table 5.4 we can see the values that we set for the standard deviations for each of the precision parameters of the second order two-dimensional model. We have analysed the one-dimensional second-order case having the b_{rw2} as standard deviation values for each of λ_c, λ_r, λ_s and λ_g. In order to apply the same degree of smoothness in the two-dimensional second-order case we know that the adjusted parameters b needs to be scaled properly for fitting best the two-dimensional equivalent. Therefore, b_{rw2D} values represent the scaled standard deviations that will be used for the modelling of the two-dimensional second order IGMRF.

5.2.6 Summary

This chapter explains analytically the use of scaling in prior distributions. In Sections 5.2.1 and 5.2.2 the use of the reference standard deviation is described in one and two dimensions respectively. Following, Section 5.2.3 compares the reference standard deviations between the different types of IGMRF in one dimension and the different types of IGMRF in two dimensions making also a comparison between the one and two dimensional cases using a number of different nodes. The next section, Section 5.2.4 defines the upper limit that is used in the scaling in order to apply the same degree of prior information to the models. Finally, Section 5.2.5 describes an application to blood pressure data both having one- and two- dimensional variables. Each step of the process is explained showing cases of using different number of years comparing each time the one-dimensional first and second order IGMRF and the two-dimensional second order IGMRF.

Chapter 6

Conclusions and Future work

6.1 Discussion

We have shown that it is possible to fit a model to mean blood pressure data in two dimensions, and in particular that it is possible to fit a non-linear surface that implicitly estimates the interaction between DBP and SBP variables. Our practical implementation will be of interest to any researchers with sparse bivariate data, but particularly to those with health-related data that vary by age and sex, over time and between countries, and where interactions are of substantial scientific importance.
This section presents the conclusions of this project as well as the challenges that we faced. It discusses the future work that can be applied and it ends with some final comments.

6.2 Conclusions

The variables of our problem are DBP, SBP and the interaction of these two. Based on the results we can notice that the interaction differs from country to country. We can observe that for some countries the interaction increases or decreases through time based on how large the variation is for each of the variables,

DBP, SBP, through time. The key challenge of the modelling is the choice of the prior distribution that will be used for the non-linear sub-model which if the modelling is successful, it provides us with useful information.
The non-linear sub-model accounts for non-linear trends over time. The time is expressed with the years under study hence, the starting and the final year are known for our study. Our prior is able to capture the relation between the years and this is important because although it is easy to have a specific pattern of relations in the central years the pattern is not the same in the boundary years. In other words, being in the first year we only have neighbours from the years ahead while being in the final year we only have information from the years before. Hence, the prior that we choose takes all these into consideration and provides a different behaviour depending on the year's position which gives a better fit for the model. Another difficulty is the volume of the data that we handle. The more years we take into consideration the more data we consider and thus, the more time we need to produce the final results. Furthermore, the two-dimensional model introduces a larger number of variables to be estimated. Moreover, the two-dimensional second order IGMRF used for one of the priors is itself computationally expensive as it accounts for every relation between the nodes of the first and the second dimension and it needs estimations for the squared number of the years under study. Furthermore, using MCMC techniques can slow down the procedure and the more data are included the more iterations will be needed to reach convergence of the algorithm.

The code that has been developed can be accessible online through the published papers as well as in the Github page in Spyropoulou [2022]. The interested user can adopt the methodology provided using any two-dimensional dataset where there is an interest in interaction.

The computation time of using the code to fit the model described in the thesis was three days and by using techniques such as Cholesky decomposition and sparse matrices memory issues were tackled. Moreover, by using parallel computing we were able to have results of 20 chains (or more, depending on the server is used and the user's choice) in three days as well focusing on the validation of the results.

6.3 Future Work

Future work could include systematic data collation and publication of estimates, as carried out previously for SBP and DBP as single variables. Additionally, there are also some interesting statistical challenges that are of note and are presented below:

1. The mean of an IGMRF
   The reason for using IGMRFs is that we wish to model the structure of the nodes in a specific random variable which is described through the precision matrix. Since IGMRFs can represent a graph we want to see which are the nodes that are connected hence dependent and which not. Therefore, all the analysis is focused on the precision matrix. However, there is not much presented in the literature about the mean of the IGMRFs. A question that might arise is what would be the case if the mean vector of an IGMRFs is different than zero. In other words, does the mean represent solely a shift (move) or there is a deeper meaning? Therefore, it would be interesting to explore the intuition and the properties of the mean as well in an IGMRF.

2. More than two dimensions
   After interacting with data which use IGMRFs in two dimensions it is normal to consider what is the extension of it. A question that may arise is what will be the procedure and methodology if we had three variables to work with. A possible suggestion is to have again all the possible combinations and those combinations will construct the precision matrix. So, for the case of three variables and each of them having five nodes then the possible combinations will be $5^3 = 125$, hence, the precision matrix will be 125×125 dimensional. Since, we have three variables there will be two kinds of interactions in the

125 combinations. A drawback that may arise is how convenient and fast will be this suggestion therefore, a limitation of this technique must be the increased dimensions of a two-dimensional matrix.

The aforementioned suggestion implies that we can express a problem of n dimensions to a problem of two dimensions. On the contrary, in machine learning techniques in order to have a solution to a problem we increase more and more the dimensions of the problem. For example, in Neural Networks or Support Vector Machines in order to achieve linearly separable groups you project your groups to higher dimensions. Furthermore, the illustration of the problem having more than two variables and more than one interaction is difficult to be applied and interpreted. It is cumbersome to extract any conclusion for more than three variables and it is confusing to get the intuition or the meaning for the relation of each variable.

3. Investigating R-INLA

   Other interesting research could explore alternatives to MCMC that would allow data for long time periods to be modelled in a tractable manner. Therefore, we are interested to see how this model can be fitted using integrated nested Laplace approximation in R (R-INLA) instead of MCMC methodology. R-INLA is an alternative and approximation of MCMC and it is used mainly in the IGMRF methods therefore, many issues theoretically could be applied more easily in the R-INLA than in the MCMC methodology.

4. Exploring the realm of PC priors

   As we have seen in Chapter 5 the prior distribution of the non-linear terms' hyperparameters need to be rescaled every time we change the number of nodes or the type of the IGMRF that we take into account. Therefore, it would be of interest to explore priors that are invariant to these changes. In Simpson et al. [2017] are proposed priors that are invariant to the change of the structure matrix which occurs when a different number of nodes is used. These are known as penalised complexity priors and their useful property is that they are unchanged with respect to the reparameterisation. Furthermore, one of the advantages of PC priors is the importance that they give to the base model, which is, the absence of random effects. It means that the new prior takes into account that there is a possibility the random effect is zero as well, i.e., there is a probability that the base model fits better. By referring to the base model we imply that the random effects are equal to zero, Simpson et al. [2017].

5. Different options of reference standard deviation
   Finally, in Chapter 5 we introduce and describe the procedure of scaling between models that can differ in dimensionality and the types of IGMRFs that are used. A possible suggestion would be to use a different function as a reference function, σ_{ref}. In Section 5.2 the reference standard deviation, σ_{ref} is calculated using a geometric mean which is considered suitable for positive numbers and hence we can explore more functions that are appropriate for positive numbers. Afterwards, it would be interesting to see the differences and the challenges that arise from using the new functions.

6.3.1 Final Comments

In conclusion, our fundamental problem is the structure of the prior for capturing the non-linear change over time. There are many factors that play a significant role in the solution of the problem. Specifically, the structure of the precision matrix depends on: 1) The number of years being dependent. In other words that is referring to the order of the random walk (IGMRF) where first or second order means that one or two years depend on the under-study year. 2) The number of dimensions. It answers the question for how many variables we want to observe the non-linear behaviour over time. After answering this question the next step is to define the number of years being dependent in each variable explained in 1. 3) The nature of dependence. That is referring to temporal or spatial dependence. It corresponds to the type of precision matrix which best explains the dependence that we want to capture and lastly, 4) The scaling of the precision parameters. After we have decided which precision matrix is suitable, care needs to be taken in defining the hyperprior. The hyperprior refers to the precision parameter and it is also susceptible to different dimensionality. Therefore, proper scaling needs to be applied tothe precision parameter going from one to two dimensions.

Bibliography

A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis (3rd ed.)*. Chapman and Hall/CRC, New York, USA, 2013.

Richard McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC, New York, USA, 2020.

Christian P Robert, George Casella, and George Casella. *Monte Carlo statistical methods*, volume 2. Springer, New York, USA, 1999.

Nicholas Metropolis and Stanislaw Ulam. The monte carlo method. *Journal of the American statistical association*, 44(247):335–341, 1949.

Andreĭ Andreevich Markov. An example of statistical investigation of the text Eugene Onegin concerning the connection of samples in chains. *Science in Context*, 19(4):591–600, 2006.

Siddhartha Chib and Edward Greenberg. Understanding the Metropolis-Hastings algorithm. *The American Statistician*, 49(4):327–335, 1995.

George Casella and Edward I George. Explaining the Gibbs sampler. *The American Statistician*, 46(3):167–174, 1992.

W Keith Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Oxford University Press*, pages 97–109, 1970.

Jarad Niemi. Metropolis-hastings explaination, 2019. Available at: https://www.jarad.me/courses/stat544/slides/Ch11/Ch11a.pdf (Accessed June 10, 2022).

Adrian FM Smith and Gareth O Roberts. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society: Series B (Methodological)*, 55(1):3–23, 1993.

A Gelman, G Roberts, and W Gilks. Efficient Metropolis Jumping Hules. *Bayesian Statistics*, (5):599–607, 1996.

Wally R Gilks, Nicky G Best, and Keith KC Tan. Adaptive rejection Metropolis sampling within Gibbs sampling. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 44(4):455–472, 1995.

S Geman and D Geman. Stochastic Relaxation, Gibbs Distributions and Bayesian Restoration of lmages. *IEEE Transactions on Patterns Analysis and Machine Intelligence*, 6:721–741, 1984.

Peter J Green and Antonietta Mira. Delayed rejection in reversible jump metropolis–hastings. *Biometrika*, 88(4):1035–1053, 2001.

Martyn Plummer et al. Jags: A program for analysis of bayesian graphical models using gibbs sampling. In *Proceedings of the 3rd international workshop on distributed statistical computing*, volume 124, pages 1–10. Vienna, Austria., 2003.

Bob Carpenter, Andrew Gelman, Matthew D Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan: A probabilistic programming language. *Journal of statistical software*, 76 (1), 2017.

Finn Lindgren and Håvard Rue. Bayesian spatial modelling with r-inla. *Journal of statistical software*, 63:1–25, 2015.

José C Pinheiro and Douglas M Bates. Linear mixed-effects models: basic concepts and examples. *Mixed-effects models in S and S-Plus*, pages 3–56, 2000.

Jonathan Bearak, Anna Popinchalk, Leontine Alkema, and Gilda Sedgh. Global, regional, and subregional trends in unintended pregnancy and its outcomes from 1990 to 2014: estimates from a bayesian hierarchical model. *The Lancet Global Health*, 6(4):e380–e389, 2018.

Andrew Gelman. Multilevel (hierarchical) modeling: what it can and cannot do. *Technometrics*, 48(3):432–435, 2006.

Widdow Quinn. Bayesian multilevel modelling using pystan, 2006. Available at: https://widdowquinn.github.io/Teaching-Stan-Hierarchical-Modelling/07-partial_pooling_intro.html (Accessed: July 11, 2022).

Florian Wilhelm. Finally!Bayesian Hierarchical Modelling at Scale, 2020. Available at: https://florianwilhelm.info/2020/10/bayesian_hierarchical_modelling_at_scale (Accessed: July 11, 2022).

Aram V Chobanian, George L Bakris, Henry R Black, William C Cushman, Lee A Green, Joseph L Izzo Jr, Daniel W Jones, Barry J Materson, Suzanne Oparil, Jackson T Wright Jr, et al. The seventh report of the joint national committee on prevention, detection, evaluation, and treatment of high blood pressure: the jnc 7 report. *Jama*, 289(19):2560–2571, 2003.

World Health Organization. Non-communicable diseases, 2021. Available at: https://www.who.int/news-room/fact-sheets/detail/noncommunicable-diseases (Accessed: June 1, 2022).

James Bentham, Gitanjali M Singh, Goodarz Danaei, Rosemary Green, John K Lin, Gretchen A Stevens, Farshad Farzadfar, James E Bennett, Mariachiara Di Cesare, Alan D Dangour, et al. Multidimensional characterization of global food supply from 1961 to 2013. *Nature Food*, 1(1):70–75, 2020.

Mariachiara Di Cesare, Maroje Sorić, Pascal Bovet, J Jaime Miranda, Zulfiqar Bhutta, Gretchen A Stevens, Avula Laxmaiah, Andre-Pascal Kengne, and James Bentham. The epidemiological burden of obesity in childhood: a worldwide epidemic requiring urgent action. *BMC Medicine*, 17(1):1–20, 2019.

Stefanie Vandevijvere, Lindsay M Jaacks, Carlos A Monteiro, Jean-Claude Moubarac, Martin Girling-Butcher, Arier C Lee, An Pan, James Bentham, and Boyd Swinburn. Global trends in ultraprocessed food and drink product sales and their association with adult body mass index trajectories. *Obesity Reviews*, 20:10–19, 2019.

