A two-stage Bayesian modelling framework with applications in spatial epidemiology

Stephen Jun Vecera Villejo

School of Mathematics and Statistics

University of Glasgow

A thesis submitted for the degree of

Doctor of Philosophy

Declaration

I, Stephen Jun Vecera Villejo, declare that this thesis, titled "A two-stage Bayesian modelling framework with applications in spatial epidemiology" and the work presented within it are my own. I confirm that all sources consulted, including the published work of others, are clearly acknowledged.

The work in Chapter 3 was presented at the 24th International Conference on Computational Statistics at the University of Bologna, Italy, in August 2022, and has been published in the *Spatial Statistics* journal (10.1016/j.spasta.2023.100744). Initial work discussed in Chapters 4 and 5 was presented at the 6th Spatial Statistics conference at the University of Colorado, Boulder, USA, in July 2023. The work in Chapter 4 has been published at the *Journal of the Royal Statistical Society Series C: Applied Statistics* (10.1093/jrsssc/qlaf012). The work in Chapters 5 and 6 is currently under peer review, with articles accessible on `arXiv` at arXiv:2506.22334 and arXiv:2502.18962, respectively.

Most of the results in Chapter 4 were conceived and written during a research visit at the Norwegian University of Science and Technology, in collaboration with Professor Sara Martino and Professor Finn Lindgren, both of whom are co-authors of the published article. The work presented in Chapter 6 had significant contributions from Professor Lindgren, who suggested the use of the simulation-based calibration method to validate the Bayesian algorithms. There were also significant contributions from Professor Martino and William Ryan, who are co-authors of the resulting publication. Finally, the initial results on fitting Mixed Data Sampling (MIDAS) regression models using INLA, presented in Chapter 7, were conceived with the guidance from Professor Håvard Rue.

And here, poor fool!

with all my lore I stand,

no wiser than before

JOHANN WOLFGANG VON GOETHE

Abstract

This thesis proposes a framework for doing two-stage modelling in spatial epidemiology, whose main goal is to understand the association between a covariate of interest, which is modelled in the first stage, and health outcomes, which is modelled in the second stage using the first-stage model predictions as inputs. A two-stage modeling framework has the advantage of being more computationally efficient than a joint modelling approach when the first-stage model is already complex in itself, and avoids the potential problem of unwanted feedback effects, which happen when the second-stage data affect first-stage model inference. Chapter 1 discusses the motivation behind this research. The specific data application of this thesis links dengue incidence and climate variables, particularly temperature, relative humidity, and rainfall, in the Philippines. Dengue is an infectious disease caused by *Aedes* mosquitoes, and which poses significant socioeconomic and disease burden in many tropical and subtropical regions of the world.

In a two-stage modelling framework, the first stage fits the model for the main covariate of interest, whose association with the health outcome is investigated. In the data application, the first-stage fits climate models, which are then used to predict the true climate field over the entire spatial domain. The data limitation, which poses challenges on the accuracy of model inference and predictions, is the sparsity in the data from weather stations. This data problem is overcome by incorporating additional data sources (referred to as *proxy data*), albeit more biased but with wider spatial coverage, and then combining the different data sources in a process called data fusion, whose main goal is the improvement of model accuracy. Chapter 3 presents an initial exploration of a data fusion Bayesian

model estimated using integrated nested Laplace approximation (INLA). Chapter 4 presents a flexible model specification of the data fusion model, which is shown to outperform benchmark approaches in terms of the accuracy of model predictions and parameter estimates. The proposed model specifies both a time-varying random field to account for the additive bias and a constant multiplicative bias parameter in the proxy data. Chapter 4 also presents the results from applying the proposed data fusion model on the meteorological data in the Philippines. The results of leave-group-out cross validation show that the data fusion model outperforms benchmark approaches.

Chapter 5 presents results from an extensive analysis on the link between climate and dengue occurrence in the Philippines. The predicted climate fields from Chapter 4 are used as inputs to the health model. To account for the uncertainty in the predictions from the climate models, a resampling approach is used, which generates samples from the first-stage model posteriors and where each sample is used as an input to the second-stage model. The final posterior estimates of second-stage model parameters are then computed using Bayesian model averaging. The results show that temperature has a non-linear relationship with dengue occurrence. In particular, temperature is generally positively related to dengue, but very hot conditions tend to have a negative impact. Moreover, the relationship between rainfall and dengue varies in space, depending on the climate type of the area. For areas with uniform and low variation in the amount of rainfall all year round, rainfall is negatively associated with dengue, while for areas with pronounced dry and wet season, rainfall is positively related with dengue. This is potentially explained by the fact that consistent rainfall tends to wash away mosquito breeding sites, while sporadic rainfall during dry season tends to create more breeding sites.

Chapter 6 investigates the correctness of the two approaches for doing two-stage modelling used in Chapter 5, particularly the crude plug-in approach, which simply plugs in the posterior mean of the first-stage

(climate) model parameters to the second-stage (health) model, and the resampling approach. I used the simulation-based calibration (SBC) approach, which tests the self-consistency property of Bayesian models, to validate the correctness of the aforementioned approaches. Results show that the crude plug-in method indeed underestimates the posterior uncertainty in the second-stage model parameters, while the resampling approach is correct. This chapter also proposes a new approach for doing uncertainty propagation, called the **Q** uncertainty method, which introduces a new model component called the *error component*. The $\mathbf{Q}^{-1}$ matrix essentially encodes the uncertainty in the first-stage latent parameters. In addition, I proposed a low rank approximation of the **Q** matrix, which can be useful for large spatio-temporal applications. I also used the SBC method to validate the correctness of the proposed method. Results of model validation on toy spatial models show that the **Q** method can be correct, but the accuracy of the posterior approximations and the computational benefits of the method depends on the coarseness of the mesh for the error component and the dimension of the first-stage model latent parameters. The main reasons for the computational bottleneck with the proposed method is that the predictor expression of the **Q** method involves non-linear model components, which does not fit quite conveniently in the INLA framework.

Finally, Chapter 7, the conclusion chapter, highlights the main contributions of this thesis and outlines potential directions for future work. In addition, I reemphasize current approaches for fitting conditional latent Gaussian models, and provide ideas on a new approach for fitting such models. Whereas the previous chapters highlight the problem of spatial misalignment, the final chapter discusses the issue of time misalignment. I provide ideas and initial results from using INLA to fit Mixed-Data-Sampling (MIDAS) models, which provide a framework for fitting a regression model on time series data with varying frequencies.

Acknowledgements

I am profoundly grateful to all those who have supported me throughout my PhD journey.

First and foremost, I am immensely grateful to my main PhD supervisor, Professor Janine Illian, for believing in me from the very beginning, offering unwavering support, and accompanying me through every stage of this journey. I am also grateful to Dr Ben Swallow, for generously sharing his expertise and time as my second PhD supervisor during the first two years of my PhD, and to Dr Daniela Castro-Camilo, for providing valuable practical advice on both PhD and post-PhD life.

I have had the privilege of collaborating with other experts in my research area. I am deeply grateful to Professor Sara Martino for her substantial contributions and generous support, and for hosting my research visit at the Norwegian University of Science and Technology. I am also grateful to Professor Finn Lindgren for reviewing my work and offering constructive suggestions that influenced the direction of my thesis.

I am grateful to my PhD examiners, Professor Michela Cameletti and Dr Oliver Stoner, for carefully reviewing my work, providing constructive feedback and suggestions, and highlighting areas for improvement. I also thank Dr Craig Anderson, the convenor of the examination committee, for ensuring that the viva ran smoothly.

I acknowledge the support from the University of the Philippines, particularly my colleagues at the School of Statistics, and to Professor Joseph Ryan Lansangan for his assistance with administrative matters and paperwork during my application for PhD admission and study leave.

I am deeply grateful to the School of Mathematics and Statistics at the

University of Glasgow for the generous Maclaurin scholarship. I also extend my gratitude to the colleagues and staff at the School with whom I had the privilege to work with in my capacity as a part-time staff member. My sincere thanks to Dr Mitchum Bock, my teaching mentor, for his kindness and support, and to Dr Eilidh Jack, Dr Wei Zhang, Dr Tereza Neocleous, Professor Claire Miller, and Professor Philipp Otto, with whom I co-lectured courses. I am also grateful to Professor Jethro Browell, who, together with Professor Claire Miller, served as one of my reviewers for the PhD Annual Progress Review.

I am grateful to my church in Glasgow, Re:Hope Church, for being my spiritual family in the city – a constant source of joy and meaningful conversations. Special thanks to my friends in the worship team, too many to mention individually, for their fellowship and support.

My stay in Glasgow has truly felt like home, thanks to my Filipino friends and community. I am especially thankful to my library buddies – Amiel Gee, and Jackie – for being constants throughout this journey. Thanks also to Philip for being such a generous and supportive friend, and to Jius for your guidance and advice in many things. I extend my heartfelt appreciation to all my other Filipino friends, with whom I shared wonderful memories; regrettably, I am unable to mention individually due to their sheer number.

I am fortunate to have worked alongside remarkable PhD colleagues who became valued friends and companions, making my time as a doctoral student both enriching and memorable. I am grateful to Robin, Daniela, William, Vanessa, Gabriel, Pietro, Giovanni, Jacopo, Alba, Iain, Toby, Weiyue, Lanxin, and Chenglei.

I am grateful to Patryk, a kind, generous, and supportive friend. I also extend my gratitude to more friends with whom I shared memorable times in Glasgow – Mario, Mathew, Robert, Tobias, and Marketa. I also thank my wonderful flatmates, Wei and Hendrik, whom I consider lifelong friends.

I am also grateful to Kai for being a loyal and thoughtful buddy.