Contributions of mean and shape of blood pressure distribution to worldwide trends and variations in raised blood pressure: a pooled analysis of 1018 population-based measurement studies with 88.6 million participants. *International journal of epidemiology*, 47(3):872–883i, 2018.

NCD Risk Factor Collaboration. Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants. *The Lancet*, 387(10027):1513–1530, 2016.

NCD Risk Factor Collaboration. Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. *The Lancet*, 389(10064):37–55, 2017.

NCD Countdown 2030. Worldwide trends in non-communicable disease mortality and progress towards sustainable development goal target 3.4. *The Lancet*, 392 (10152):1072–1088, 2018.

Benjamin R Fletcher, Jaime Hartmann-Boyce, Lisa Hinton, and Richard J McManus. The effect of self-monitoring of blood pressure on medication adherence and lifestyle factors: a systematic review and meta-analysis. *American journal of hypertension*, 28(10):1209–1221, 2015.

World Health Organization. Global Status Report on Noncommunicable Diseases 2018. Technical report, World Health Organization, Geneva, Switzerland, 2018.

Mariel M Finucane, Gretchen A Stevens, Melanie J Cowan, Goodarz Danaei, John K Lin, Christopher J Paciorek, Gitanjali M Singh, Hialy R Gutierrez, Yuan Lu, Adil N Bahalim, et al. National, regional, and global trends in body-mass index since 1980: systematic analysis of health examination surveys and epidemiological studies with 960 country-years and 9·1 million participants. *The Lancet*, 377(9765):557–567, 2011.

Goodarz Danaei, Mariel M Finucane, John K Lin, Gitanjali M Singh, Christopher J Paciorek, Melanie J Cowan, Farshad Farzadfar, Gretchen A Stevens, Stephen S Lim, Leanne M Riley, et al. National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. *The Lancet*, 377(9765):568–577, 2011.

Goodarz Danaei, Gitanjali M Singh, Christopher J Paciorek, John K Lin, Melanie J Cowan, Mariel M Finucane, Farshad Farzadfar, Gretchen A Stevens, Leanne M Riley, Yuan Lu, et al. The global cardiovascular risk transition: associations of four metabolic risk factors with national income, urbanization, and western diet in 1980 and 2008. *Circulation*, 127(14):1493–1502, 2013.

Jonathan Pearson-Stuttard, Bin Zhou, Vasilis Kontis, James Bentham, Marc J Gunter, and Majid Ezzati. Retracted: Worldwide burden of cancer attributable to diabetes and high body-mass index: a comparative risk assessment. *The Lancet Diabetes and Endocrinology*, 6:95–104, 2018.

NCD Risk Factor Collaboration. Rising rural body-mass index is the main driver of the global obesity epidemic in adults. *Nature*, 569(7755):260–264, 2019.

World Obesity Federation. Global Atlas on Childhood Obesity. Technical report, World Obesity Federation, London, UK, 2019.

Development Initiatives. Global Nutrition Report: Action on Equity to End Malnutrition. Technical report, Development Initiatives, Bristol, UK, 2020.

Mariel M Finucane, Christopher J Paciorek, Goodarz Danaei, and Majid Ezzati. Bayesian estimation of population-level trends in measures of health status. *Statistical Science*, pages 18–25, 2014.

Havard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC Press, New York, USA, 2005.

Yu Yue and Paul L Speckman. Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*, 19(1):96–116, 2010.

Kevin Thon, Håvard Rue, Stein Olav Skrøvseth, and Fred Godtliebsen. Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 56(1):49–61, 2012.

NCD Risk Factor Collaboration. Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *The Lancet*, 390(10113):2627–2642, 2017.

NCD Risk Factor Collaboration. Repositioning of the global epicentre of non-optimal cholesterol. *Nature*, 582(7810):73–77, 2020.

World Health Organization. Global Report on Diabetes. Technical report, World Health Organization, Geneva, Switzerland, 2016.

Petros Dellaportas and Gareth O Roberts. An introduction to MCMC. In *Spatial statistics and computational methods*, pages 1–41. Springer, 2003.

Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

Andrew Gelman and Donald B Rubin. Markov chain Monte Carlo methods in biostatistics. *Statistical Methods in Medical Research*, 5(4):339–355, 1996.

Nicholas J Higham. Analysis of the Cholesky decomposition of a semi-definite matrix. *Oxford University Press*, 1990.

Mahdi Jadaliha, Jinho Jeong, Yunfei Xu, Jongeun Choi, and Junghoon Kim. Fully bayesian prediction algorithms for mobile robotic sensors under uncertain localization using gaussian markov random fields. *Sensors*, 18(9):2866, 2018.

StackExchange. Torus graph, 2021. Available at: https://tex.stackexchange.com/questions/148510/how-to-plot-a-lattice-of-points-on-the-surface-of-a-torus.svg (Accessed: November 25, 2021).

Demetri Terzopoulos. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):417–438, 1988.

Paul Damien, Petros Dellaportas, Nicholas G Polson, and David A Stephens. *Bayesian theory and applications*. OUP Oxford, 2013.

Jim Griffin and Phil Brown. Hierarchical shrinkage priors for regression models. *Bayesian Analysis*, 12(1):135–159, 2017.

Erich L Lehmann and George Casella. *Theory of point estimation*. Springer Science & Business Media, 2006.

Guan-Hua Huang. Model identifiability. *Encyclopedia of Statistics in Behavioral Science*, 2005.

Alan E Gelfand and Sujit K Sahu. Identifiability, improper priors, and gibbs sampling for generalized linear models. *Journal of the American Statistical Association*, 94(445):247–253, 1999.

Andrew Gelman, Walter R Gilks, and Gareth O Roberts. Weak convergence and optimal scaling of random walk Metropolis algorithms. *The Annals of Applied Probability*, 7(1):110–120, 1997.

Andrew Gelman, Jessica Hwang, and Aki Vehtari. Understanding predictive information criteria for Bayesian models. *Statistics and Computing*, 24(6):997–1016, 2014.

Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.

Aki Vehtari. Comparison of MCMC effective sample size estimators. *Github, https://avehtari.github.io/rhat_ess/ess_comparison.html*, 2021.

Sigrunn Holbek Sørbye and Håvard Rue. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8:39–51, 2014.

Maria-Zafeiria Spyropoulou and James Bentham. Scaling priors in two dimensions for intrinsic Gaussian Markov random fields. *Statistica Neerlandica, to be submitted (preprint: https://arxiv.org/abs/2111.09003)*, 2022a.

Finn Lindgren and Håvard Rue. On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, 35(4):691–700, 2008.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

Maria-Zafeiria Spyropoulou and James Bentham. A two-dimensional intrinsic Gaussian Markov random field for blood pressure data. *The Annals of Applied Statistics under review (preprint: https://arxiv.org/abs/2111.07848)*, 2022b.

Sara Martino, Rupali Akerkar, and Håvard Rue. Approximate Bayesian inference for survival models. *Scandinavian Journal of Statistics*, 38(3):514–528, 2011.

Sigrunn H Sørbye and Håvard Rue. Simultaneous credible bands for latent Gaussian models. *Scandinavian Journal of Statistics*, 38(4):712–725, 2011.

Maria-Zafeiria Spyropoulou. Thesis code, 2022. Available at: https://github.com/mariza22 (Accessed: January 28, 2023).

Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G Martins, Sigrunn H Sørbye, et al. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28, 2017.

Appendix A: Paper under review by Annals of Applied Statistics

Submitted to the Annals of Applied Statistics

A FLEXIBLE TWO-DIMENSIONAL BAYESIAN HIERARCHICAL MODEL FOR BLOOD PRESSURE DATA

BY MARIA-ZAFEIRIA SPYROPOULOU[1,a], JAMES BENTHAM[2,b]

[1] School of Sport and Exercise Science, University of Kent, UK, [a]mzs2@kent.ac.uk

[2] School of Mathematics, Statistics and Actuarial Science, University of Kent, UK, [b]j.bentham@kent.ac.uk

Many phenomena are naturally bivariate, including blood pressure, which has systolic and diastolic levels. Here, we develop and fit a Bayesian hierarchical model that estimates trends in these values and their interaction simultaneously, using sparse data that vary substantially between countries, over time and by age and sex. A key element of the model is a two-dimensional second-order intrinsic Gaussian Markov random field, which captures non-linear trends in the variables and their interaction. The model is fitted using Markov chain Monte Carlo methods, with a block Metropolis-Hastings algorithm providing efficient updates. Performance is demonstrated using simulated and real data.

1. Introduction. Blood pressure is bivariate: systolic blood pressure (SBP) is a measurement of the force exerted by the heart when it beats, and diastolic blood pressure (DBP) is a measurement of resistance to flow in the blood vessels. Raised blood pressure (RBP), defined as SBP $\geq$ 140 mmHg or DBP $\geq$ 90 mmHg, is estimated to affect more than one billion people worldwide (NCD Risk Factor Collaboration (2017*a*)), and is a key risk factor for cardiovascular diseases, cancers and diabetes, which are responsible for approximately 70% of global deaths each year (World Health Organization (2018)).

In the past decade, Bayesian hierarchical modelling has become an established method in global health research, and has been used to make national-level estimates for variables including blood pressure, body-mass index, and diabetes (NCD Risk Factor Collaboration (2017*a*), NCD Risk Factor Collaboration (2017*b*), NCD Risk Factor Collaboration (2016)). The models are sufficiently robust for the World Health Organization (WHO) to report these estimates in the Global Diabetes Report (World Health Organization (2016)), and for the results to be included in the Global Atlas on Childhood Obesity (World Obesity Federation (2019)) and the Global Nutrition Report (Development Initiatives (2020)). However, while influential, this work has allowed estimation only of single variables at a time. Given that interactions between disease risk factors are complex and vary over time, by age and sex, and between countries, these models have not been able to capture important information.

In particular, although estimates of national-level trends in SBP, DBP and RBP are available (NCD Risk Factor Collaboration (2017*a*)), there is no detailed understanding of how the interaction between SBP and DBP varies over time, and by country, age and sex. Here, we extend existing methodology (Danaei *and others* (2011), Finucane *and others* (2014)) to the two-dimensional case, specifying and fitting a Bayesian hierarchical model that allows SBP, DBP and their interaction to be estimated simultaneously. A key development is to extend the random walk (Rue and Held (2005)), which modelled non-linear trends in individual variables, to two dimensions, using an intrinsic Gaussian Markov random field or IGMRF (Rue and Held (2005), Yue and Speckman (2010), Thon *and others* (2012)) as the prior distribution for the precision matrix of a two-dimensional random walk.

Keywords and phrases: Bayesian hierarchical modelling, block Metropolis-Hastings sampling, canonical parametrisation, health data, intrinsic Gaussian Markov random fields, Markov chain Monte Carlo methods.

The paper is organised as follows: Section 2 introduces the model, Section 3 describes the two-dimensional IGMRF, and Section 4 presents model implementation. Section 5 provides results from a simulation study, Section 6 describes an application to real data, and Section 7 summarises our findings and proposes future work.

2. Model and Assumptions. Data for mean DBP, mean SBP, and their interaction, the mean of the product of the DBP and SBP measurements, are available from study i carried out in country j for age group h. The model has a hierarchy with four levels, where countries are nested in regions, super-regions and the world as a whole. We model these data, $\mathbf{y}$, as shown in (1), where the D, S and I subscripts refer to DBP, SBP and their interaction, respectively, and $j[i]$ refers to the country j in which study i was carried out.

$$\begin{bmatrix} y_{h,i,D} \\ y_{h,i,S} \\ y_{h,i,I} \end{bmatrix} \sim \mathcal{N}\Bigg(\begin{bmatrix} a_{j[i],D} \\ a_{j[i],S} \\ a_{j[i],I} \end{bmatrix} + \begin{bmatrix} b_{j[i],D}t_i \\ b_{j[i],S}t_i \\ b_{j[i],I}t_i \end{bmatrix} + \begin{bmatrix} \mathbf{u}_{j[i],(D,S),t_i} \end{bmatrix} + \begin{bmatrix} X_{i,D}.\boldsymbol{\beta} \\ X_{i,S}.\boldsymbol{\beta} \\ X_{i,I}.\boldsymbol{\beta} \end{bmatrix}$$

$$(1) \qquad + \begin{bmatrix} \gamma_{i,D}(z_h) \\ \gamma_{i,S}(z_h) \\ \gamma_{i,I}(z_h) \end{bmatrix} + \begin{bmatrix} e_{i,D} \\ e_{i,S} \\ e_{i,I} \end{bmatrix}, \text{diag} \begin{bmatrix} SD^2_{D,h,i}/n_{h,i} + \tau^2_{i,D} \\ SD^2_{S,h,i}/n_{h,i} + \tau^2_{i,S} \\ SD^2_{I,h,i}/n_{h,i} + \tau^2_{i,I} \end{bmatrix} \Bigg)$$

This model has the same components as the earlier one-dimensional model, which fitted the complex patterns observed in SBP and DBP data well (Danaei *and others* (2011), Finucane *and others* (2014), NCD Risk Factor Collaboration (2017*a*)), but extends them to the two-dimensional case. It includes linear intercepts and slopes, $\mathbf{a}_{j[i]}$ and $\mathbf{b}_{j[i]}$, non-linear terms that vary over time, $\mathbf{u}_{j[i]}$, covariate effects, $\boldsymbol{\beta}$, terms in age, $\boldsymbol{\gamma}_i(z_h)$, study-specific random effects, $\mathbf{e}_i$, and noise, $\boldsymbol{\epsilon}_{h,i}$, assumed to be iid Gaussian. More succinctly, the model can be expressed as

$$(2) \qquad \mathbf{y}_{h,i} \sim \mathcal{N}(\mathbf{a}_{j[i]} + \mathbf{b}_{j[i]}t_i + \mathbf{u}_{j[i],t_i} + \mathbf{X}_i\boldsymbol{\beta} + \boldsymbol{\gamma}_i(z_h) + \mathbf{e}_i, \mathbf{SD}^2_{h,i}/n_{h,i} + \boldsymbol{\tau}^2_i)$$

and to simplify the MCMC implementation, we rewrite the likelihood as shown in (3) (Danaei *and others* (2011), Finucane *and others* (2014)):

$$(3) \qquad \mathbf{y} \sim \mathcal{N}(\mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u} + \boldsymbol{\gamma}(z_h), \boldsymbol{\Sigma})$$

The vector $\boldsymbol{\theta}$ contains the $\mathbf{a}_j$, $\mathbf{b}_j$, $\boldsymbol{\beta}$ and $\mathbf{e}_i$ terms, $\mathbf{F}$ is a mapping matrix that contains the covariates $\mathbf{X}$ and that assigns each $\mathbf{a}_j$, $\mathbf{b}_j$, $\boldsymbol{\beta}$ and $\mathbf{e}_i$ to their corresponding studies, and $\boldsymbol{\Sigma}$ denotes the diagonal matrix of $\mathbf{SD}^2/n + \boldsymbol{\tau}^2$ values which express the residual variance within and between age groups respectively in each study i.

It can be seen in (1) and (2) that there is a single non-linear sub-model $\boldsymbol{u}_{j[i],t_i}$, which provides simultaneous estimates of non-linear trends in SBP and DBP, and implicit estimates of trends in their interaction, using an IGMRF as described below. The non-linear term for country j, $\mathbf{u}_{j[i]}$, can be represented as

$$\boldsymbol{u}_j = \boldsymbol{u}^c_j + \boldsymbol{u}^r_{k[j]} + \boldsymbol{u}^s_{l[j]} + \boldsymbol{u}^g$$

and in matrix form, the non-linear sub-model for all country-year combinations is

$$\boldsymbol{M}\boldsymbol{u} = \mathbf{M}^c\boldsymbol{u}^c + \mathbf{M}^r\boldsymbol{u}^r + \mathbf{M}^s\boldsymbol{u}^s + \mathbf{M}^g\boldsymbol{u}^g$$

Here, $\boldsymbol{M}^c$ is a translation matrix of dimensionality $I \times J \times T^2$ that assigns the elements of $\mathbf{u}^c$ to I studies observed in $J \times T^2$ country-year combinations, while the $\mathbf{M}^r$, $\mathbf{M}^s$ and $\mathbf{M}^g$ matrices correspond to region-year and super-region-year combinations, and years for the globe.

The age sub-model is defined as

$$\gamma(z_h) = \mathbf{Z}\boldsymbol{\psi} + (\mathbf{F}\boldsymbol{\theta} + \mathbf{M}\mathbf{u}) \odot (\mathbf{Z}\boldsymbol{\phi}) + (\mathbf{C}\mathbf{c} \odot \mathbf{Z})1_5$$

where $\mathbf{Z} = (N_{0,3}(t), N_{1,3}(t), \ldots, N_{4,3}(t))'$ expresses the age data using cubic B-splines, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_5)'$ are constant terms, and $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_5)'$ are mean terms. The $\mathbf{C}$ term is an $I \times J$ translation matrix assigning each of the I studies to the J countries, while $\mathbf{c}$ is a $J \times 5$ matrix expressing country-specific random effects for each of the five terms in the cubic B-splines. The $\odot$ symbol denotes component-wise matrix-vector multiplication.

3. Intrinsic Gaussian Markov Random Fields. A random vector $\mathbf{x} = (x_1, ..., x_n)^T \in \mathcal{R}^n$ is called a Gaussian Markov random field (GMRF) linked to a labelled graph $G = (V, E)$ with mean $\boldsymbol{\mu}$ and precision matrix $\mathbf{Q} > 0$ iff its density has the form

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

and

$$Q_{ij} \neq 0 \Leftrightarrow \{i, j\} \in E, \quad \forall i \neq j \tag{4}$$

A GMRF is any Gaussian distribution with a symmetric positive definite covariance matrix and vice versa (Rue and Held (2005)). The properties of a GMRF are described by its precision matrix $\mathbf{Q}$ since its structure indicates connections between nodes; as shown in (4), non-zero values in $\mathbf{Q}$ correspond to an edge in G. Where $\mathbf{Q}$ is dense, the graph is fully connected, while sparsity implies conditional independence.

IGMRFs are improper GMRFs, as they have a precision matrix that is not of full rank, and since the precision matrix is not of full rank, its inverse does not exist. This implies that IGMRFs do not have well-defined mean or covariance matrices. However, they do have the property that the mean of an IGMRF of order T is defined up to the addition of a polynomial of order $T-1$ (Rue and Held (2005)).

3.1. *One-dimensional IGMRFs.* One-dimensional IGMRFs are described in detail elsewhere (Rue and Held (2005)). For a first-order random walk, the rank of the precision matrix is $T-1$, and the behaviour in the first and last rows is different from the rest of the matrix, as shown in (5). These are the boundary constraints, and in this case, imposing the constraint $\sum_i x_i = 0$ on the mean of the intercepts gives a proper joint density (Sørbye and Rue (2014)). For a second-order IGMRF, the rank of the precision matrix is $T-2$, and the behaviour in the first two and last two rows is different from the rest of the matrix, as shown in (6). There are now two constraints, $\sum_i x_i = 0$ and $\sum_i i x_i = 0$, corresponding to the means of the intercepts and slopes.

$$\mathbf{Q}_T = \lambda \begin{pmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & -1 & 2 & -1 \\ & & & -1 & 1 \end{pmatrix} \tag{5}$$

$$\mathbf{Q}_T = \lambda \begin{pmatrix} 1 & -2 & 1 & & \\ -2 & 5 & -4 & 1 & \\ 1 & -4 & 6 & -4 & 1 \\ & 1 & -4 & 5 & -2 \\ & & 1 & -2 & 1 \end{pmatrix} \tag{6}$$

3.2. *Two-dimensional IGMRFs.* To estimate non-linear trends in DBP, SBP and their interaction simultaneously, we use a two-dimensional IGMRF with precision matrix $\mathbf{Q}_i$ as a prior distribution at each level of the hierarchy. If the values $\mathbf{u}_i$, $i : \{c, r, s, g\}$, represent non-linearity at country, region, super-region and global level, respectively, we have

$$\pi(\mathbf{u}_c) \propto \exp\left(-\frac{1}{2}\mathbf{u}_c^T \mathbf{Q_c} \mathbf{u}_c\right)$$

and similar results for $\mathbf{u}_r$, $\mathbf{u}_s$ and $\mathbf{u}_g$. The posterior density of $\mathbf{u}_c$ is:

$$\log(\text{posterior}) \propto \log(\text{likelihood}) + \log(\text{prior}) + \log(\text{hyperprior})$$

$$\log P(\mathbf{u}_c|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\gamma}) \propto \log P(\mathbf{y}|\mathbf{u}_c, \lambda_c, \boldsymbol{\theta}, \boldsymbol{\gamma}) + \log P(\mathbf{u}_c|\lambda_c, \boldsymbol{\theta}, \boldsymbol{\gamma}) + \log P(\log \lambda_c) \tag{7}$$

$$\log P(\mathbf{u}_c|\mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\gamma}) \propto \log P(\mathbf{y}|\mathbf{u}_c, \boldsymbol{\theta}, \boldsymbol{\gamma}) + \log P(\mathbf{u}_c|\lambda_c) + \log P(\log \lambda_c)$$

The $\mathbf{u}_r$, $\mathbf{u}_s$ and $\mathbf{u}_g$ terms display the same behaviour as (7), with precision parameters λ_r, λ_s, and λ_g in analogous form. In two dimensions, the behaviour of $\mathbf{Q}$ is more complex than in the one-dimensional case, as its entries are now block-circulant matrices rather than single numbers (Thon *and others* (2012)). Moreover, although the order of an IGMRF can be expressed as the rank deficiency of its precision matrix (Rue and Held (2005)), we now have two variables and so we need to impose three constraints, $\sum x_{ij} = 0$, $\sum i x_{ij} = 0$ and $\sum j x_{ij} = 0$, to give a proper joint distribution and a finite marginal distribution. These constraints correspond to the mean of the intercepts and the two means of the linear trends. Each block matrix inside the precision matrix is $T \times T$ dimensional and therefore the precision matrix is $T^2 \times T^2$ dimensional.

3.2.1. *Constructing the Precision Matrix.* The precision matrix $\mathbf{Q} = \lambda \times \mathbf{C}^T\mathbf{C}$, where $\mathbf{C}$ is a block-circulant matrix, λ is a precision parameter and we have a so-called structure matrix, $\mathbf{P} = \mathbf{C}^T\mathbf{C}$. We consider the case of variables i and j, and set $T = 5$ as an example to illustrate general behaviour. The sub-matrices used to construct $\mathbf{C}$ are:

$$\mathbf{A}_1 = \begin{pmatrix} -4 & 1 & & & 1 \\ 1 & -4 & 1 & & \\ & 1 & -4 & 1 & \\ & & 1 & -4 & 1 \\ 1 & & & 1 & -4 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} 1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & 1 & \\ & & & & 1 \end{pmatrix}$$

with

$$\mathbf{C} = \begin{pmatrix} \mathbf{A_1} & \mathbf{A_2} & \mathbf{0} & \mathbf{0} & \mathbf{A_2} \\ \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_2} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_2} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A_2} & \mathbf{A_1} & \mathbf{A_2} \\ \mathbf{A_2} & \mathbf{0} & \mathbf{0} & \mathbf{A_2} & \mathbf{A_1} \end{pmatrix}$$

Here $\mathbf{0}$ represents $T \times T$ zeroes, with the full matrices shown in Figures 1 and 2 respectively.

The construction of $\mathbf{C}$ is based on forward differences for each variable, with increments:

$$\Delta^2_{(1,0)} f(D_i, S_j) + \Delta^2_{(0,1)} f(D_i, S_j)$$

Here $\Delta_{(1,0)}$ and $\Delta_{(0,1)}$ are the first-order forward difference operators, which can be written as:

$$\Delta_{(1,0)} f(D_i, S_j) = f(D_i, S_j) - f(D_{i-1}, S_j)$$
$$\Delta_{(0,1)} f(D_i, S_j) = f(D_i, S_j) - f(D_i, S_{j-1})$$

1.1	-4	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.
1.2	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.
1.3	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.
1.4	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.
1.5	1	.	.	1	-4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1
2.1	1	.	.	.	.	-4	1	.	.	1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
2.2	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
2.3	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
2.4	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
2.5	.	.	.	.	1	1	.	.	1	-4	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
3.1	.	.	.	.	.	1	.	.	.	.	-4	1	.	.	1	1	.	.	.	.	.	.	.	.	.
3.2	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.	.
3.3	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.	.
3.4	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.	.	.	.
3.5	.	.	.	.	.	.	.	.	.	1	1	.	.	1	-4	.	.	.	.	1	.	.	.	.	.
4.1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-4	1	.	.	1	1	.	.	.	.
4.2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.	.
4.3	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.	.
4.4	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.	.	1	.
4.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	-4	.	.	.	.	1
5.1	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-4	1	.	.	1
5.2	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.	.
5.3	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1	.
5.4	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	1	-4	1
5.5	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	1	.	.	1	-4

FIG 1. *Matrix* $\mathbf{C}$ *for 5 years of SBP and DBP combinations*

1.1	20	-8	1	1	-8	-8	2	.	.	2	1	.	.	.	.	1	.	.	.	.	-8	2	.	.	2
1.2	-8	20	-8	1	1	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	.
1.3	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.
1.4	1	1	-8	20	-8	.	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2
1.5	-8	1	1	-8	20	2	.	.	2	-8	.	.	.	.	1	.	.	.	.	1	2	.	.	2	-8
2.1	-8	2	.	.	2	20	-8	1	1	-8	-8	2	.	.	2	1	.	.	.	.	1	.	.	.	.
2.2	2	-8	2	.	.	-8	20	-8	1	1	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.
2.3	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.
2.4	.	.	2	-8	2	1	1	-8	20	-8	.	.	2	-8	2	.	.	.	1	.	.	.	.	1	.
2.5	2	.	.	2	-8	-8	1	1	-8	20	2	.	.	2	-8	.	.	.	.	1	.	.	.	.	1
3.1	1	.	.	.	.	-8	2	.	.	2	20	-8	1	1	-8	-8	2	.	.	2	1	.	.	.	.
3.2	.	1	.	.	.	2	-8	2	.	.	-8	20	-8	1	1	2	-8	2	.	.	.	1	.	.	.
3.3	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.
3.4	.	.	.	1	.	.	.	2	-8	2	1	1	-8	20	-8	.	.	2	-8	2	.	.	.	1	.
3.5	.	.	.	.	1	2	.	.	2	-8	-8	1	1	-8	20	2	.	.	2	-8	.	.	.	.	1
4.1	1	.	.	.	.	1	.	.	.	.	-8	2	.	.	2	20	-8	1	1	-8	-8	2	.	.	2
4.2	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	.	-8	20	-8	1	1	2	-8	2	.	.
4.3	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.
4.4	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	1	1	-8	20	-8	.	.	2	-8	2
4.5	.	.	.	.	1	.	.	.	.	1	2	.	.	2	-8	-8	1	1	-8	20	2	.	.	2	-8
5.1	-8	2	.	.	2	1	.	.	.	.	1	.	.	.	.	-8	2	.	.	2	20	-8	1	1	-8
5.2	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	.	-8	20	-8	1	1
5.3	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1
5.4	.	.	2	-8	2	.	.	.	1	.	.	.	.	1	.	.	.	2	-8	2	1	1	-8	20	-8
5.5	2	.	.	2	-8	.	.	.	.	1	.	.	.	.	1	2	.	.	2	-8	-8	1	1	-8	20