I am deeply grateful to specific friends who bridged distance and time zone differences. A special mention to Rutcher and Janronel, who were always present, serving as constant source of encouragement and trusted confidantes. I also thank Barton for his generosity and unwavering reliability, and Diam and Maggie, both of whom I consider family.

Finally, I am forever grateful to my family in the Philippines – my Mama Helen and Papa Junior, my siblings Eden and Exequiel, my brother-in-law Jeff, and nephew Arrow – as well as my aunties, uncles, and cousins, for their unwavering support throughout my entire PhD journey.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

This thesis presents a two-stage modelling framework for statistical modelling. This approach is relevant for addressing complex modelling problems by decomposing them into stages, thereby mitigating the computation burden that would arise from doing model inference and prediction when done otherwise. Two motivating examples in the area of spatial epidemiology are presented in Section 1.2. In such applications, the first stage fits a model for covariates of interest, such as climatic factors and pollutant concentrations. This stage may involve the integration of multiple data sources – for example, data from monitoring stations, outputs from numerical models, and satellite imagery. The second stage then models the health outcome, using the predictions from the first-stage model as inputs. While the two-stage modelling approach offers computational advantages, it necessitates the appropriate propagation of model uncertainty from the first stage to the second stage. Thus, this thesis examines the uncertainty propagation problem in two-stage models. It also proposes an approach for integrating multiple data sources – an essential feature of the first-stage model – through a process known as *data fusion*. A concrete data application is presented in the context of exploring the link between climate and dengue, an overview of which is provided in Section 1.2.1. Section 1.3 elaborates on the two-stage modelling framework, and Section 1.5 outlines the main objectives of this thesis.

1.2 Motivating examples

1.2.1 Climate and dengue

Dengue fever, an infectious disease caused by the dengue arbovirus and commonly transmitted by two mosquito species (*Aedes aegypti* and *Aedes albopictus*), poses a strong public health threat across tropical regions. This infectious disease imposes a substantial socioeconomic and disease burden in many tropical and subtropical areas around the world (Murray et al., 2013). Over half of the world's population live in areas at risk of the disease. From 1990 to 2019, the estimated increase in dengue incidence is 85.5%; and from 2021, the number of global cases has been reported to double each year (eClinicalMedicine, 2024). It is estimated that by 2085, 50-60% of the world population will be at risk of dengue due to climate change scenarios, holding all risk factors constant (Hales et al., 2002). According to the European Centre for Disease Prevention and Control, dengue is the most significant mosquito-borne viral disease affecting humans globally. Each year, tens of millions of cases are reported, leading to an estimated 20 000 to 25 000 deaths, mainly among children (ECDC, 2023).

Dengue is classified as a neglected tropical disease (NTD), a group of diseases caused by pathogens and primarily affecting impoverished areas in tropical countries (WHO, 2023b). *Aedes* mosquitoes breed in small water bodies – even as small as containers, car tyres, etc., in and around houses. This allows the virus to be easily transmitted and spread among and within communities in urban locations despite the relatively limited flight range of the vector. Its complex epidemiology presents significant challenges for public health control. It is important to understand and identify risk factors of dengue, in order to inform public health policy aimed at controlling disease transmission and predicting future outbreaks.

The pathogen of dengue is transmitted between human hosts by mosquito vectors; thus, dengue is considered a vector-borne disease and is also classified as an *indirectly transmitted disease* (McMichael, 2003). Since pathogens for indirectly transmitted diseases exist in the external environment during their life cycles, this class of dis-

eases is susceptible to climatic factors. The association between dengue and climate variables, particularly temperature, rainfall, and relative humidity, has been extensively studied in the literature (Abdullah et al., 2022; Colón-González et al., 2021; Couper et al., 2021; Murray et al., 2013; Naish et al., 2014; Xu et al., 2017); details are presented in Section 5.3.

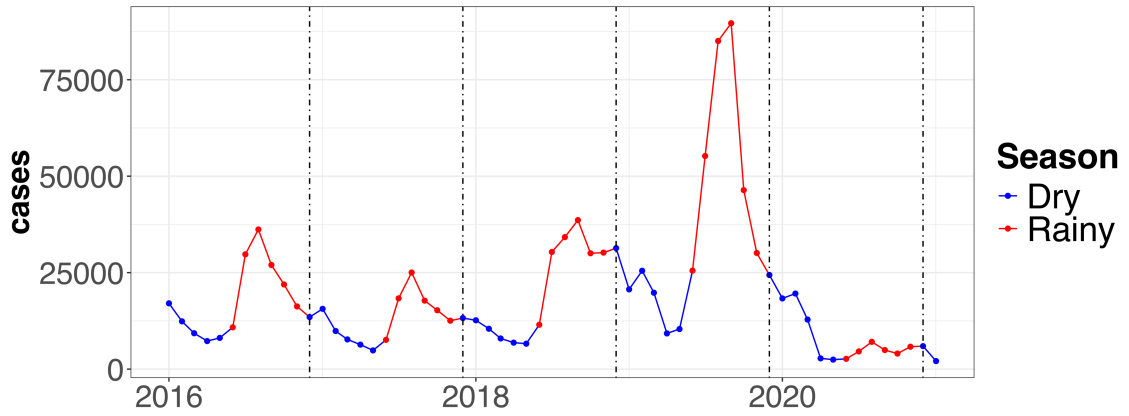


Figure 1.1: Time series plot of the number of dengue cases in the Philippines from January 2016 to January 2021

Figure 1.1 shows a plot of the monthly number of dengue cases in the Philippines from January 2016 to January 2021. The plot shows seasonality in dengue incidence, where the cases are generally higher during rainy season - which corresponds to months June to November (PAGASA, 2023). Around July 2019, the country declared a national dengue alert and dengue epidemic due to a surge in cases and deaths (BBC, 2019). Thus, it is noticeable from the plot that there is an unusually high number of cases around August to October 2019. In fact, for the entire year 2019, the number of cases was higher compared to the previous years. Another important observation is that during 2020, which is the start of the COVID-19 pandemic, the number of reported cases is very low. The decline in cases was a global phenomenon due to the pandemic and the low reporting rate (WHO, 2023a). This can be explained by two primary reasons. Firstly, there was a reduced mobility of the population and several studies have shown that a reduced household movement is associated with a reduced transmission (Stoddard et al., 2013). Secondly, this was a result of reporting hesitancy because of the fear of contracting the COVID-19 virus when visiting a health center or a hospital (Seposo, 2021). Moving forward, in 2023, there was an upsurge in the dengue cases globally, with a simultaneous occurrence of multiple outbreaks even in

regions previously unaffected by dengue (WHO, 2023a).

Figure 1.2 shows a map of annual (2016 to 2020) total number of dengue cases in the country. It is apparent that 2019 had the most number of cases. It also shows specific areas with the most recorded cases. In particular, the region badly hit by the 2019 dengue epidemic is the island in the western central part of the country, which is referred to as the Western Visayas region. Additional areas that can be identified in the plot are several contiguous provinces in the north, referred to as the Calabarzon region, and several areas located in the south. These were specific areas identified by the Health Department of the country which needed immediate emergency attention (BBC, 2019).

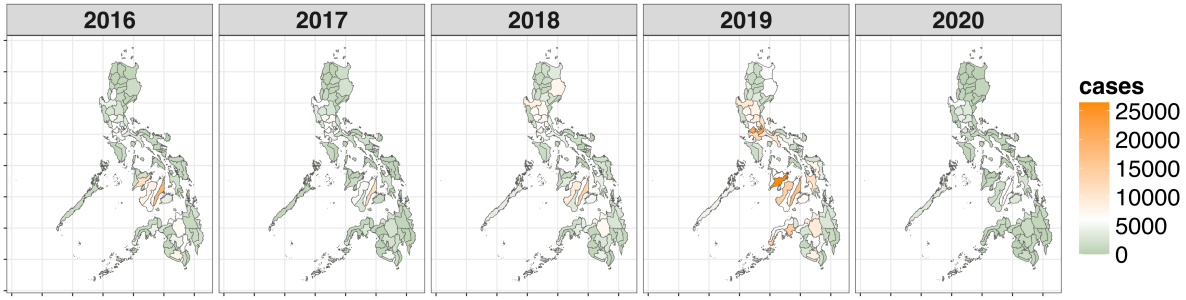
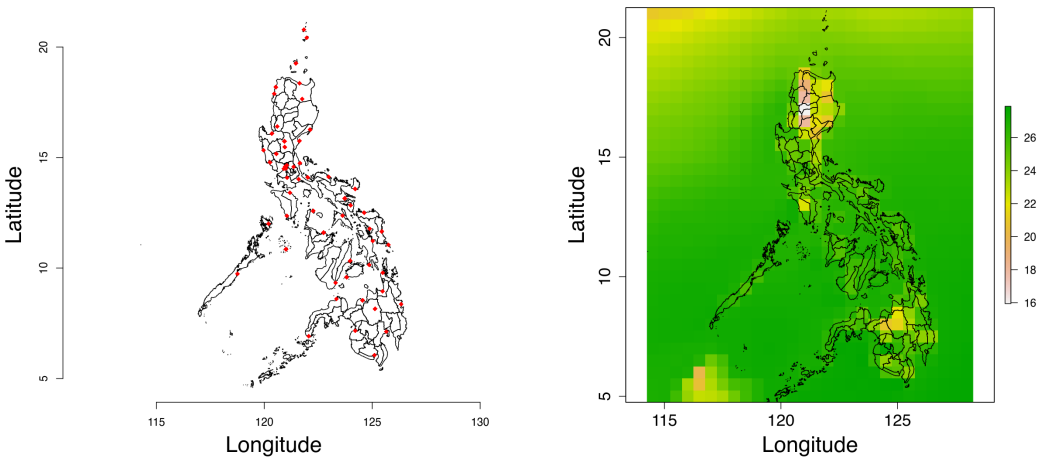


Figure 1.2: Plot of total dengue cases yearly (2016 – 2020) in the Philippines

Furthermore, Figure 1.3a shows the location of the weather synoptic stations in the Philippines. There are 57 stations which is very sparse relative to the size of the archipelago. Figure 1.3b shows an output of the *Global Spectral Model* (GSM), a weather forecast model maintained by the Japan Meteorological Agency. They provide forecast outputs of up to 132 hours four times a day (with initial times 0000, 0600, 1200, and 1800 UTC) within 4 hours of the initial time, and up to 264 hours twice a day (with initial time 0000 and 1200 UTC) within 7 hours of the initial time. The data from the weather forecast model have a very wide spatial coverage and are also observed at higher frequency in time.