FIG 2. *Structure matrix* $\mathbf{P} = \mathbf{C}^T\mathbf{C}$ *for 5 years of SBP and DBP combinations*

Similarly, the second-order forward differences for each variable are defined as:

$$\Delta^2_{(1,0)} f(D_i, S_j) = \Delta_{(1,0)}(\Delta_{(1,0)} f(D_i, S_j)) = f(D_i, S_j) - 2f(D_{i-1}, S_j) + f(D_{i-2}, S_j)$$

$$\Delta^2_{(0,1)} f(D_i, S_j) = \Delta_{(0,1)}(\Delta_{(0,1)} f(D_i, S_j)) = f(D_i, S_j) - 2f(D_i, S_{j-1}) + f(D_i, S_{j-2})$$

To construct the structure matrix, we require the square of these differences since $\mathbf{P} = \mathbf{C}^T\mathbf{C}$, and so each row of $\mathbf{P}$ will have increments:

$$\begin{aligned} -(\Delta^2_{(1,0)} f(D_i, S_j) + \Delta^2_{(0,1)} f(D_i, S_j))^2 = \\ -(\Delta^4_{(1,0)} f(D_i, S_j) + \Delta^4_{(0,1)} f(D_i, S_j) + \Delta^2_{(0,1)}\Delta^2_{(1,0)} f(D_i, S_j)) \end{aligned}$$

Applying these forward differences, we have the following entries for each row:

$$\begin{aligned} f(D_{i-2}, S_j) + 2f(D_{i-1}, S_{j-1}) - 8f(D_{i-1}, S_j) + 2f(D_{i-1}, S_{j+1}) + f(D_i, S_{j-2}) \\ - 8f(D_i, S_{j-1}) + 20f(D_i, S_j) - 8f(D_i, S_{j+1}) + f(D_i, S_{j+2}) \\ + 2f(D_{i+1}, S_{j-1}) - 8f(D_{i+1}, S_j) + 2f(D_{i+1}, S_{j+1}) + f(D_{i+2}, S_j) \end{aligned}$$

This leads to a structure matrix with a torus form (Thon *and others* (2012)). This is not appropriate for our analyses given that time is linear, and so we apply boundary constraints.

3.2.2. *Applying Boundary Constraints.* A structure matrix $\mathbf{P}$ with boundary constraints is available (Yue and Speckman (2010)), and this specification has been used in our two-dimensional blood pressure model. Here, we define the block matrices of $\mathbf{P}$ such that each row and column sums to zero, taking into account only data in neighbouring years. We use sub-matrices:

$$\mathbf{A}_1 = \begin{pmatrix} 6 & -5 & 1 & & \\ -5 & 12 & -6 & 1 & \\ 1 & -6 & 12 & -6 & 1 \\ & 1 & -6 & 12 & -5 \\ & & 1 & -5 & 6 \end{pmatrix}, \quad \mathbf{A}_2 = \begin{pmatrix} -5 & 2 & & & \\ 2 & -7 & 2 & & \\ & 2 & -7 & 2 & \\ & & 2 & -7 & 2 \\ & & & 2 & -5 \end{pmatrix}$$

$$\mathbf{A_3} = \mathrm{diag}(1,1,1,1,1)$$

$$\mathbf{A}_4 = \begin{pmatrix} 12 & -7 & 1 & & \\ -7 & 20 & -8 & 1 & \\ 1 & -8 & 20 & -8 & 1 \\ & 1 & -8 & 20 & -7 \\ & & 1 & -7 & 12 \end{pmatrix}, \quad \mathbf{A}_5 = \begin{pmatrix} -6 & 2 & & & \\ 2 & -8 & 2 & & \\ & 2 & -8 & 2 & \\ & & 2 & -8 & 2 \\ & & & 2 & -6 \end{pmatrix}$$

They are used to construct a precision matrix:

$$\lambda\mathbf{P} = \lambda \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \mathbf{A}_3 & & \\ \mathbf{A}_2 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 & \\ \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_5 & \mathbf{A}_3 \\ & \mathbf{A}_3 & \mathbf{A}_5 & \mathbf{A}_4 & \mathbf{A}_2 \\ & & \mathbf{A}_3 & \mathbf{A}_2 & \mathbf{A}_1 \end{pmatrix} \tag{8}$$

This matrix has both overall boundary constraints and constraints within the block matrices, with ranks of $T^2 - 3$ as described above, and so $\mathbf{Q}$ has three eigenvalues with values of zero (Paciorek (2009)). The matrix is shown in detail in Figure 3, where for the case $T = 5$, only the 13^{th} row has no boundary constraints. The diagonal elements represent the conditional precision of DBP and SBP in a particular year i, while scaled off-diagonal elements describe the conditional correlation between DBP and SBP in years i and j (Rue and Held (2005)). These relationships are shown in detail in Figure 4, where the number of neighbours depends on the combinations of years for which DBP and SBP are being estimated; for example, when the DBP and SBP variables are at $(i, j) = (1, 1)$, only information from the years ahead is available. For a two-dimensional second-order IGMRF, neighbours are up to two steps in each direction (i.e., forward or backward either vertically or horizontally in Figure 4), where moving one step diagonally comprises two steps, one vertical and one horizontal, and the numbers of possible neighbours are 6, 8, 9, 11 or 12 at the boundaries and 13 elsewhere. The differences calculated taking into account this behaviour at the boundaries are (Yue and Speckman (2010)):

1.1	6	-5	1	.	.	-5	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.
1.2	-5	12	-6	1	.	2	-7	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.
1.3	1	-6	12	-6	1	.	2	-7	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.
1.4	.	1	-6	12	-5	.	.	2	-7	2	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.
1.5	.	.	1	-5	6	.	.	.	2	-5	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.
2.1	-5	2	.	.	.	12	-7	1	.	.	-6	2	.	.	.	1	.	.	.	.	.	.	.	.	.
2.2	2	-7	2	.	.	-7	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	.	.	.	.
2.3	.	2	-7	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.	.	.	.	.
2.4	.	.	2	-7	2	.	1	-8	20	-7	.	.	2	-8	2	.	.	.	1	.	.	.	.	.	.
2.5	.	.	.	2	-5	.	.	1	-7	12	.	.	.	2	-6	.	.	.	.	1	.	.	.	.	.
3.1	1	.	.	.	.	-6	2	.	.	.	12	-7	1	.	.	-6	2	.	.	.	1	.	.	.	.
3.2	.	1	.	.	.	2	-8	2	.	.	-7	20	-8	1	.	2	-8	2	.	.	.	1	.	.	.
3.3	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-8	2	.	.	.	1	.	.
3.4	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-7	.	.	2	-8	2	.	.	.	1	.
3.5	.	.	.	.	1	.	.	.	2	-6	.	.	1	-7	12	.	.	.	2	-6	.	.	.	.	1
4.1	.	.	.	.	.	1	.	.	.	.	-6	2	.	.	.	12	-7	1	.	.	-5	2	.	.	.
4.2	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	.	-7	20	-8	1	.	2	-7	2	.	.
4.3	.	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-8	1	.	2	-7	2	.
4.4	.	.	.	.	.	.	.	.	1	.	.	.	2	-8	2	.	1	-8	20	-7	.	.	2	-7	2
4.5	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-6	.	.	1	-7	12	.	.	.	2	-5
5.1	.	.	.	.	.	.	.	.	.	.	1	.	.	.	.	-5	2	.	.	.	6	-5	1	.	.
5.2	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-7	2	.	.	-5	12	-6	1	.
5.3	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-7	2	.	1	-6	12	-6	1
5.4	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-7	2	.	1	-6	12	-5
5.5	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	.	.	2	-5	.	.	1	-5	6

FIG 3. *Structure matrix for 5 years of SBP and DBP combinations,* $\boldsymbol{P} = \boldsymbol{C}^T \boldsymbol{C}$

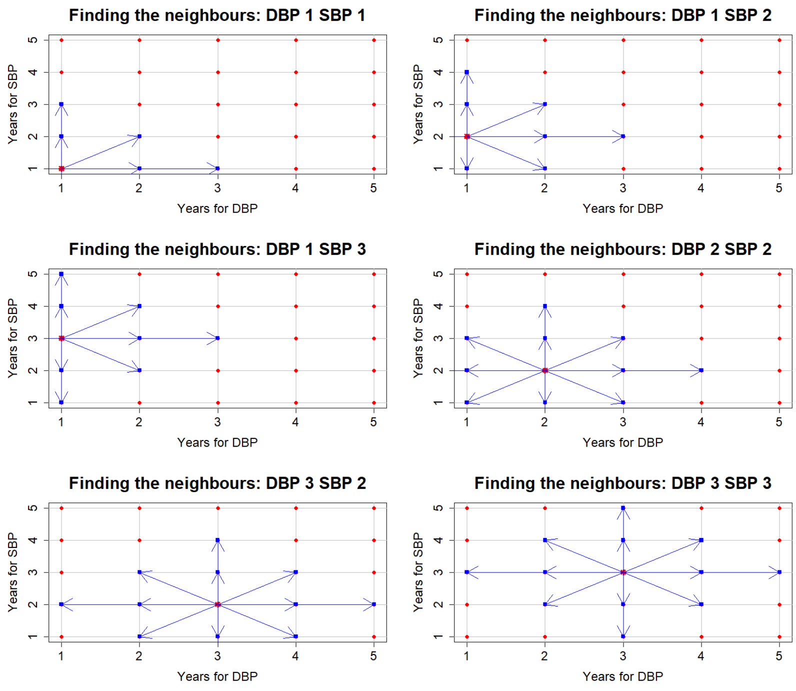

FIG 4. *Neighbouring years for DBP and SBP for the case* $T = 5$

$$\sum_{i=2}^{T-1}\sum_{j=2}^{T-1}\{\Delta_0^2 f(D_i, S_j)\}^2 + \{\Delta_1 f(D_1, S_1)\}^2 + \{\Delta_2 f(D_T, S_1)\}^2$$

$$+ \{\Delta_3 f(D_1, S_T)\}^2 + \{\Delta_4 f(D_T, S_T)\}^2 + \sum_{i=2}^{T}(\{\Delta_5 f(D_i, S_1)\}^2 + \{\Delta_6 f(D_i, S_T)\}^2)$$

$$+ \sum_{j=2}^{T}(\{\Delta_7 f(D_1, S_j)\}^2 + \{\Delta_8 f(D_T, S_j)\}^2)$$

Here, each of the forward differences can be written as:

$$\Delta_0^2 f(D_i,S_j) = (\Delta^2_{(1,0)} + \Delta^2_{(0,1)})f(D_i,S_j) = \Delta^2_{(1,0)} f(D_i,S_j) + \Delta^2_{(0,1)} f(D_i,S_j)$$

$$\{\Delta_0^2 f(D_i,S_j)\}^2 = (\Delta^2_{(1,0)} + \Delta^2_{(0,1)})^2 f(D_i,S_j) = (\Delta^4_{(1,0)} + \Delta^4_{(0,1)} + 2\Delta^2_{(1,0)}\Delta^2_{(0,1)})f(D_i,S_j)$$

and we have:

$$\begin{aligned}
\{\Delta_1 f(D_1,S_1)\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 f(D_1,S_1)\\
\{\Delta_2 f(D_T,S_1)\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 f(D_T,S_1)\\
\{\Delta_3 f(D_1,S_T)\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 f(D_1,S_T)\\
\{\Delta_4 f(D_T,S_T)\}^2 &= (\Delta_{(1,0)} + \Delta_{(0,1)})^2 f(D_T,S_T)\\
\{\Delta_5 f(D_i,S_1)\}^2 &= (\Delta^2_{(1,0)} + \Delta_{(0,1)})^2 f(D_i,S_1)\\
\{\Delta_6 f(D_i,S_T)\}^2 &= (\Delta^2_{(1,0)} + \Delta_{(0,1)})^2 f(D_i,S_T)\\
\{\Delta_7 f(D_1,S_j)\}^2 &= (\Delta_{(1,0)} + \Delta^2_{(0,1)})^2 f(D_1,S_j)\\
\{\Delta_8 f(D_T,S_j)\}^2 &= (\Delta_{(1,0)} + \Delta^2_{(0,1)})^2 f(D_T,S_j)
\end{aligned}$$

When $T = 5$, we have 25 combinations of differences, each of which has different behaviour depending on its position. For example, in (8), the middle row of the precision matrix is:

$$\begin{aligned}
f(D_{i-2},S_j) &+ 2f(D_{i-1},S_{j-1}) - 8f(D_{i-1},S_j) + 2f(D_{i-1},S_{j+1}) + f(D_i,S_{j-2})\\
&- 8f(D_i,S_{j-1}) + 20f(D_i,S_j) - 8f(D_i,S_{j+1}) + f(D_i,S_{j+2})\\
&+ 2f(D_{i+1},S_{j-1}) - 8f(D_{i+1},S_j) + 2f(D_{i+1},S_{j+1}) + f(D_{i+2},S_j)
\end{aligned}$$

4. Implementation using Metropolis-within-Gibbs Sampling.

4.1. *Block-sampling Metropolis-Hastings algorithm.* For complex Bayesian hierarchical models, block updating can be used to avoid slow convergence and poor mixing caused by strong dependence between parameters and hyperparameters (Rue and Held (2005)). For our model, there is particularly large dependence between the parameters and hyperparameters of the non-linear model and so block updating is used in those updates.