The questions of interest are the following:

- Can observations from weather synoptic stations and simulated outcomes from the weather forecast model be combined to produce more accurate predictions of the latent process of interest—such as the true temperature field—compared to approaches that rely solely on data from the stations?



(a) Weather synoptic stations

(b) Sample GSM outcomes

Figure 1.3: Climate data sources for the Philippines: (a) 57 weather synoptic stations (b) Global Spectral Model (GSM), a numerical weather prediction model maintained by the Japan Meteorological Agency

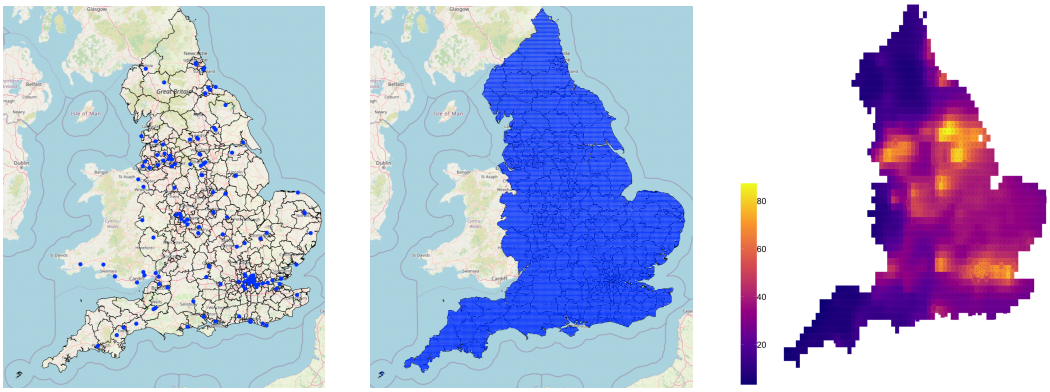
- Can we provide a link, using statistical approaches, between the climate variables and dengue? What is the impact of temperature, relative humidity, and rainfall on dengue incidence?
- How do we correctly propagate the uncertainty in the climate models into the health model?

1.2.2 Air pollution and respiratory diseases

Another relevant application is linking air pollution and respiratory diseases. In the United Kingdom, air pollution remains a public health problem and is known to cause premature deaths (Lee et al., 2017). The link between air pollution and respiratory diseases is well-established in the literature (Nascimento et al., 2016; Simkovich et al., 2019; Tran et al., 2023). A commonly used approach to perform analysis is to use spatio-temporal areal unit studies which use population-level data (Greven et al., 2011; Lee et al., 2009, 2017).

Figure 1.4a shows the locations of the stations under the Automatic Urban and Rural Network (AURN) in England and Wales. These stations record the concentration of certain pollutants such as nitrogen dioxide (NO_2), ozone (O_3), and particles less than 10 μm (PM_{10}) and 2.5 μm ($PM_{2.5}$). In addition to data from stations, dispersion models – which are based on deterministic differential equations that simulate the spread of pollutants in the atmosphere – are also commonly utilized. An

example is the Air Quality Unified Model (AQUM), which is a weather and chemical transport model that provides hourly estimates of pollutant concentrations in a 12 km^2 grid all over England (Lee et al., 2017). Figure 1.4b shows the locations from where the output of the numerical model is simulated. It obviously has a very wide spatial coverage compared to the network of monitoring stations. Figure 1.4c shows a sample outcome of the numerical model for NO_2 concentration for the month of January 2007.



(a) Monitors in England and Wales **(b)** AQUM data locations **(c)** Sample AQUM Data

Figure 1.4: Air pollution data sources for England

The questions of interest are the following:

- How can we combine the AURN and AQUM data to yield more accurate estimates and predictions of the true concentration of certain air pollutants in England, compared with relying solely on AURN data?
- Can we provide a link, using statistical approaches, between health outcomes, which are usually observed as counts at the level of Local Unitary Authorities, and air pollution concentration, which are point-referenced?
- How do we correctly propagate the uncertainty in the air pollution model into the health model?

1.3 Two-stage modelling framework

A two-stage modelling framework is widely used in different areas of statistical modelling, such as longitudinal data analysis, survival analysis, and spatial statistics. For instance, in survival analysis, it is common practice to first model longitudinal biomarkers, which are biological characteristics or medical signs – such as blood sugar level and cholesterol levels – and subsequently use the estimated trends of these biomarkers as inputs in a survival model (Rustand et al., 2024; Ye et al., 2008). In the area of spatial statistics, a two-stage framework is often used to address a spatial misalignment problem, e.g., when the response variable and covariates have different spatial supports (Gryparis et al., 2009; Szpiro et al., 2011). This is discussed in more detail in the context of spatial epidemiology in Section 1.3.1.

1.3.1 Two-stage modelling in spatial epidemiology

An example in spatial epidemiology (Blangiardo et al., 2016; Cameletti et al., 2019; Lee et al., 2017; Liu et al., 2017) is shown in Figure 1.5, where the objective is to understand the link between case counts of a disease and exposure variables, such as pollution and toxin levels, and meteorological variables. The case counts are observed in areas/blocks/polygons, while the exposure variable is a spatially continuous phenomenon, measured at a finite number of spatial locations (e.g. weather stations, monitoring stations). The first step involves fitting a spatial model (*first-stage model*) that is used to predict the exposure surface on a fine grid. Given the predictions on the grid, spatial averages are then computed over blocks consistent with the spatial support of the health data. The next step is to fit another spatial model (*second-stage model*) to link the health outcome variable and the block-level estimates of exposures. A simple formulation of this two-stage model is as follows:

$$\textbf{First stage}: \quad x(\mathbf{s}) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}) + \xi(\mathbf{s}) \tag{1.1}$$

$$\mathrm{w}(\mathbf{s}_i) = x(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad \epsilon(\mathbf{s}_i) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2\big) \quad i = 1, \ldots, n_{\mathrm{w}} \tag{1.2}$$

$$\textbf{Second stage}: \quad \mathrm{y}(B_j)\big|\mu_{B_j} \overset{\text{iid}}{\sim} \text{Poisson}\big(\mu_{B_j}\big), \quad \mu_{B_j} = \mathbb{E}\big[\mathrm{y}(B_j)\big] \tag{1.3}$$

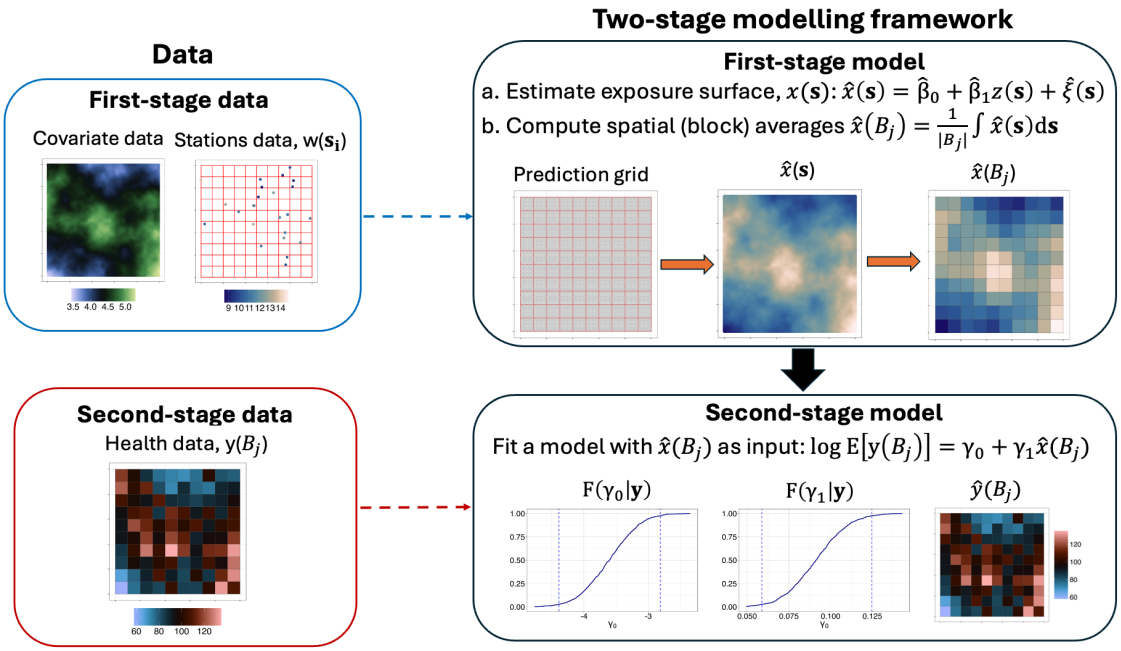


Figure 1.5: A two-stage modelling framework in spatial epidemiology: linking health outcomes, such as areal data on case counts of a disease, and pollution and/or meteorological variables observed at finite number of spatial locations or stations. *First stage*: fitting a spatial model for the true exposure surface. *Second stage*: fitting the health model.

$$\log\left(\mu_{B_j}\right) = \gamma_0 + \gamma_1 x(B_j) \tag{1.4}$$

$$x(B_j) = \frac{1}{|B_j|}\int_{B_j} x(\mathbf{s})d\mathbf{s}, \quad j = 1, \ldots, n_{\mathrm{y}} \tag{1.5}$$

where $x(\mathbf{s})$ is the first-stage latent process of interest, e.g., true temperature field or true pollution concentration field, $\mathrm{z}(\mathbf{s})$ is a known covariate and is available for the whole spatial domain, $\xi(\mathbf{s})$ is a random (spatial) field to account for extra variation in the data unexplained by $\mathrm{z}(\mathbf{s})$, $\mathrm{w}(\mathbf{s}_i)$ is the observed value of $x(\cdot)$ at a spatial location $\mathbf{s}_i$, $\epsilon(\mathbf{s}_i)$ is a measurement error term, and $\mathrm{y}(B_j)$ denotes the observed count of the disease in block B_j which is assumed to follow the Poisson distribution. The above model assumes that the health outcomes $\mathrm{y}(B_j)$ are linked to $x(\mathbf{s})$ via the spatial averages $x(B_j) = \frac{1}{|B_j|}\int_{B_j} x(\mathbf{s})d\mathbf{s}$, where $|B_j|$ is the size of block B_j (Equation (1.5)). The model unknowns are γ_0, γ_1, β_0, β_1, σ^2, and the parameters in the assumed model for $\xi(\mathbf{s})$.

The first stage would fit Equations (1.1) and (1.2) in order to obtain the estimated values of $x(\mathbf{s})$, say, $\hat{x}(\mathbf{s})$. This allows us to obtain the spatial averages $\hat{x}(B_j)$, defined

as

$$\hat{x}(B_j) = \frac{1}{|B_j|}\int_{B_j} \hat{x}(\mathbf{s})d\mathbf{s} = \frac{1}{|B_j|}\int_{B_j} \Big(\hat{\beta}_0 + \hat{\beta}_1 \mathrm{z}(\mathbf{s}) + \hat{\xi}(\mathbf{s})\Big)d\mathbf{s}, \tag{1.6}$$

where $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\xi}(\mathbf{s})$ are estimated values, say, the posterior means. The second stage then fits the model where $\hat{x}(B_j)$ is an input in Equation (1.4) in order to obtain the posterior estimates of γ_0 and γ_1.

One might wonder, why we do not fit Equations (1.1) – (1.5) simultaneously?. Firstly, fitting the above model simultaneously, also referred to as a *joint modelling approach* or *fully Bayesian approach*, can be computationally challenging and expensive (Gryparis et al., 2009; Liu et al., 2017). A joint modelling approach may not be a practical approach especially with the increase in the volume of available and accessible data nowadays, which implies that fitting the first-stage model can be complex in itself. Secondly, given that the first-stage model is already computationally demanding, doing multiple health effect analyses or running multiple candidate epidemiological models, which require refitting the first-stage model several times in a joint modelling framework, will be computationally expensive (Blangiardo et al., 2016; Liu et al., 2017). Thirdly, a potential problem with a joint modelling approach is that it could cause potential 'feedback' effects wherein the data $\mathrm{y}(B_j)$ influence and distort the model for $x(\mathbf{s})$, which consequently may compromise the $\Big(x(B_j), \mathrm{y}(B_j)\Big)$ relationship (Gryparis et al., 2009; Shaddick and Wakefield, 2002; Wakefield and Shaddick, 2006). This could happen when the data to inform about $x(\mathbf{s})$ are sparse (Gryparis et al., 2009) or due to model misspecification (Yucel and Zaslavsky, 2005). One way to 'cut' the feedback between the two stages is by introducing a cut function in an MCMC algorithm (Plummer, 2015). The cut function essentially simplifies the full conditional distribution of a graphical model into smaller modules that interact more weakly than in a full Bayesian analysis (Bayarri et al., 2009; Spiegelhalter et al., 2003). However, this approach may not converge to a well-defined limiting distribution unless tempered transitions are introduced (Plummer, 2015). Also, it is well-known to be difficult to implement and computationally expensive (Chakraborty et al., 2023). Fourthly, performing a two-stage modelling approach has an intuitive physical interpretation since there is a clear one-directional relationship between $x(\cdot)$

and y(·), e.g., climate and pollution levels affect disease risks but not the other way around.

Although a two-stage modelling framework is simpler, since it breaks down the problem into two separate but connected stages, there is one important issue that needs to be addressed: the problem of uncertainty propagation, which is elaborated in Section 1.3.1.1. The uncertainty propagation is intrinsic in a joint modelling approach, but not in a two-stage modelling framework and hence needs to be specifically addressed here.

1.3.1.1 Uncertainty propagation problem

When adopting a two-stage modelling approach, it is important to account for the uncertainty in the first-stage model results when fitting the second-stage model. This is referred to as the uncertainty propagation problem. The problem is due to the fact that the predicted values $\hat{x}(B_j)$, shown in Equation (1.6), from the first-stage model are subject to some uncertainty due to estimation error or model misspecification error. This uncertainty must be considered and correctly propagated from the first stage to the second stage. The problem of uncertainty propagation in a two-stage modelling framework is formally discussed in Chapter 6

There are two existing and commonly used approaches to performing two-stage Bayesian modelling: 1) a crude plug-in method which does not account for the uncertainty in $\hat{x}(B_j)$, and 2) the posterior sampling approach (Blangiardo et al., 2016; Cameletti et al., 2019; Lee et al., 2017; Liu et al., 2017). A crude plug-in method, which simply plugs-in the posterior means from the first-stage model into the second-stage model, would potentially underestimate the true posterior uncertainty of the second-stage model parameters; on the other hand, the posterior sampling approach does account for the first-stage model uncertainty but can be computationally expensive. The posterior sampling approach generates several samples from the first-stage model posteriors, and then each sample is used as an input to the second-stage model. The final posterior estimates of the second-stage model parameters are then computed using Bayesian model averaging.

One of the goals of this work is, therefore, to validate the correctness of the two

aforementioned approaches. Another goal is to propose a new method for uncertainty propagation which does not do resampling, and hence is potentially more computationally efficient. This method is called the **Q** *uncertainty method*, and is introduced in Chapter 6.

1.3.1.2 Combining multiple data sources $x(\mathbf{s})$

Another aspect of the two-stage modelling framework that this work investigates is the use of several data sources to estimate an unknown field $x(\mathbf{s})$ (see Equation (1.1)). The primary data source to estimate the field $x(\mathbf{s})$ are direct measurements at a finite number of point locations. In the context of environmental sciences and meteorology in particular, data on e.g. air quality, environmental pollution, surface temperature, or rainfall are collected through a network of monitoring stations and weather stations (Arab et al., 2014; Blangiardo et al., 2016; Chien and Yu, 2014; Greven et al., 2011; Jaya and Folmer, 2022; Lawson et al., 2016; Lee et al., 2015, 2017). These data are used for prediction, and to improve the understanding of the spatio-temporal dynamics of the underlying processes and the impact of these variables on potential outcomes of interest, such as health outcomes. However, due to high maintenance costs, the monitoring networks are typically spatially sparse (Lawson et al., 2016). Increasingly, data from additional sources derived from satellite images or outcomes of numerical models with a high spatial resolution are available. These can be used jointly with the stations data to improve the accuracy of predictions in a process that combines information from different data sources and is often referred to as *data fusion* or *data assimilation* (Bauer et al., 2015; Gettelman et al., 2022; Lawson et al., 2016). The goal is to exploit the better spatial resolution of the additional data to fill gaps in areas only sparsely covered by the stations in order to predict and map the variables into space with more accuracy at smaller scales than based on the stations data alone. However, a general issue that is common in attempts to combine data from more than one source is that the various data streams differ in their quality. Satellite data and outcomes of numerical models, specifically, are often biased due to calibration issues, and these biases have to be accounted for in the modelling process. Moreover, while the data from the stations are point-referenced, simulated outcomes

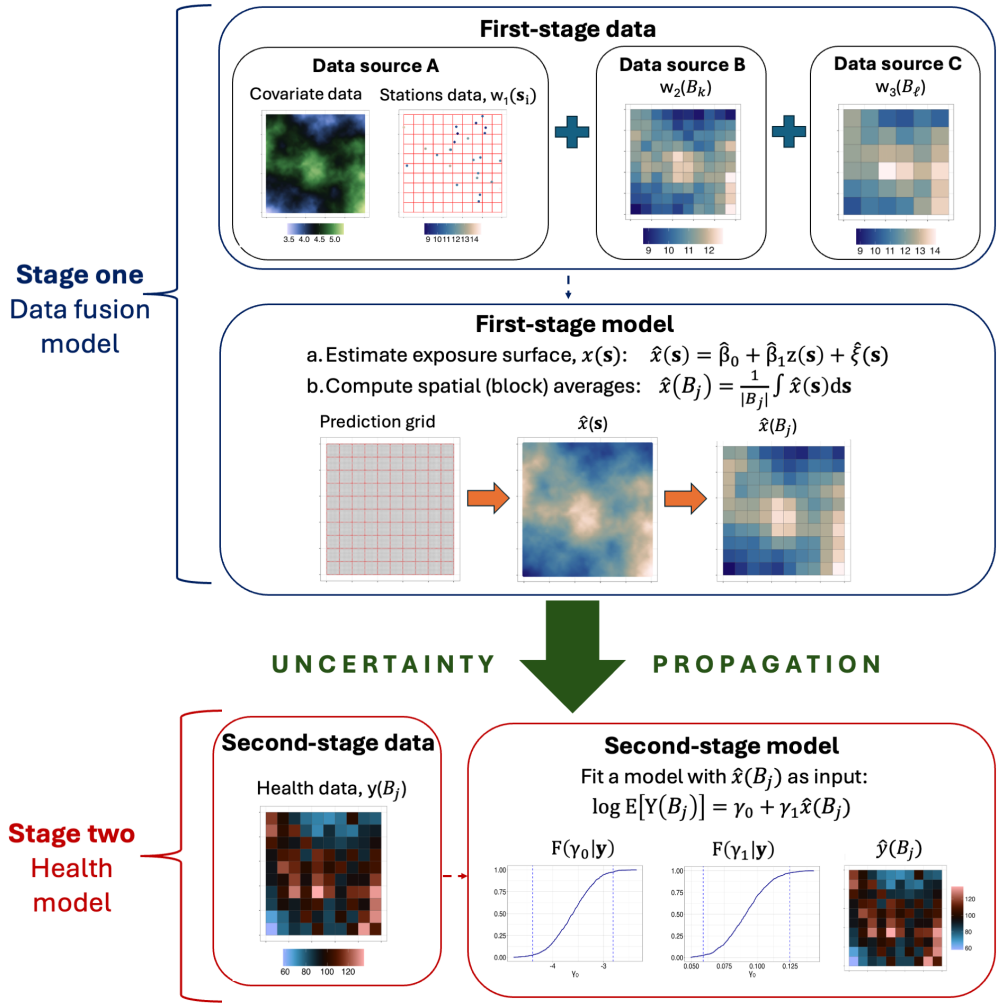


Figure 1.6: Extended two-stage modelling framework to emphasize two things: data fusion challenge in the first stage, and the uncertainty propagation from the first-stage model to the second-stage model.