The prior distributions of the non-linear parameters are:

$$\mathbf{u}^c_j \sim \mathcal{N}(\mathbf{0}, \lambda_c \mathbf{P}), \quad \mathbf{u}^r_{k[j]} \sim \mathcal{N}(\mathbf{0}, \lambda_r \mathbf{P}), \quad \mathbf{u}^s_{l[j]} \sim \mathcal{N}(\mathbf{0}, \lambda_s \mathbf{P}), \quad \mathbf{u}^g \sim \mathcal{N}(\mathbf{0}, \lambda_g \mathbf{P})$$

Here, $\mathbf{0}$ is a vector of length T^2, $\mathbf{P}$ is a $T^2 \times T^2$ matrix, and each of the non-linear terms is a vector that comprises all possible combinations of years for DBP and SBP; for instance, for the country-level non-linear term, we have a vector of length T^2, $\mathbf{u}^c_j = (u_{11}, u_{12}, \ldots, u_{32}, u_{34}, u_{35}, \ldots, u_{TT})$, for each country. However, while the non-linear model allows estimation for T^2 combinations, information that informs the likelihood is only available when data are present, and this is only the case when both DBP and SBP are in the same year since measurements are taken for individuals simultaneously.

Here, we describe the proposal and posterior distributions at the country level of the hierarchy, but the distributions for the other levels are analogous. The posterior distribution is the joint distribution of the parameter and hyperparameter and is defined as:

$$\begin{aligned}
\log(P(\mathbf{u}^c, \lambda_c | \mathbf{y}, \lambda_c, \boldsymbol{\theta}, \gamma)) \propto\ & (\mathbf{y} - \mathbf{F}(\mathbf{Z}\boldsymbol{\phi} + 1)\boldsymbol{\theta} - \mathbf{M}_c(\mathbf{Z}\boldsymbol{\phi} + 1)\mathbf{u}^c - \mathbf{Z}\boldsymbol{\psi} - (\mathbf{C}_c\mathbf{Z})1_5)^T \boldsymbol{\Sigma}^{-\mathbf{1}}\\
& (\mathbf{y} - \mathbf{F}(\mathbf{Z}\boldsymbol{\phi} + 1)\boldsymbol{\theta} - \mathbf{M}_c(\mathbf{Z}\boldsymbol{\phi} + 1)\mathbf{u}^c - \mathbf{Z}\boldsymbol{\psi} - (\mathbf{C}_c\mathbf{Z})1_5)\\
& + \log(P(\mathbf{u}^c|\lambda_c)) + \log P(\log \lambda_c)
\end{aligned}$$

The proposal distribution depends solely on $\mathbf{u}^c$, hence we derive the proposal distribution for $\mathbf{u}^c$ using the canonical parametrisation:

$$\mathbf{u}^c|\boldsymbol{\mu}_{u^c}, \mathbf{Q}_{u^c} \sim \mathcal{N}(\mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c}, \mathbf{Q}_{u^c})$$

where:

$$\mathbf{Q}_{u^c} = (\mathbf{M}_c * (\mathbf{Z}\phi + 1))^T \boldsymbol{\Sigma}^{-1}(\mathbf{M}_c * (\mathbf{Z}\phi + 1)) + I_j \otimes \lambda_c \mathbf{P}$$

$$\mathbf{Q}_{u^c}\boldsymbol{\mu}_{u^c} = (\mathbf{M}_c * (\mathbf{Z}\phi + 1))^T \boldsymbol{\Sigma}^{-1}[\mathbf{y} - \mathbf{F}(\mathbf{Z}\phi + 1)\boldsymbol{\theta} - (\mathbf{Z}\psi - ((\mathbf{C}c)\mathbf{Z})1_5)]$$

4.2. *Defining constraints.* To construct the proposal and define the parameters, we need to apply appropriate constraints. For the hyperprior, we propose $\log \lambda_c^*$ from the $\mathcal{N}(\log \lambda_c^{i-1}, \omega^2_{\log \lambda_c})$ distribution, while $\mathbf{Q}_{u^c}$ is a block diagonal matrix describing full conditional within-country correlations. Each block corresponds to a country and describes the correlations that exist between data at different time points. Therefore, we have J blocks corresponding to the countries, each containing T^2 year combinations. For the block corresponding to a country j, there are four possible scenarios, where the country has:

1. *No data*: The likelihood makes no contribution, hence $\mathbf{Q}_{u^c}$ depends solely on the prior, so we propose $\mathbf{u}_j^{c*} \sim \mathcal{N}(0, (\lambda_c \mathbf{P})^{-1})$. The rank of $\mathbf{P}$ is $T^2 - 3$, corresponding to infinite prior variance on the intercept and the two linear trends of $\mathbf{u}_j^c$. We constrain these three linear combinations of $\mathbf{u}_j^c$ to zero by taking the generalised inverse of $\mathbf{P}$, setting the last three eigenvalues of $\mathbf{P}$ to infinity.
2. *One year of data*: $\mathbf{Q}_{u^c}$ has rank $T^2 - 2$ because the intercept is defined by the data but the linear trends are not. We constrain the linear trends of $\mathbf{u}_j^c$ to zero by taking the generalised inverse of $\mathbf{Q}_{u^c}$, setting the last two eigenvalues to infinity. For identifiability, we constrain the mean of the intercept to zero.
3. *Two years of data*: $\mathbf{Q}_{u^c}$ has rank $T^2 - 1$ because the mean and one linear trend of $\mathbf{u}_j^c$ are defined by the data. We constrain the other linear trend of $\mathbf{u}_j^c$ to zero by taking the generalised inverse of $\mathbf{Q}_{u^c}$, setting the last eigenvalue to infinity. For identifiability, we constrain the mean of the intercept and the mean of the first linear trend to zero.
4. *Three or more years of data*: $\mathbf{Q}_{u^c}$ has rank T^2 because the mean and both of the linear trends are identified by the data. For identifiability, we constrain the means of the intercept and both linear trends to zero.

We accept $\mathbf{u}^{c*}$ and λ_c^* with probability r:

$$(9) \quad r = \frac{P(\mathbf{y}|\mathbf{u}^{c*}, .)P(\mathbf{u}^{c*}|\lambda_c^*)P(\log \lambda_c^*)P(\mathbf{u}^{c(i-1)}|\mathbf{A}\mathbf{u}_j^{c(i-1)} = 0\,\forall j, \lambda_c^{i-1}, \mathbf{y}, .)}{P(\mathbf{y}|\mathbf{u}^{c(i-1)}, .)P(\mathbf{u}^{c(i-1)}|\lambda_c^{(i-1)})P(\log \lambda_c^{(i-1)})P(\mathbf{u}^{c*}|\mathbf{A}\mathbf{u}_j^{c*} = 0\,\forall j, \lambda_c^{i-1}, \mathbf{y}, .)}$$

Here, $\mathbf{A}$ is a $3 \times T^2$ constraint matrix, where the first row is a vector $\mathbf{1}$, the second row is a vector of centred time values for the DBP variable, and the third row is a vector of centred time values for the SBP variable.

5. Simulation study. To validate model fit and to establish that the complex MCMC sampler did not require excessive computational time, we simulated a dataset for a large number of countries. Specifically, we used the NCD Risk Factor Collaboration regional classification for blood pressure (NCD Risk Factor Collaboration (2017*a*)). We simulated data for three countries from 20 regions, and data for Japan, the only country simulated from its region. The simulated data were therefore similar in structure to earlier real data for which estimates were made separately for SBP and DBP (NCD Risk Factor Collaboration (2017*a*)). We simulated data for eight age groups and five years, with values based on published estimates by country. Different changes in mean blood pressure by age were simulated, both with values that increased monotonically with age, and others with more complex patterns to reflect the effects of interventions such as anti-hypertensive medication. These simulations produced substantial sparsity in the data, which tested the models ability to borrow strength appropriately.

We carried out five folds of cross-validation, with the average coverage and relative errors shown for super-regions in Table 1. The coverage of actual uncertainty by the estimated credible intervals was generally more than 95% (i.e. conservative), although with some cases where the coverage was less for the interactions (i.e., anti-conservative). As shown in Table 1, the relative errors were small, demonstrating that the model fits well. This can also be seen in Figure 5, which provides an example of model fitting for 50-year-old females in Belgium, where the patterns observed in the data are complex.

TABLE 1
Coverage and relative errors in cross-validation

	Coverage			Relative Errors		
Regions	DBP	SBP	Int	DBP	SBP	Int
Total	0.984	0.986	0.948	-0.002	-0.002	-0.045
Central and Eastern Europe	0.994	1	0.952	-0.003	-0.001	-0.017
C. Asia, Middle East & N. Africa	0.990	0.987	0.865	-0.002	-0.005	0.031
East and South East Asia	0.955	0.997	0.897	-0.003	-0.001	-0.030
High-income Western countries	0.998	0.994	0.996	0.000	-0.001	0.046
Latin America and Caribbean	0.978	0.982	1	-0.001	0.000	-0.151
Oceania	0.970	1	0.955	-0.002	-0.002	-0.050
South Asia	0.990	0.983	0.830	-0.004	-0.003	0.050
Sub-Saharan Africa	0.990	0.982	0.954	-0.002	-0.002	-0.095

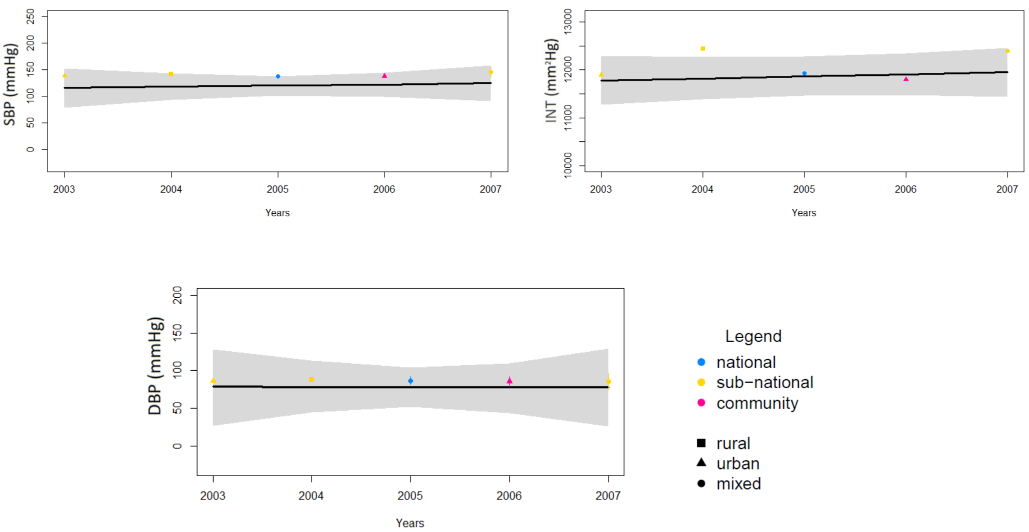

FIG 5. *Fit plots for 50-year-old females in Belgium from cross validation. The solid line represents the posterior mean and the shaded area the 95% credible interval.*

6. Application to real data. Performance of the model using real blood pressure data was tested using the STEPwise Approach to NCD Risk Factor Surveillance (STEPS) studies, whose results are made publicly available by WHO (https://www.who.int/teams/noncommunicable-diseases/surveillance/data). We downloaded data from 38 countries for the period 2004 to 2014 that comprised 478 country-year-sex-age combinations. Data were available from each year but were unevenly spread over time, with particularly large numbers of data points in 2007 and 2014. While a large proportion of the data comprised national-level estimates, 3% of the data points were at subnational level, and 6% were from rural-only or urban-only samples. Data were available for ages from 15 to 74 years, with the largest samples being for age groups in the range from 25 to 64 years. Sample sizes ranged from 46 to 3,036, while the ranges for DBP and SBP were 59.4-96.7 mmHg and 93.5-155.2 mmHg, respectively.