from numerical models or remote-sensed data are considered areal data on regular grids (Bruno et al., 2016). Thus, in order to optimize the integration of different data sources, two important aspects need to be considered: the flexibility of the model to account for the disparity in the accuracy levels of the different data sources, and the efficiency of the model to overcome the problem of spatial misalignment.

Figure 1.6 shows an extension of Figure 1.5 to account for the two aforementioned vital components of the two-stage modelling framework that this work investigates. Firstly, the proposed framework specifies a data fusion model in the first stage to obtain more accurate predictions of the latent field of interest $x(\mathbf{s})$. Figure 1.6 assumes three data sources for the first-stage model: the point-referenced data (source A), such as stations data, and two additional data sources which are areal and of different resolutions (sources B and C). Secondly, it emphasizes the need to propagate the uncertainty from the first-stage model to the second-stage model. The first-stage

model follows Equations (1.1) – (1.2), while the second-stage model follows Equations (1.4) – (1.5), similar to Figure 1.5.

1.4 Research Gaps

The following is a summary of the challenges and/or gaps in the literature that this thesis aims to address.

1. Firstly, this work extends the proposed model in Cameletti et al. (2019) which addresses a spatial misalignment problem to link health and air pollution data. The model in Cameletti et al. (2019) uses a two-stage modelling framework to link health and air pollution. This thesis then extends their approach by incorporating a data fusion model in the first-stage model. This is a more realistic and potentially better model since in many real-life scenarios, the data coming from stations are typically sparse, and we want to use proxy data, such as outcomes from numerical models and/or remote-sensed data, to improve model accuracy. This is addressed in Chapter 3.

2. This thesis adopts the Bayesian melding model, which is an approach to perform data fusion and is formally presented in Section 2.7.1 of Chapter 2. It assumes that all observed data are a function of the same latent process. In many recent studies that adopted the same idea, such as Moraga et al. (2017), Zhong and Moraga (2023), and Forlani et al. (2020), calibration biases for the proxy data are not accounted for. Accounting for these biases, particularly both an additive and multiplicative bias, is important since there are more data coming from the biased proxy data and we want to avoid these data to dominate parameter estimation. This work uses the Bayesian model averaging with INLA as the estimation approach since it provides numerical stability. All materials and results which tries to address the challenges and gaps in this research area are shown in Chapter 4.

3. The next research gap this work tries to address is the validation of commonly used methods for doing two-stage modelling, particularly a plug-in method and

posterior resampling method. These two methods are used to link dengue and climate in the Philippines, which are presented in Chapter 5. The two aforementioned methods are formally discussed in Chapter 6. Both are validated using the simulation-based calibration (SBC) method, which is a method to test for the self-consistency property of Bayesian models. To the best of my knowledge, no published work has applied this approach to validate uncertainty propagation methods in two-stage modelling applications. A formal discussion of the SBC method is also presented in Chapter 6 .

4. This study addresses another gap in the literature by proposing a novel method for uncertainty propagation in two-stage models, called the **Q** uncertainty method. The new approach tries to address the limitations of the two baseline methods. Unlike the plug-in method which ignores the posterior uncertainty of the first-stage model, the proposed method accounts for the first-stage model uncertainty via the **Q** matrix. This matrix is then incorporated in the predictor expression of the second-stage model. Moreover, unlike the posterior sampling method which fits the second-stage model multiple times, the proposed method fits the second-stage model once, and can therefore be more computationally efficient.

5. Finally, a research gap this thesis addresses is to provide additional empirical evidence in the literature on the link between climate and dengue in the Philippines. I use a novel Bayesian spatio-temporal model that incorporates a complex specification for structured and unstructured effects in space and time, including their interactions. The model specifies a spatial effect which accounts for the archipelagic geography of the country. Moreover, the climate predictions used as input to the dengue model are based on the climate data fusion models presented in Chapter 4. This extends much of the existing work in the literature, particularly for the Philippines, where analyses have primarily relied on sparse data from weather stations. All the details of the methodology and results are presented in Chapter 5.

1.5 Objectives of the thesis

The objectives of this PhD work are as follows:

1. The first primary objective is to propose a data fusion approach in a two-stage spatio-temporal model. To achieve this objective, I start with the case of a simple measurement error process and investigate the performance of the integrated nested Laplace approximation (INLA) and the stochastic partial differential equations (SPDE) approach through a simulation study. The first objective is an initial exploration of doing data fusion in the context of two-stage spatio-temporal modelling. The focus is on exploring the capabilities of the INLA-SPDE method to fit the proposed data fusion model. Here, I do not intend to compare the performance of the data fusion model with benchmark models, such as a stations-only model. This will be dealt with in Objective (2). The issue of uncertainty propagation is not yet formally dealt with as well. In order to account for the uncertainty in the first-stage model in this initial stage of the work, I use the posterior sampling approach. All relevant materials and results addressing Objective (1) are provided in Chapter 3.

2. The second primary objective is to extend the proposed data fusion model addressed in Objective (1) to provide flexibility in accounting for the biases in the different data sources. The following are the specific goals:

    - To propose a flexible data fusion model extending the capabilities of the proposed data fusion model in Objective (1)
    - To conduct a simulation study to compare the performance of the following approaches: a stations-only model, a regression (statistical) calibration model, and the proposed data fusion model
    - To apply the proposed model on the meteorological data in the Philippines (motivating example in Section 1.2.1)
    - To compare the performance of the proposed data fusion model and the two aforementioned benchmark approaches in the data application using the leave-group-out cross-validation (LGOCV)

The second primary objective also focuses on the data fusion problem, i.e., the first-stage model of the two-stage modelling framework presented in Section 1.3. Here, I perform a proper comparison between the proposed method and benchmark approaches. I then present an extensive analysis on a data application motivated by meteorological data in the Philippines. All relevant materials and results addressing Objective (2) are presented and discussed in Chapter 4.

3. The third primary objective is to link climate and dengue in the Philippines (see motivating example in Section 1.2.1). In particular, I investigate three climate variables: temperature, relative humidity, and rainfall. Here, I explore both a stations-only model and a data fusion model as input in the health (second-stage) model. In order to account for the uncertainty in the first-stage models, I use the posterior sampling approach, which is also used in Objective (1). All materials and results for this chapter are presented and discussed in Chapter 5. This chapter does not propose any methodological innovation; rather it provides an extensive case study, which showcases the proposed framework in Figure 1.6.

4. The fourth primary objective is to evaluate the correctness of commonly used two-stage modelling approaches and to propose a new method to do uncertainty propagation in two-stage Bayesian hierarchical models. In particular, the following are the specific objectives:

   (a) To evaluate the correctness of two existing and commonly used approaches for doing two-stage modelling: a crude plug-in approach and posterior sampling approach

   (b) To propose a variation in the implementation of the simulation-based calibration approach, which is the approach used to perform the model validation in (a)

   (c) To propose a new approach for doing uncertainty propagation in two-stage Bayesian hierarchical models

   (d) To propose a low-rank approximation of the proposed method in (c)

   (e) To demonstrate a comparison of the different uncertainty propagation ap-

proaches on toy two-stage spatial models

(f) To illustrate the proposed methods on a simple spatial model to link climate and dengue in the Philippines (motivating example in Section 1.2.1)

The fourth main objective of this research focuses on the uncertainty propagation problem presented in Section 1.3.1.1 (also see Figure 1.6). To validate the different approaches for two-stage modelling, I use the simulation-based calibration (SBC) approach, which is an approach for testing the self-consistency property of Bayesian models. This chapter proposes a methodological innovation in doing two-stage modelling. The proposed methods are also validated using the SBC method. All relevant results and materials which investigates Objective (4) are presented and discussed in Chapter 6.

Chapter 2

Statistical methods

This chapter presents statistical concepts and methods that are relevant in this PhD thesis. Section 2.1 presents the problem of spatial misalignment, and a related problem called the *change of support problem* (COSP). Section 2.2 discusses classical models for point-referenced spatial data, while Section 2.3 discusses some classical models for areal data. The former is relevant for the first-stage model in Figure 1.6, while the latter is relevant for the second-stage model. The generalized linear model (GLM) or generalized linear mixed model (GLMM) specification of the spatial models are discussed in Section 2.4. As mentioned in Section 1.5, I perform Bayesian inference on the models in this PhD thesis, particularly using the integrated nested Laplace approximation (INLA) approach. Section 2.5 discusses the INLA methodology, both the classical and modern INLA, which are presented in Sections 2.5.2 and 2.5.3, respectively. Another important method is the stochastic partial differential equations (SPDE) method, which I use to estimate Gaussian fields of the Matérn type. This is discussed in Section 2.6. Finally, some data fusion approaches are discussed in Section 2.7.

2.1 Problem of spatial misalignment

Chapter 1 introduced the two-stage modelling framework for this PhD thesis. A motivation for doing this is that the spatial data for analysis have different spatial supports, i.e., they are spatially misaligned. This was discussed in Section 1.3.1 and

was clearly seen in the motivating examples in Sections 1.2.1 and 1.2.2. In particular, the response variable in the first stage is point-referenced while the response variable in the second stage is areal. The point-referenced data, denoted by $\mathrm{w}(\mathbf{s}_i), i = 1, ..., n_{\mathrm{w}}$ are assumed to follow a latent stochastic process and are observed at finite locations $\{\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n\}$, while the areal data $\mathrm{y}(\cdot)$ are observed at the blocks/polygons $B_j, j = 1, ..., n_{\mathrm{y}}$. This scenario is just one of the various types/examples of spatially misaligned data.

The problem of spatial misalignment necessitates doing inference and/or prediction on a spatial scale different from that of the observed data. This problem is referred to as the *change of support problem* (COSP) in spatial statistics (Gelfand et al., 2010; Gotway and Young, 2002). The term *spatial support* refers to the geometrical size, shape, volume, and orientation of the units or regions associated with the measurements (Schabenberger and Gotway, 2017). If the data are point-referenced, and the target spatial support are also points, e.g., a prediction grid, this is referred to as *kriging* (Cressie, 1988; Montero et al., 2015). If the data are point-referenced while the target spatial support consists of areas/block/polygons, this is referred to as *block kriging* or *up-scaling*. This is discussed further in Section 2.2.5, since this type of COSP is relevant in this PhD work. Additional types and examples of COSP are presented in Gotway and Young (2002).

Suppose that $x(\mathbf{s})$ is the unknown first-stage field. We wish to upscale $x(\mathbf{s})$ to $x(B_j)$, which represents the block-level value of $x(\mathbf{s})$. This is relevant in this application, since one way to perform the two-stage modelling, as shown in Figure 1.6, is to estimate $x(B_j)$, and then plug-in that value into the second-stage model. $x(B_j)$ is related to $x(\mathbf{s})$, but has different statistical and spatial properties. An intuitive way to upscale is via spatial averaging. The spatial averages of the process are meaningful if $x(\mathbf{s})$ is an *intensive* spatial quantity. Intensive spatial variables are variables that do not have values proportional to the spatial support (Pebesma and Bivand, 2023). This means that if an area is split into smaller areas, the values of the variable are not split similarly, i.e., values may vary within but on average remain the same. Examples of intensive variables are temperature, air pollution concentration, and population density. On the other hand, variables whose values are associated with

a physical size are called *extensive* variables. An example of an extensive variable is population count, since if an area is split into smaller areas, the population count needs to be split too. The split of the values is not necessarily done proportional to the sub-areas since population is rarely uniform over space, but the sum of the population count across all subareas need to equal that of the total.

In this PhD thesis's motivating examples (see Section 1.2), $x(\mathbf{s})$ refers to climate fields and air pollution concentration fields, which are intensive spatial variables. Thus, the up-scaled value $x(B_j)$, which is the spatial average of $x(\mathbf{s})$, is given by

$$x(B_j) = \int_{B_j} \lambda(\mathbf{s})\mu(\mathbf{s})d\mathbf{s}, \tag{2.1}$$

where $\lambda(\mathbf{s})$ is an intensity function which weights the values of $\mu(\mathbf{s})$, and such that $\int_{B_j} \lambda(\mathbf{s})d\mathbf{s} = 1$ (Gelfand et al., 2010; Pebesma and Bivand, 2023).

2.2 Classical models for geostatistical data

This section presents classical methods for geostatistical or point-referenced data. This is relevant for the first-stage model in the framework presented in Figure 1.6.

Suppose $\mathbf{w} = \begin{pmatrix} \mathrm{w}(\mathbf{s}_1) & \mathrm{w}(\mathbf{s}_2) & \cdots & \mathrm{w}(\mathbf{s}_{n_\mathrm{w}}) \end{pmatrix}^\intercal, \mathbf{s} \in \mathbb{R}^d$, is the observed data for a continuously-indexed spatial process. The process is observed at a finite set of locations $\{\mathbf{s}_1, \ldots, \mathbf{s}_{n_\mathrm{w}}\}$. A commonly assumed model structure for $\mathrm{w}(\mathbf{s}_i)$ is given by

$$\mathrm{w}(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \xi(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad i = 1, \ldots, n_\mathrm{w} \tag{2.2}$$

where $\mu(\mathbf{s}_i)$ is a large-scale trend, $\xi(\mathbf{s}_i)$ is a random small-scale variation, and $\epsilon(\mathbf{s}_i)$ is a measurement error. Cressie (2015) proposed an additional term to account for *micro-scale* variation in $\mathrm{w}(\mathbf{s}_i)$. But in this review, we focus on the model specification in Equation (2.2). Also, note that Equations (1.1) and (2.2) are equivalent when $x(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \xi(\mathbf{s}_i)$.

The large-scale trend $\mu(\mathbf{s}_i)$ is assumed as a function of a fixed set of covariates $\mathbf{z}(\mathbf{s}_i) = \begin{pmatrix} \mathrm{z}(\mathbf{s}_{i1}) & \mathrm{z}(\mathbf{s}_{i2}) & \cdots \mathrm{z}(\mathbf{s}_{ip}) \end{pmatrix}^\intercal$ and a vector of parameters $\boldsymbol{\beta}$, particularly, $\mu(\mathbf{s}_i) = \mathbf{z}(\mathbf{s}_i)^\intercal\boldsymbol{\beta}$. The variation and covariation in $\mathrm{w}(\mathbf{s}_i)$ is explained by the stochastic

properties of $\xi(\mathbf{s}_i)$ and $\epsilon(\mathbf{s}_i)$.

$\xi(\mathbf{s}_i)$ is typically assumed a stationary process with a covariance function, say, $c_\xi(\cdot)$. These are formally defined in Section 2.2.1. Moreover, the measurement error $\epsilon(\mathbf{s}_i)$ is a white noise process with mean 0 and variance $\mathbb{V}[\epsilon(\mathbf{s}_i)] = \sigma^2_\epsilon$. $\xi(\mathbf{s})$ and $\epsilon(\mathbf{s}_i)$ are assumed independent. The sum $\mu(\mathbf{s}) + \xi(\mathbf{s})$ is referred to as the *signal*, and comprises the model terms which are spatially structured either in a deterministic or stochastic fashion (Schabenberger and Gotway, 2017). This is the main interest in applications of spatial prediction, not $\mathrm{w}(\mathbf{s}_i)$ which is the noisy version of the signal. On the other hand, the term $\xi(\mathbf{s}) + \epsilon(\mathbf{s})$ is sometimes referred to as the *error process* of the model. However, for some applications, $\xi(\mathbf{s})$ can be part of the mean process to account for local stochastic fluctuations of the process. This explains the famous adage 'one modeler's mean function is another modeler's covariance structure'.

Thus, a classical model specification for $\mathbf{w}$ is given by

$$\begin{aligned} \mathbf{w} &= \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\xi} + \boldsymbol{\epsilon} \\ \mathbb{E}[\mathbf{w}] &= \mathbf{Z}\boldsymbol{\beta} \\ \mathbb{V}[\mathbf{w}] &= \mathbb{V}[\boldsymbol{\xi} + \boldsymbol{\epsilon}] \equiv \boldsymbol{\Sigma}(\boldsymbol{\theta}) = \boldsymbol{\Sigma}_\xi(\boldsymbol{\theta}_\xi) + \sigma^2_\epsilon \mathbb{I}_{n_\mathrm{w}} \\ (\boldsymbol{\Sigma}_\xi)_{ij} &= c_\xi(\mathbf{s}_i, \mathbf{s}_j), \end{aligned} \tag{2.3}$$

where $(\boldsymbol{\Sigma}_\xi)_{ij}$ denotes the $(i,j)^{\text{th}}$ element of $\boldsymbol{\Sigma}_\xi(\boldsymbol{\theta}_\xi)$, $\boldsymbol{\theta}_\xi$ are the parameters of the covariance function $c_\xi(\cdot)$, $\mathbb{I}_{n_\mathrm{w}}$ is an identity matrix of dimension $n_\mathrm{w} \times n_\mathrm{w}$, $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_\xi & \sigma^2_\epsilon \end{pmatrix}^\intercal$, $\mathbf{Z}$ is the matrix of known covariates with i^{th} row as $\mathbf{z}(\mathbf{s}_i)$, $\boldsymbol{\xi} = \begin{pmatrix} \xi(\mathbf{s}_1) & \cdots & \xi(\mathbf{s}_{n_\mathrm{w}}) \end{pmatrix}^\intercal$, and $\boldsymbol{\epsilon} = \begin{pmatrix} \epsilon(\mathbf{s}_1) & \cdots & \epsilon(\mathbf{s}_{n_\mathrm{w}}) \end{pmatrix}^\intercal$.