The data were from low-income and middle-income countries in central, south, southeast and east Asia, eastern Europe, Latin America and the Caribbean, the Middle East and north Africa, Oceania and sub-Saharan Africa. Although we had data for 38 countries, we made estimates for 48 countries in total from those regions, with cases where countries had data for zero, one, two or three or more years, i.e., each of the cases described in Section 4.2. The precision matrix described in (8) was $11^2 \times 11^2$ dimensional, and while this increased the time required for the MCMC algorithm to converge and mix, it was still tractable using a high-performance computing system. Figure 6 shows an example model fit for Myanmar, which exhibits realistically smooth but non-linear trends in the three blood pressure variables simultaneously. These effects, and others seen for different countries, would not have been captured fully by the models used in previous estimation of blood pressure trends.

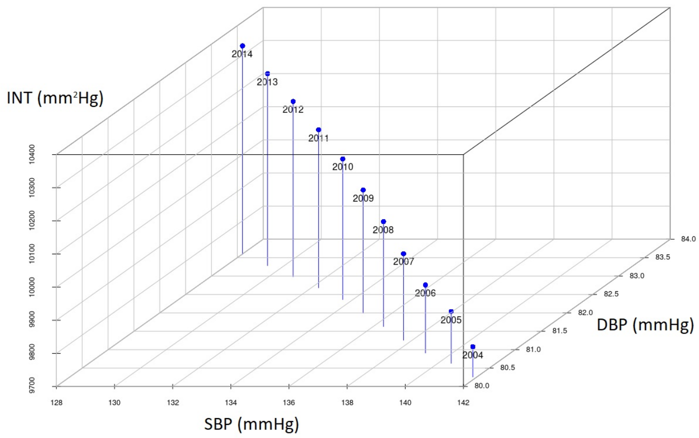

FIG 6. *Blood pressure estimates for 50-year old males in Myanmar from 2004 to 2014.*

7. Summary and future work. We have shown that it is possible to fit a model to mean blood pressure data in two dimensions, and in particular that it is possible to fit a non-linear surface that implicitly estimates the interaction between these variables. Our practical implementation will be of interest to any researchers with sparse bivariate data, but particularly to those with health-related data that vary by age and sex, over time and between countries, and where interactions are of substantial scientific importance.

Future work could include systematic data collation and publication of estimates, as carried out previously for SBP and DBP as single variables, and there are also interesting statistical challenges. For example, given that the hyperparameters need to be parametrised each time the number of time points or the order of the IGMRF is changed, it would be of interest to explore the use of penalised complexity priors. Other interesting research could explore alternatives to MCMC that would allow data for long time periods to be modelled in a tractable manner, or modelling to be carried out in more than two dimensions.

8. Software. R code, simulated data, and documentation are available from the corresponding author.

Acknowledgments. Specialist and High Performance Computing systems were provided by Information Services at the University of Kent. *Conflict of Interest*: None declared.

Funding. The first author was supported by a Vice Chancellor's Research Scholarship.

REFERENCES

DANAEI, GOODARZ, FINUCANE, MARIEL M, LIN, JOHN K, SINGH, GITANJALI M, PACIOREK, CHRISTOPHER J *and others*. (2011). National, regional, and global trends in systolic blood pressure since 1980: Systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. *Lancet* **377**(9765), 568–577.

DEVELOPMENT INITIATIVES. (2020). Global Nutrition Report: Action on Equity to End Malnutrition. *Technical Report*, Development Initiatives, Bristol, UK.

FINUCANE, MARIEL M, PACIOREK, CHRISTOPHER J, DANAEI, GOODARZ AND EZZATI, MAJID. (2014). Bayesian estimation of population-level trends in measures of health status. *Statistical Science* **29**(1), 18–25.

NCD RISK FACTOR COLLABORATION. (2016). Worldwide trends in diabetes since 1980: a pooled analysis of 751 population-based studies with 4·4 million participants. *Lancet* **387**(10027), 1513–1530.

NCD RISK FACTOR COLLABORATION. (2017*a*). Worldwide trends in blood pressure from 1975 to 2015: a pooled analysis of 1479 population-based measurement studies with 19·1 million participants. *Lancet* **389**(10064), 37–55.

NCD RISK FACTOR COLLABORATION. (2017*b*). Worldwide trends in body-mass index, underweight, overweight, and obesity from 1975 to 2016: a pooled analysis of 2416 population-based measurement studies in 128·9 million children, adolescents, and adults. *Lancet* **390**(10113), 2627–2642.

PACIOREK, CHRISTOPHER J. (2009). Understanding intrinsic Gaussian Markov random field spatial models, including intrinsic conditional autoregressive models. *Technical Report*, University of California, Berkeley, California, USA.

RUE, HÅVARD AND HELD, LEONHARD. (2005). *Gaussian Markov random fields: theory and applications*. CRC press, Boca Raton, Florida, USA.

SØRBYE, SIGRUNN HOLBEK AND RUE, HÅVARD. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics* **8**, 39–51.

THON, KEVIN, RUE, HÅVARD, SKRØVSETH, STEIN OLAV AND GODTLIEBSEN, FRED. (2012). Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. *Computational Statistics & Data Analysis* **56**(1), 49–61.

WORLD HEALTH ORGANIZATION. (2016). Global Report on Diabetes. *Technical Report*, World Health Organization, Geneva, Switzerland.

WORLD HEALTH ORGANIZATION. (2018). Global Status Report on Noncommunicable Diseases 2018. *Technical Report*, World Health Organization, Geneva, Switzerland.

WORLD OBESITY FEDERATION. (2019). Global Atlas on Childhood Obesity. *Technical Report*, World Obesity Federation, London, UK.

YUE, YU AND SPECKMAN, PAUL L. (2010). Nonstationary spatial Gaussian Markov random fields. *Journal of Computational & Graphical Statistics* **19**(1), 96–116.

Appendix B: Paper to be submitted to Statistica Neerlandica before viva

ORIGINAL ARTICLE

Scaling priors for Intrinsic Gaussian Markov Random Fields applied to blood pressure data

Maria-Zafeiria Spyropoulou[1] | James Bentham[2]

[1]School of Sport and Exercise Science, University of Kent, UK

[2]School of Statistics, Mathematics, and Actuarial Science, University of Kent, UK

Correspondence
Maria-Zafeiria Spyropoulou, School of Sport and Exercise Science, University of Kent, Canterbury, CT2 7FS, UK
Email: mzs2@kent.ac.uk

Present address
School of Sport and Exercise Science, University of Kent, Canterbury, CT2 7FS, UK

Funding information
Vice-Chancellor's Research Scholarship, University of Kent

Intrinsic Gaussian Markov Random Fields (IGMRFs) can be used to induce conditional dependence in Bayesian hierarchical models. IGMRFs have both a precision matrix, which defines the neighbourhood structure of the model, and a precision, or scaling, parameter.

Previous studies have shown the importance of selecting the prior of this scaling parameter appropriately for different types of IGMRF, as it can have a substantial impact on posterior estimates. Here, we focus on cases in one and two dimensions, where tuning of the prior is achieved by mapping it to the marginal standard deviation of an IGMRF of corresponding dimensionality. We compare the effects of scaling various IGMRFs, including applications to simulated and real blood pressure data using MCMC methods.

Key words: Hyperpriors, Intrinsic Gaussian Markov Random Fields, MCMC, Precision, Scaling, Two-dimensional data.

1 | INTRODUCTION

Intrinsic Gaussian Markov Random Fields (IGMRFs) are used widely as prior distributions in Bayesian hierarchical models, particularly for modelling spatial or temporal data, as they capture conditional dependence through their precision matrices (Rue and Held, 2005). We examine one- and two-dimensional IGMRFs; the latter can capture dependence between a pair of variables at multiple time points. IGMRFs are of various types, and can be specified to induce particular neighbourhood structures for the precisions, either by varying weights, introducing certain behaviour at boundaries or within the precision matrix, or by considering different sets of neighbours (Terzopoulos, 1988).

We present results for the two-dimensional case that generalise previous work on scaling different types of field in one dimension. In that case, an IGMRF was used as the prior for capturing non-linear trends, with a hyperprior for the precision parameter (Sørbye and Rue, 2014); we select precision parameters so that the same degree of scaling is applied to bivariate data as in the one-dimensional case. We have shown previously that appropriate model fitting to simulated and real two-dimensional blood pressure data can be achieved when applying MCMC methodology (Spyropoulou and Bentham, Under review (preprint: https://arxiv.org/abs/2111.07848). In this paper, we show that scaling requires particular care in two dimensions, where differences in behaviour between IGMRFs may be larger than in a single dimension.

The paper is structured as follows. Section 2 describes the behaviour of IGMRFs in one and two dimensions, while Section 3 describes the mapping between the precision parameter and the marginal standard deviation for two-dimensional IGMRFs. Analyses using simulated and real blood pressure data are presented in Section 4, with a discussion of our findings and suggestions for future work in Section 5.

2 | USE OF IGMRFS AS PRIORS

2.1 | Motivation

Blood pressure is bivariate, with measurements comprising systolic and diastolic values (SBP and DBP, respectively). While a realistic one-dimensional model of trends in mean SBP or DBP at national level has been developed (Danaei et al., 2011; Finucane et al., 2014), it cannot make estimates for both variables simultaneously, and no information is captured on interactions between them. We have developed a two-dimensional extension, including analogous terms to the original model (Spyropoulou and Bentham, Under review (preprint: https://arxiv.org/abs/2111.07848). Specifically, a vector $\mathbf{y}_{h,i}$ of the means of SBP, DBP, and their interactions, indexed by age group h and study i in country j, is assumed to be distributed

$$\mathbf{y}_{h,i} \sim \mathcal{N}(\mathbf{a}_{j[i]} + \mathbf{b}_{j[i]} t_i + \mathbf{u}_{j[i],t_i} + \mathbf{X}_i \boldsymbol{\beta} + \gamma_i(z_h) + \mathbf{e}_i, \mathbf{SD}^2_{h,i}/n_{h,i} + \tau_i^2) \tag{1}$$

The model includes country-level linear intercepts and slopes, $\mathbf{a}_{j[i]}$ and $\mathbf{b}_{j[i]}$, time-varying non-linear terms, $\mathbf{u}_{j[i],t_i}$, covariate effects $\boldsymbol{\beta}$, terms in age γ_i, study-specific random effects $\mathbf{e}_i$, age-varying study-specific random effects $\mathbf{w}_{h,i}$ corresponding to τ_i^2, and noise $\epsilon_{h,i}$, assumed iid Gaussian. In the earlier work, a one-dimensional second-order IGMRF was used as a prior for the $\mathbf{u}_{j[i]}$ terms, which we have extended to the two-dimensional case, including changes in the degree of scaling applied.

2.2 | IGMRFs of one and two dimensions

We begin by comparing the behaviour of one-dimensional first-order and second-order IGMRFs with the two-dimensional second-order case (Rue and Held, 2005). An IGMRF can be defined as

$$\pi(\mathbf{x}) = (2\pi)^{-(n-k)/2}(|\mathbf{Q}^*|) \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^T \mathbf{Q}(\mathbf{x}-\boldsymbol{\mu})\right) \tag{2}$$

where k denotes the order of the IGMRF, n is the total number of nodes and the rank is defined as $n-k$. As described previously (Rue and Held, 2005), for a vector of observations $\mathbf{u}$ of length n, the one-dimensional first-order model assumes independent increments

$$\Delta u_s = u_{s+1} - u_s \sim \mathcal{N}(0, \lambda^{-1}), \quad s = 1, \ldots, n-1 \tag{3}$$

with joint density

$$\pi(\mathbf{u}|\lambda) \propto \lambda^{(n-1)/2} \exp\left(-\frac{\lambda}{2}\sum_{s=1}^{n-1}(u_{s+1}-u_s)^2\right) \tag{4}$$

The second-order model assumes independent increments

$$\Delta^2 u_s = u_{s+2} - 2u_{s+1} + u_s \sim \mathcal{N}(0, \lambda^{-1}), \quad s = 1, \dots, n-2 \tag{5}$$

with joint density

$$\pi(\mathbf{u}|\lambda) \propto \lambda^{(n-2)/2} \exp\left(-\frac{\lambda}{2}\sum_{s=1}^{n-2}(u_{s+2}-2u_{s+1}+u_s)^2\right) \tag{6}$$

In two dimensions, a second-order model constructed on a torus assumes independent two-dimensional second-order increments (Rue and Held, 2005), and for variables indexed d and s we have

$$\begin{aligned}\Delta_0^2 u_{d,s} &= (\Delta^2_{(1,0)} + \Delta^2_{(0,1)})u_{d,s}\\ &= u_{d+2,s} - 2u_{d+1,s} + 2u_{d,s} - 2u_{d,s+1} + u_{d,s+2}\\ &= u_{d+1,s} - 4u_{d,s} + u_{d-1,s} + u_{d,s+1} + u_{d,s-1} \sim \mathcal{N}(0, \lambda^{-1})\end{aligned} \tag{7}$$

with joint density

$$\pi(\mathbf{u}|\lambda) \propto \lambda^{(n_1\times n_2-3)/2} \exp\left(-\frac{\lambda}{2}\sum_{d=2}^{n_1-1}\sum_{s=2}^{n_2-1}(\Delta^2_{(1,0)}u_{d,s} + \Delta^2_{(0,1)}u_{d,s})^2\right) \tag{8}$$

where n_1 and n_2 represent the total number of nodes for each variable.