2.2.1 Model specification for $\xi(\mathbf{s})$

The stationary assumption on $\xi(\mathbf{s}_i)$ facilitates and simplifies inference. A restrictive kind of stationarity, called *strict stationarity*, means that the (joint) probability distribution of $\begin{pmatrix} \xi(\mathbf{s}_1) & \xi(\mathbf{s}_2) & \cdots & \xi(\mathbf{s}_{n_\mathrm{w}}) \end{pmatrix}^\intercal$ is the same as $\begin{pmatrix} \xi(\mathbf{s}_1 + \mathbf{h}) & \xi(\mathbf{s}_2 + \mathbf{h}) & \cdots & \xi(\mathbf{s}_{n_\mathrm{w}} + \mathbf{h}) \end{pmatrix}^\intercal$, for any $\mathbf{h} \in \mathbb{R}^d$. A less restrictive kind, called *weak stationarity* or *second-order stationarity*, assumes that $\xi(\mathbf{s}_i)$ has a constant mean and that the covariance between

$\xi(\mathbf{s}_i)$ and $\xi(\mathbf{s}_i + \mathbf{h})$ only depends on $\mathbf{h}$, i.e., $\text{Cov}\big(\xi(\mathbf{s}_i), \xi(\mathbf{s}_i + \mathbf{h})\big) = c_\xi(\mathbf{h})$, for any $\mathbf{h} \in \mathbb{R}^d$. Second-order stationary implies that the covariance function is a function only of the difference or spatial lag between two locations.

A third kind of stationarity, which arose in the traditional kriging literature, is called *intrinsic stationarity* and is satisfied if $\mathbb{V}\big[\xi(\mathbf{s}_i + \mathbf{h}) - \xi(\mathbf{s}_i)\big] = 2\gamma(\mathbf{h})$. Intrinsic stationarity does not provide a probability model (likelihood) for the data (Banerjee et al., 2014). The function $\gamma(\mathbf{h})$ is called the *semivariogram* and is related to the covariance function $c_\xi(\cdot)$ via

$$c_\xi(\mathbf{h}) = c_\xi(\mathbf{0}) - \gamma(\mathbf{h}).$$

Given $c_\xi(\cdot)$, $\gamma(\cdot)$ can be easily recovered; hence, weakly stationarity implies intrinsic stationarity. However, the converse is not true, unless the $\lim_{\|\mathbf{h}\| \to \infty} \gamma(\mathbf{h})$ exists, where $\|\mathbf{h}\|$ is the length of $\mathbf{h}$.

Another common model assumption is that $c_\xi(\cdot)$ or $\gamma(\cdot)$ depends only on $\|\mathbf{h}\|$, which is referred to as *isotropy*. An extensive list of isotropic covariance functions and isotropic semivariograms, and their properties, are in Banerjee et al. (2014). Examples are the exponential, Gaussian, and Matérn covariance functions / semivariograms.

2.2.2 Estimation approaches

This section describes classical estimation approaches for the spatial model given in Equation (2.3). If $\boldsymbol{\theta}$ is known, the *generalized least squares estimator* for $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_{\text{gls}}$ is

$$\hat{\boldsymbol{\beta}}_{\text{gls}} = \left(\mathbf{Z}^\intercal \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{Z}\right)^{-1} \mathbf{Z}^\intercal \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1} \mathbf{w}.$$

If $\boldsymbol{\xi}$ and $\boldsymbol{\epsilon}$ are assumed to be Gaussian, then $\hat{\boldsymbol{\beta}}_{\text{gls}}$ is equivalent to the maximum likelihood estimator for $\boldsymbol{\beta}$. This implies that $\hat{\boldsymbol{\beta}}_{\text{gls}}$ is a consistent estimator for $\boldsymbol{\beta}$. For the non-Gaussian case, the consistency property is not guaranteed. Conditions to ensure the consistency of $\hat{\boldsymbol{\beta}}_{\text{gls}}$ are discussed in Schabenberger and Gotway (2017).

If $\boldsymbol{\theta}$ is unknown, there are two main approaches: an iterative reweighted generalized least squares approach, and a likelihood-based approach. The first one iteratively

estimates $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Given an initial estimate of $\boldsymbol{\beta}$, the $\boldsymbol{\theta}$ is estimated using the residuals $\mathbf{w} - \mathbf{Z}\boldsymbol{\beta}$. The 'updated' estimate of $\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\beta}}_{\text{egls}} = \left(\mathbf{Z}^\intercal \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})^{-1}\mathbf{Z}\right)^{-1}\mathbf{Z}^\intercal \boldsymbol{\Sigma}(\tilde{\boldsymbol{\theta}})\mathbf{Z},$$

where $\tilde{\boldsymbol{\theta}}$ is the estimated value for $\boldsymbol{\theta}$. The process repeats by estimating $\boldsymbol{\theta}$ using the 'updated' residuals. The iteration stops when there are minimal changes in the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. $\hat{\boldsymbol{\beta}}_{\text{egls}}$ is called the *estimated generalized least squares estimator* (EGLS) for $\boldsymbol{\beta}$. Under certain regularity conditions, $\hat{\boldsymbol{\beta}}_{\text{egls}}$ is a consistent estimator for $\boldsymbol{\beta}$. A main concern with this approach is the impact of using estimated covariance parameters in the plug-in expression for $\hat{\boldsymbol{\beta}}_{\text{egls}}$. This ignores extra uncertainty from estimating the covariance parameters and can lead to understimated standard errors.

The parameters $\boldsymbol{\theta}$ are estimated based on the assumed semivariogram or covariance function. A classical approach is to fit a parametric semivariogram model to the empirical semivariogram using least squares method. One way to obtain the empirical semivariogram is the *Matheron* estimator, which provides an unbiased empirical estimate and visualization of the semivariogram. However, this gives unstable estimates when data is sparse, and is highly sensitive to outliers. A robust but biased estimator is the *Cressie-Hawkins* estimator. Given the empirical semivariogram, the estimate for $\boldsymbol{\theta}$ is then derived using the least squares approach based on the assumed parametric model for $\gamma(\mathbf{h})$. The derivation of the (weighted) least squares estimator for $\gamma(\mathbf{h})$ is detailed in Cressie (1985). Another approach for estimating $\boldsymbol{\theta}$ is the use of generalized estimating equations and composite likelihood approaches, both of which can be viewed as generalizations of the least squares approach (Schabenberger and Gotway, 2017). Both approaches derive an unbiased score function as a function of $\boldsymbol{\theta}$, and do not rely on an exact likelihood model for the data, but nevertheless give a consistent estimator for $\boldsymbol{\theta}$ under certain regularity conditions.

The second approach maximizes the likelihood function given the distributional assumption of the data, which is typically Gaussian. The maximum likelihood approach simultaneously estimates $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$, unlike the EGLS estimator. Since $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is a nonlinear function of $\boldsymbol{\theta}$, then nonlinear optimization techniques, such as the Newton-

Rhapson or Quasi-Newton, are used to obtain maximum likelihood estimates. The size of the optimization problem is substantially reduced using *profiling* techniques. The MLE for $\boldsymbol{\beta}$ takes the same form as $\hat{\boldsymbol{\beta}}_{\text{egls}}$, except that $\boldsymbol{\theta}$ is evaluated at its MLE. An advantage of this approach is that it provides the variance-covariance matrix of the parameter estimates based on the Hessian matrix. The drawback of this approach is that the MLE for $\boldsymbol{\theta}$ is biased. The bias arises from a loss of degrees of freedom due to the simultaneous estimation of $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. Correction for bias is done using the *restricted maximum likelihood estimation* (REML) technique. A disadvantage of both the ML and REML technique is that the computation of the standard errors for $\hat{\boldsymbol{\beta}}$ involves plugging in the estimated covariance parameters.

2.2.3 Spatial prediction

The terms *prediction* and *estimation* are distinct concepts, although both are related. The former is used when the goal is to determine the value of a random quantity, while the latter is used to determine the value of a fixed quantity. In the classical geostatistical model in Equation (2.3), the goal is either to estimate $\mathbb{E}[\mathrm{w}(\mathbf{s}_0)]$ or to predict $\mathrm{w}(\mathbf{s}_0)$, where $\mathbf{s}_0 \in \mathbb{R}^2$ denotes a new point location of interest. This section presents classical approaches for doing spatial prediction.

The derivation of the optimal approach for prediction starts with specifying a loss function. The most commonly used loss function is the squared-error loss function, since statistical properties are easily examined in such scenarios. An important result is that the predictor which minimizes the expected squared error loss, also termed the *mean-squared prediction error*, is the conditional expectation of $\mathrm{w}(\mathbf{s}_0)$ given the observed data. In the Gaussian case, the conditional expectation and conditional variance, which quantifies the uncertainty in the prediction, have closed-form expressions. This framework for doing spatial prediction is referred to as *kriging*.

In traditional kriging, the derivation of the optimal predictor usually restricts to the class of unbiased and linear predictors; hence, the optimal predictor is usually called *best linear unbiased predictor* (BLUP). When the mean of $\mathrm{w}(\cdot)$ is known, the optimal linear predictor, under squared-error loss, is called the *simple kriging* predictor. With the Gaussian assumption on the data, the simple kriging predictor is

equivalent to the conditional mean of a Gaussian random field; hence, it is the best predictor in the class of both linear and non-linear estimators (under squared-error loss). If the data is not Gaussian, the simple kriging predictor is only the best in the class of linear functions of the data.

When the mean of $w(\cdot)$ is unknown, one typically starts by estimating the mean using ordinary least squares, and then perform simple kriging on the residuals. However, this approach underestimates the uncertainty in the predictions, since the uncertainty from estimating the mean in the first stage is not taken into account in the second stage. The predictor is unbiased but is not the BLUP.

Moreover, when the mean of $w(\cdot)$ is unknown but constant over space, the BLUP is referred to as the *ordinary kriging* estimator. When the mean is not constant, an approach for fitting Equation (2.3) is via *trend surface models*, which models the mean using a highly parametrized fixed effects structure. The issue with trend surface models is that it requires a large number of regression coefficients to capture even simple spatial structure. An alternative is to do localized estimation, such as kernel estimation, and is also referred to as non-parametric regression. Such models depend on the kernel function, the degree of the local polynomial, and the bandwidth. A third approach extends the ideas from simple kriging, which derives the BLUP also under squared error loss but accounts for the unknown mean structure in the data. This approach is referred to as *universal kriging*. The universal kriging predictor is a function of $\hat{\boldsymbol{\beta}}_{\text{gls}}$.