Our models are time-varying, so the assumption of an IGMRF on a torus is not appropriate, and the two-dimensional second-order density in this case (Yue and Speckman, 2010) has the more complex form

$$\begin{aligned}\pi(\mathbf{u}|\lambda) \propto \lambda^{(n_1\times n_2-3)/2} \exp\Bigg(&-\frac{\lambda}{2}\sum_{d=2}^{n_1-1}\sum_{s=2}^{n_2-1}\{\Delta_0^2 u_{d,s}\}^2 + \{\Delta_1 u_{1,1}\}^2 + \{\Delta_2 u_{n_1,1}\}^2\\ &+ \{\Delta_3 u_{1,n_2}\}^2 + \{\Delta_4 u_{n_1,n_2}\}^2 + \sum_{d=2}^{n_1}(\{\Delta_5 u_{d,1}\}^2 + \{\Delta_6 u_{d,n_2}\}^2) + \sum_{s=2}^{n_2}(\{\Delta_7 u_{1,s}\}^2 + \{\Delta_8 u_{n_1,s}\}^2)\Bigg)\end{aligned} \tag{9}$$

A special case of (9) arises when the variables have the same number of nodes, i.e., $n = n_1 = n_2$

$$\begin{aligned}\pi(\mathbf{u}|\lambda) \propto \lambda^{(n^2-3)/2} \exp\Bigg(&-\frac{\lambda}{2}\sum_{d=2}^{n-1}\sum_{s=2}^{n-1}\{\Delta_0^2 u_{d,s}\}^2 + \{\Delta_1 u_{1,1}\}^2 + \{\Delta_2 u_{n,1}\}^2\\ &+ \{\Delta_3 u_{1,n}\}^2 + \{\Delta_4 u_{n,n}\}^2 + \sum_{d=2}^{n}(\{\Delta_5 u_{d,1}\}^2 + \{\Delta_6 u_{d,n}\}^2) + \sum_{s=2}^{n}(\{\Delta_7 u_{1,s}\}^2 + \{\Delta_8 u_{n,s}\}^2)\Bigg)\end{aligned} \tag{10}$$

In each case, we have time-varying non-linear effects $\mathbf{u} \sim \mathcal{N}(0, (\lambda\mathbf{P})^{-1})$. They follow an IGMRF with structure matrix, $\mathbf{P}$, and precision parameter, λ, which is always a scalar (Yue and Speckman, 2010).

3 | SPECIFYING HYPERPRIORS FOR TWO-DIMENSIONAL IGMRFS

IGMRFs have structure matrices and marginal variances that change depending on their type. Hyperpriors need to be chosen and assigned appropriate ranges for particular models based on their structure, particularly their dimensionality and numbers of nodes (Sørbye and Rue, 2014). For our model, we need to scale λ appropriately for a two-dimensional second-order IGMRF with boundaries and up to 40 nodes. Here, we derive reference standard deviations and use them to select appropriate values for specific hyperpriors.

3.1 | Reference standard deviation

We can describe an IGMRF using an alternative definition where

$$\sigma^2_\lambda(u_i) = \lambda^{-1}\Sigma^*_{ii} \tag{11}$$

Here, Σ_{ii} is the diagonal element of the covariance matrix in position i. For the standard normal distribution, we have

$$\mathbf{u}\sqrt{\lambda} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{P}^{-1}\right), \quad \lambda\sigma^2_\lambda(u_i) = \Sigma^*_{ii} \tag{12}$$

Therefore for $\lambda = 1$, we have

$$\sigma^2_{\{\lambda=1\}}(u_i) = \Sigma^*_{ii} \tag{13}$$

Combining the results in (11) and (13), we have

$$\sigma^2_\lambda(u_i) = \lambda^{-1}\Sigma^*_{ii} = \lambda^{-1}\sigma^2_{\{\lambda=1\}}(u_i) \tag{14}$$

This implies that for any fixed precision λ, the marginal standard deviation of the components of a Gaussian vector $\mathbf{u}$ can be expressed as a function of λ by (Sørbye and Rue, 2014)

$$\sigma_\lambda(u_i) = \lambda^{-1/2}\sigma_{\{\lambda=1\}}(u_i), \quad i = 1, \ldots, n \tag{15}$$

For a given IGMRF $\mathbf{u}$ with random precision λ, we can calculate a reference standard deviation for fixed $\lambda = 1$, and then approximate the marginal standard deviation for each component of $\mathbf{u}$ by (Sørbye and Rue, 2014)

$$\sigma_\lambda(u_i) \approx \lambda^{-1/2}\sigma_{ref}(\mathbf{u}), \quad i = 1, \ldots, n \tag{16}$$

The reference standard deviation is calculated using the geometric mean, an appropriate measure for a set of positive numbers (Sørbye and Rue, 2014). The reference standard deviation for $\mathbf{u}$ in the one-dimensional case is then

$$\sigma_{ref}(\mathbf{u}) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\log\sigma_{\{\lambda=1\}}(u_i)\right) = \exp\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2}\log\Sigma^*_{ii}\right) \tag{17}$$

where the values Σ^*_{ii} denote the diagonal elements of the inverse matrix $\mathbf{\Sigma}^* = \mathbf{Q}^{-1}$ calculated when $\lambda = 1$. Specifically, this is calculated as $\mathbf{Q}^{-1} = \mathbf{\Gamma}^T \Lambda^{-1} \mathbf{\Gamma}$, where $\mathbf{\Gamma}$ are the eigenvectors and $\mathbf{\Lambda}$ the eigenvalues of $\mathbf{Q}$ when $\lambda = 1$. Since $\mathbf{Q}$ and $\mathbf{\Sigma}^*$ are both $n \times n$ dimensional for one-dimensional IGMRFs, the n diagonal values are used to calculate the geometric mean.

We extend the calculation of $\sigma_{ref}(\mathbf{u})$ to two-dimensional second-order IGMRFs. The precision matrix is now $(n_1 \times n_2) \times (n_1 \times n_2)$ dimensional, where n_1 and n_2 are the total number of nodes for the first and second variables respectively. The scaling is no longer for n_1 or n_2 values, but their product $n_1 \times n_2$, with a special case when $n = n_1 = n_2$

$$\sigma_{ref}(\mathbf{u}) = \exp\left(\frac{1}{n^2}\sum_{i=1}^{n^2} \log \sigma_{\{\lambda=1\}}(u_i)\right) = \exp\left(\frac{1}{n^2}\sum_{i=1}^{n^2} \frac{1}{2} \log \mathbf{\Sigma}^*_{ii}\right) \tag{18}$$

Again, $\mathbf{\Sigma}^*_{ii}$ denotes the diagonal elements of the inverse matrix $(\mathbf{\Sigma}^*)^{1/2} = (\mathbf{Q}^{-1})^{1/2}$, while $\mathbf{\Sigma}^* = \mathbf{Q}^{-1} = \mathbf{\Gamma}^T \Lambda^{-1} \mathbf{\Gamma}$ for $\lambda = 1$. The precision matrix $\mathbf{Q}$, and therefore $\mathbf{\Sigma}^*$, is $n^2 \times n^2$ dimensional, hence there are n^2 elements in the diagonal.

It has been shown previously (Sørbye and Rue, 2014) that using 100 nodes, the reference standard deviations in one dimension are $\sigma_{ref}(\mathbf{u}) = 3.89$ and $\sigma_{ref}(\mathbf{u}) = 41.39$ for the first-order and second-order cases, respectively; applying the result in (18), we find that $\sigma_{ref}(\mathbf{u}) = 7.24$ for the two-dimensional second-order case for 100 nodes in each dimension. As shown in Figure 1, the behaviour by node of the marginal standard deviations also varies substantially between the three IGMRFs. For a particular hyperprior, larger variances would be allowed for the one-dimensional second-order IGMRF than its two-dimensional equivalent, and both would have larger variances than the one-dimensional first-order case. Equivalently, to allow the same variance we need to impose an upper limit on the marginal standard deviation

$$Pr(\sigma(u_i) > U) \approx Pr\left(\frac{\lambda}{\sigma^2_{ref}(\mathbf{u})} < \frac{1}{U^2}\right) = \alpha \tag{19}$$

where α is a fixed small probability (Sørbye and Rue, 2014). By assigning a hyperprior to $\lambda(\sigma^2_{ref}(\mathbf{u}))^{-1}$, the interpretation of the hyperprior remains the same for the different models.

These results complement others (Lindgren and Rue, 2008; Lindgren et al., 2011), where k equally sized subintervals are created between original nodes $u_1, u_2, \ldots$, to give equidistant nodes $u'_1, u'_2, \ldots, u'_{k+1}$. In the first-order one-dimensional case, the precision using the new nodes is $(k\lambda)^{-1}$, for the second-order equivalent, the precision using the new nodes is $(k^3\lambda)^{-1}$, and finally, for the second-order two-dimensional IGMRF, the precision using the new nodes is $(k^2\lambda)^{-1}$.

3.2 | Specifications using Gaussian hyperpriors

Applying a Gaussian hyperprior, the upper limit expressed in probabilistic form in (19) is

$$U = \left(\frac{b\sigma^2_{ref}(\mathbf{u})}{F^{-1}(\alpha, \mu, 1)}\right)^{1/2} \tag{20}$$

where $F^{-1}(\cdot)$ denotes the quantiles of the Gaussian distribution (Sørbye and Rue, 2014). For a given value of α, we can then interpret the mean and standard deviation parameters, μ and b, in terms of this upper limit.

To recalculate hyperpriors for different IGMRFs, we can use the same mean parameter μ for each model and

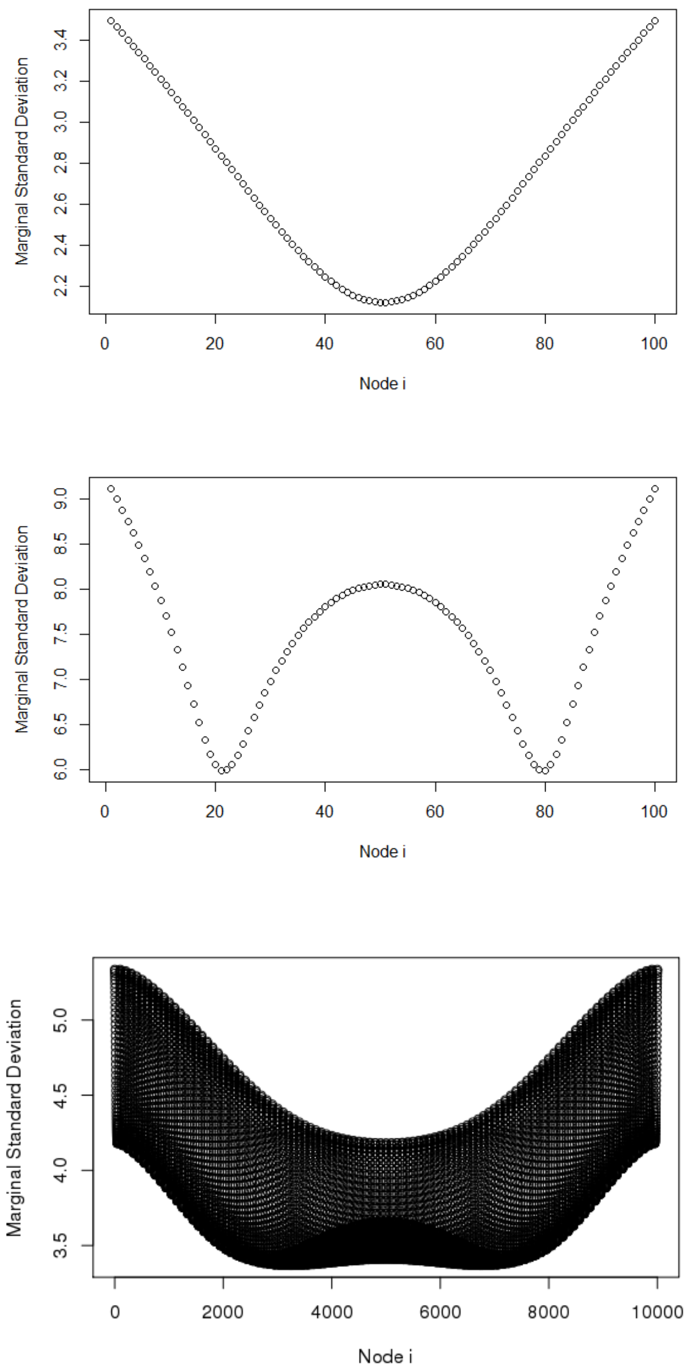

FIGURE 1 Marginal standard deviations of one-dimensional first-order and second-order (Sørbye and Rue, 2014), and two-dimensional second-order IGMRFs, calculated using fixed precision $\lambda = 1$.

calculate a new standard deviation parameter. By using the upper limit provided in (20), the new standard deviation parameter is expressed as

$$b_{new} = \frac{U^2 F^{-1}(\alpha, \mu, 1)}{\sigma^2_{ref}(\mathbf{u})} \tag{21}$$

We see that the new value of the standard deviation in (21) depends on σ^2_{ref}, which captures the precision matrix for a specific type of IGMRF. It is then only necessary to recalculate the standard deviation parameter, b, to account for the different shapes and sizes of the graph for a specific IGMRF (Sørbye and Rue, 2014). This can be done for the three types of IGMRF considered using

$$b_{rw2} = b_{rw1} \times \frac{\sigma^2_{ref}(\mathbf{u}_{rw1})}{\sigma^2_{ref}(\mathbf{u}_{rw2})}, \quad b_{rw2D} = b_{rw2} \times \frac{\sigma^2_{ref}(\mathbf{u}_{rw2})}{\sigma^2_{ref}(\mathbf{u}_{rw2D})}, \quad b_{rw2D} = b_{rw1} \times \frac{\sigma^2_{ref}(\mathbf{u}_{rw1})}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} \tag{22}$$

Here, *rw1* and *rw2* refer to the one- and two-dimensional first-order IGMRFs (Sørbye and Rue, 2014), and *rw2D* to the two-dimensional second-order IGMRF.

3.3 | Types of two-dimensional second-order IGMRFs

To provide a guide for researchers applying models to real data, we compare IGMRFs with fixed order but different numbers of nodes and boundary conditions, and IGMRFs of varying order and dimensionality. We do so firstly for two-dimensional second-order IGMRFs with four structure matrices: Torus 1 and Torus 2 (Rue and Held, 2005; Thon et al., 2012), and Bound 1 (Yue and Speckman, 2010) and Bound 2 (Terzopoulos, 1988). Torus 1 has a structure matrix defined on a torus, while Torus 2 has a similar structure matrix but with boundaries at its four corners, $u_{1,1}, u_{n_1,1}, u_{1,n_2}, u_{n_1,n_2}$. Bound 1 and Bound 2 have boundary effects and induce the same neighbours in the structure matrix for each node, but give different weightings to these neighbours. For our application to time-varying mean blood pressure data and for similar applications, the bounded IGMRFs are appropriate and should be used.

In Table 1, we see that Torus 2 consistently has the lowest reference standard deviation, with the changes in each IGMRF being similar proportionally when the number of nodes is increased. Bound 2 has the largest reference standard deviation, followed by Bound 1, which we used in our two-dimensional model of blood pressure (Spyropoulou and Bentham, Under review (preprint: https://arxiv.org/abs/2111.07848). These findings show that it is clearly necessary to scale the hyperparameter each time the precision matrix or number of nodes is changed, especially when boundary conditions are introduced.

Table 2 shows the effect of changing the order or dimensionality of the IGMRFs. For small numbers of nodes, the one-dimensional first-order random walk has the largest reference standard deviation. However, this increases slowly with the number of nodes, and for 10 nodes or more there is a consistent pattern with the one-dimensional second-order random walk having the largest reference standard deviations, while the two-dimensional second-order random walk behaves more similarly to the one-dimensional first-order random walk. These results emphasise the importance of scaling the hyperparameter when dimensionality or order is changed.

TABLE 1 Reference standard deviations σ_{ref} for second-order two-dimensional IGMRFs.

Nodes	Torus 1	Torus 2	Bound 1	Bound 2
5	0.30	0.05	0.41	0.53
6	0.35	0.06	0.48	0.63
8	0.44	0.08	0.62	0.81
10	0.54	0.09	0.76	1.00
12	0.63	0.11	0.90	1.19
14	0.73	0.12	1.04	1.39
16	0.82	0.14	1.19	1.58
18	0.93	0.16	1.33	1.77
20	1.02	0.17	1.47	1.96
25	1.25	0.21	1.83	2.44
30	1.51	0.25	2.19	2.92
40	2.00	0.33	2.91	3.88
50	2.50	0.42	3.63	4.84
100	5.00	0.83	7.24	9.64

TABLE 2 Reference standard deviations, σ_{ref}, for one-dimensional first-order and second-order IGMRFs, and two-dimensional second-order IGMRFs.

Nodes	$\sigma_{ref}(\mathbf{u}_{rw1})$	$\sigma_{ref}(\mathbf{u}_{rw2})$	$\sigma_{ref}(\mathbf{u}_{rw2D})$
5	0.85	0.53	0.41
6	0.94	0.65	0.48
8	1.09	0.96	0.62
10	1.22	1.35	0.76
12	1.34	1.75	0.90
14	1.45	2.20	1.04
16	1.55	2.68	1.19
18	1.65	3.19	1.33
20	1.74	3.73	1.47
25	1.94	5.20	1.83
30	2.13	6.82	2.19
40	2.46	10.49	2.91
50	2.75	14.65	3.63
100	3.89	41.39	7.24

4 | APPLICATION TO BLOOD PRESSURE DATA

4.1 | Simulation study

4.2 | Application using real data

We compare hyperprior scaling for one- and two-dimensional second-order IGMRFs using real blood pressure data (Spyropoulou and Bentham, Under review (preprint: https://arxiv.org/abs/2111.07848). The scaling varies both by dimensionality and number of nodes, which in our data corresponds to the number of years considered. The hyperpriors must be set for each of the four precision parameters, λ_c, λ_r, λ_s and λ_g, that are used at different levels of a Bayesian hierarchical model, with countries nested in regions, super-regions and the globe.

When using 40 years of data, we can calculate the following values

$$\sigma_{ref}(\mathbf{u}_{rw2}) = 10.49 \qquad \sigma_{ref}(\mathbf{u}_{rw2D}) = 2.91 \tag{23}$$

The distribution under consideration for the country-level precision parameter is $\lambda_c \sim \mathcal{N}(\mu, b)$, with μ and b parameters that are assigned the values

$$\mu = 7, \qquad b = 2, \qquad \alpha = 0.001 \tag{24}$$

Here, b is the adjusted parameter to which we must apply the correct scaling. From (23) and (24), the upper bounds for the one-dimensional and two-dimensional second-order IGMRFs are

$$\begin{aligned} U_{rw2} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2})}{F^{-1}(\alpha, \mu, 1)}\right)^{1/2} = 7.5 \\ U_{rw2D} &= \left(\frac{b\sigma^2_{ref}(\mathbf{u}_{rw2D})}{F^{-1}(\alpha, \mu, 1)}\right)^{1/2} = 2.08 \end{aligned} \tag{25}$$

There are no negative values in Gaussian quantiles, so to reproduce earlier results (Sørbye and Rue, 2011) using a Gaussian rather than a Gamma distribution, we need to proceed as if we have truncation below at zero. By taking the median, we have

$$\text{median}(U_{rw2}, U_{rw2D}) = U = 4.79 \tag{26}$$

The new standard deviation parameters for the hyperpriors are:

$$\begin{aligned} b_{rw2} &= \frac{U^2 F^{-1}(\alpha, \mu, 1)}{\sigma^2_{ref}(\mathbf{u}_{rw2})} = 0.81 \\ b_{rw2D} &= \frac{U^2 F^{-1}(\alpha, \mu, 1)}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 10.59 \end{aligned} \tag{27}$$

Alternatively, knowing $b_{rw2} = 0.81$,

$$b_{rw2D} = b_{rw2}\frac{\sigma^2_{ref}(\mathbf{u}_{rw2})}{\sigma^2_{ref}(\mathbf{u}_{rw2D})} = 10.59$$

We also observe clear patterns as the adjusted parameter, b, is varied. In earlier work on blood pressure modelling (Danaei et al., 2011), the standard deviation of a one-dimensional second-order IGMRF, b_{rw2}, was set to 3. In the case of five nodes, scaling makes this equivalent to 5.01 for the two-dimensional case, b_{rw2D}, while the adjusted parameter, b, is also equal to 3, as shown in Table 3. We can see variations in the tuning of b_{rw1}, b_{rw2} and b_{rw2D} as the number of nodes changes, and in particular cases, they each coincide with the adjusted parameter b. For example, we see that the adjusted parameter, b is equal to b_{rw1} when the number of nodes is 10; for five nodes, the adjusted parameter is equal to b_{rw2}; for 30 nodes, the adjusted parameter is equal to b_{rw2D}. Table 4 shows the scaling applied to standard deviations for a model with 11 nodes. With an adjusted parameter $b = 0.90$, scaling λ_c gives 0.53 and 1.84 for the one- and two-dimensional second-order models, respectively, for example. Together, these results allowed us to apply the same degree of smoothness in the two-dimensional second-order case as in the earlier work, scaling b_{rw2D} correctly given the variation in the blood pressure data.

TABLE 3 Scaling standard deviation parameters as the adjusted parameter b and number of nodes are varied.

Nodes	b=1			b=2			b=3		
	b_{rw1}	b_{rw2}	b_{rw2D}	b_{rw1}	b_{rw2}	b_{rw2D}	b_{rw1}	b_{rw2}	b_{rw2D}
5	0.39	1.00	1.67	0.78	2.00	3.34	1.17	3.00	5.01
6	0.48	1.00	1.84	0.96	2.00	3.67	1.43	3.00	5.49
8	0.78	1.00	2.40	1.55	2.00	4.80	2.33	3.00	7.19
10	1.00	0.82	2.58	2.00	1.63	5.16	3.00	2.45	7.74
12	1.00	0.59	2.22	2.00	1.17	4.43	3.00	1.76	6.65
14	1.00	0.43	1.94	2.00	0.87	3.89	3.00	1.30	5.83
16	1.00	0.33	1.70	2.00	0.67	3.40	3.00	1.00	5.09
18	1.00	0.27	1.54	2.00	0.53	3.08	3.00	0.80	4.62
20	1.00	0.22	1.40	2.00	0.43	2.80	3.00	0.65	4.20
25	1.00	0.14	1.12	2.00	0.28	2.25	3.00	0.42	3.37
30	1.06	0.10	1.00	2.11	0.21	2.00	3.17	0.31	3.00
40	1.40	0.08	1.00	2.80	0.15	2.00	4.20	0.23	3.00
50	1.74	0.06	1.00	3.48	0.12	2.00	5.23	0.18	3.00
100	3.46	0.03	1.00	6.93	0.06	2.00	10.39	0.09	3.00

5 | SUMMARY AND FUTURE WORK

We have shown the importance of correct scaling of hyperpriors for precision parameters in IGMRFs. This scaling varies with the dimensionality, order, and size of the IGMRFs, and also depends strongly on the structure of the precision matrices. We have presented general results in two dimensions, and specific applications to one- and two-dimensional models of blood pressure using simulated and real data.

TABLE 4 Scaling standard deviation parameters for one- and two-dimensional IGMRFs with 11 nodes.

λ	b	b_{rw1}	b_{rw2}	b_{rw2D}
λ_c	0.90	0.77	0.53	1.84
λ_r	1.20	1.03	0.71	2.45
λ_s	1.59	1.36	0.94	3.24
λ_g	3.55	3.04	2.10	7.24

Future work could include applying penalised complexity (PC) priors as precision parameters for two-dimensional random effects (Simpson et al., 2017). The precision parameter corresponds to a second-order IGMRF, with $\mathbf{u} \sim \mathcal{N}(0, \lambda^{-1}\mathbf{P}^{-1})$. However, a model can have two types of random effects, constructed and unconstructed. They have dependent precision parameters, and so a joint bivariate or multivariate distribution should express this dependence. For the two-dimensional second-order IGMRF, the precision parameter is univariate but we could investigate the use of PC priors, which have the property that no further scaling is required as the number of nodes is varied.

Acknowledgements Specialist and High Performance Computing systems were provided by Information Services at the University of Kent.

References

Goodarz Danaei, Mariel M Finucane, John K Lin, Gitanjali M Singh, Christopher J Paciorek, et al. National, regional, and global trends in systolic blood pressure since 1980: systematic analysis of health examination surveys and epidemiological studies with 786 country-years and 5·4 million participants. *Lancet*, 377(9765):568–577, 2011.

Mariel M Finucane, Christopher J Paciorek, Goodarz Danaei, and Majid Ezzati. Bayesian estimation of population-level trends in measures of health status. *Statistical Science*, 29(1):18–25, 2014.

Finn Lindgren and Håvard Rue. On the second-order random walk model for irregular locations. *Scandinavian Journal of Statistics*, 35(4):691–700, 2008.

Finn Lindgren, Håvard Rue, and Johan Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011.

Håvard Rue and Leonhard Held. *Gaussian Markov random fields: theory and applications*. CRC Press, New York, USA, 2005.

Daniel Simpson, Håvard Rue, Andrea Riebler, Thiago G Martins, and Sigrunn H Sørbye. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, 32(1):1–28, 2017.

Sigrunn H Sørbye and Håvard Rue. Simultaneous credible bands for latent Gaussian models. *Scandinavian Journal of Statistics*, 38(4):712–725, 2011.

Sigrunn H Sørbye and Håvard Rue. Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, 8:39–51, 2014.

Maria-Zafeiria Spyropoulou and James Bentham. A two-dimensional intrinsic Gaussian Markov random field for blood pressure data. *Annals of Applied Statistics*, Under review (preprint: https://arxiv.org/abs/2111.07848).

Demetri Terzopoulos. The computation of visible-surface representations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):417–438, 1988.

Kevin Thon, Håvard Rue, Stein O Skrøvseth, and Fred Godtliebsen. Bayesian multiscale analysis of images modeled as Gaussian Markov random fields. *Computational Statistics & Data Analysis*, 56(1):49–61, 2012.

Yu Yue and Paul L Speckman. Nonstationary spatial Gaussian Markov random fields. *Journal of Computational and Graphical Statistics*, 19(1):96–116, 2010.