The discussion of the previous classical kriging methods tacitly assumes that the variance-covariance structure of the data is known, i.e., the covariance parameters are known, and the variance of the measurement error is also known. However, in practice, this is not the case. The implication of substituting an estimate of $\boldsymbol{\theta}$ in the kriging predictor is that the obtained predictor is no longer the BLUP, but only an estimate of it, which is referred to as the EBLUP. Moreover, substituting the estimate for $\boldsymbol{\theta}$ in the variance expression of the BLUP only provides an estimate of the prediction error of the BLUP, not the actual prediction error of the EBLUP. Strategies to provide a correction for the prediction errors when covariance parameters are estimated were proposed in Harville and Jeske (1992), Kackar and Harville (1984),

and Prasad and Rao (1990).

Kriging approaches honor the data, i.e., the predicted surface passes through the data points. In some applications, the goal is to make inference on the signal of the model, i.e., to predict a less noisy version of the data that removes measurement error. In such a case, the prediction is often viewed as a *filtering* problem, since it predicts a filtered version of the data which removes random noise, i.e, the aim is to recover the signal from the noisy observations.

The ideas above extend to scenarios with more complex structures. One example is spatial prediction for binary data, which is referred to as *indicator kriging*. It views the binary data as a transformation of a latent continuously-indexed spatial process. Another extension is *cokriging*, which performs linear prediction based on more than one interrelated spatial processes. An in-depth discussion of other kriging methods are in Cressie (2015), Wackernagel (2003), Chiles and Delfiner (2012), Stein (2012); Waller and Gotway (2004).

2.2.4 Mixed models

Another classical framework for fitting spatial models and doing spatial prediction is by introducing *random effects*. In this framework, we can write Equations (2.3) as

$$\begin{aligned} \mathbf{w} &= \mathbf{Z}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\alpha} + \boldsymbol{\epsilon} \\ \mathbb{E}[\boldsymbol{\alpha}] &= \mathbf{0}, \mathbb{V}[\boldsymbol{\alpha}] = \mathbf{G}, \\ \mathbb{E}[\boldsymbol{\epsilon}] &= \mathbf{0}, \mathbb{V}[\boldsymbol{\epsilon}] = \mathbf{R} \end{aligned} \tag{2.4}$$

where $\mathbf{U}$ is a known matrix and $\boldsymbol{\alpha}$ is a $k \times 1$ vector of random effects. A special case is when $\mathbf{U}$ is an identity matrix so that each element of $\boldsymbol{\alpha}$, which in this case is of dimension n_{w}, is uniquely mapped to an element of $\mathbf{w}$. In relation to Equation (2.3), we have $\boldsymbol{\xi} \equiv \mathbf{U}\boldsymbol{\alpha}$. Equation (2.4) explicitly estimates $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$, also referred to as the *fixed effects* and *random effects*, respectively. The model in Equation (2.4) is a general form of a linear mixed model (LMM).

The development of estimation strategies for Equation (2.4) starts by assuming that $\mathbf{G}$ and $\mathbf{R}$ are known. The first approach obtains optimal estimates using least

squares theory (Henderson, 1950). The second approach assumes a Gaussian distribution for $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$, and then maximizes the joint likelihood of the random components to obtain the MLEs of $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. Not surprisingly, the MLE for $\boldsymbol{\beta}$ is equivalent to the generalized least squares estimator for $\boldsymbol{\beta}$ using the known form of $\mathbb{V}[\mathbf{w}]$. Also, the obtained MLE for $\boldsymbol{\alpha}$ is the BLUP under a squared error loss.

When $\mathbf{G}$ and $\mathbf{R}$ are unknown, the estimation of Equation (2.4) is typically done by assuming Gaussianity of $\boldsymbol{\alpha}$ and $\boldsymbol{\epsilon}$ and assuming a parametric model for the elements of $\mathbf{G}$ and $\mathbf{R}$, so that both matrices are parameterized by a set of few parameters, say $\boldsymbol{\theta}$. It is then straightforward to fit the model using maximum likelihood estimation or restricted MLE approach. The estimators for $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$ then take the EGLS form. A difficulty in this approach is the identifiability between the parameters describing $\mathbf{G}$ and $\mathbf{R}$.

A special case of Equation (2.4) is when $\mathbf{R} = \sigma_\epsilon^2 \mathbb{I}$ and $\mathbf{G} = \sigma^2 \mathbb{I}$, which under the squared error loss yields an objective function closely related to the objective function minimized in spline smoothing. For $\mathbf{s} \in \mathbf{R}^1$, truncated line functions (basis functions) of degree p are usually used, while for $\mathbf{s} \in \mathbf{R}^d$, radial basis functions are typically used as the spline basis. For all cases, the connection between the spline smoothing models and Equation (2.4) is that the spline basis functions are specified in the $\mathbf{U}$ matrix, while $\boldsymbol{\alpha}$ are viewed as the spline coefficients. An excellent discussion of this topic is given in Ruppert et al. (2003). Since $\mathbb{V}[\boldsymbol{\alpha}] = \sigma^2 \mathbb{I}$, then the spatial structure in the signal is encoded in the spline basis $\mathbf{U}$, which depends on the spatial configuration of the data and the knots. If $\mathbf{U}$ is assumed to contain additional spatial dependence, the computational demand increases quickly. The number of knots also has an impact on the computational requirements.

Another special case of the LMM framework is to rewrite Equation (2.3) as follows:

$$\begin{aligned} \mathbf{w} &= \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\xi} + \boldsymbol{\epsilon} \\ \boldsymbol{\xi} &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\xi), \quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \sigma_\epsilon^2 \mathbb{I}), \end{aligned} \tag{2.5}$$

where $\text{Cov}[\boldsymbol{\xi}, \boldsymbol{\epsilon}] = \mathbf{0}$. Here, it is also assumed that $\boldsymbol{\Sigma}_\xi$ is parameterized by a few set of parameters $\boldsymbol{\theta}_\xi$ via the assumed covariance function under the weak stationarity assumption. The moments of Equation (2.5) are equal to the moments of Equation

(2.3), i.e., $\mathbb{E}[\mathbf{w}] = \mathbf{Z}\boldsymbol{\beta}$ and $\mathbb{V}[\mathbf{w}] = \boldsymbol{\Sigma}_\xi(\boldsymbol{\theta}_\xi) + \sigma^2_\epsilon\mathbb{I}$. The solutions for $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\alpha}}$ are based on Henderson's mixed model equations, so that $\hat{\boldsymbol{\beta}}$ takes the form of the GLS estimator for $\boldsymbol{\beta}$, and $\hat{\boldsymbol{\alpha}}$ is the BLUP for $\boldsymbol{\alpha}$ under a squared error loss. The form of the predictor for $\mathrm{w}(\mathbf{s}_0)$ is equivalent to the universal kriging predictor with filtered measurement error, which is not a surprise.

2.2.5 Change of support for w(s)

As seen from the motivating problems in Sections 1.2.1 and 1.2.2, the change of support problem involves changing the support of a point-referenced spatial process of the first-stage model to the spatial support of the second-stage model which is an areal process. In the traditional geostatistical framework, this is referred to as *block kriging*.

In the following description of the block kriging method, we assume that $\mathrm{w}(\mathbf{s}_i) = \mu + \xi(\mathbf{s}_i)$, where $\xi(\mathbf{s}_i)$ is a weakly stationary process. This is a simplification from Equation (2.3), which assumes a non-constant mean and that $\mathrm{w}(\mathbf{s}_i)$ has a measurement error. Suppose $\mathrm{w}(B_j)$ denotes the block-level (average) measurement of $\mathrm{w}(\mathbf{s})$ over B_j. This quantity is defined as $\mathrm{w}(B_j) = \frac{1}{|B_j|}\int_{B_j} \mathrm{w}(\mathbf{s})d\mathbf{s}$. Similar to the previous kriging approaches, a squared error loss function is used and the optimal estimator is restricted to the class of linear unbiased estimators. The optimal predictor for $\mathrm{w}(B_j)$ has a similar form as the ordinary kriging predictor, except that we need $\mathrm{Cov}\big[\mathrm{w}(B_j), \mathrm{w}(\mathbf{s}_i)\big], i = 1, \ldots, n_\mathrm{w}$, instead of $\mathrm{Cov}\big[\mathrm{w}(\mathbf{s}_0), \mathrm{w}(\mathbf{s}_i)\big], i = 1, \ldots, n_\mathrm{w}$ (Chiles and Delfiner, 2012; Journel and Huijbregts, 1976). In particular, $\mathrm{Cov}\big[\mathrm{w}(B_j), \mathrm{w}(\mathbf{s}_i)\big]$ is defined as

$$\mathrm{Cov}\big[\mathrm{w}(B_j), \mathrm{w}(\mathbf{s}_i)\big] = \frac{1}{|B_j|}\int_{B_j} c(\mathbf{u} - \mathbf{s}_i)d\mathbf{u}. \tag{2.6}$$

The corresponding block kriging variance is also similar to the variance of the ordinary kriging predictor, except that we need $\mathrm{Cov}\big[\mathrm{w}(B_j), \mathrm{w}(B_j)\big]$ instead of $\mathrm{Cov}\big[\mathrm{w}(\mathbf{s}_i), \mathrm{w}(\mathbf{s}_i)\big]$. In particular, $\mathrm{Cov}\big[\mathrm{w}(B_j), \mathrm{w}(B_j)\big]$ is defined as

$$\mathrm{Cov}\big[\mathrm{w}(B_j), \mathrm{w}(B_j)\big] = \mathbb{V}\big[\mathrm{w}(B_j)\big] = \frac{1}{|B_j|^2}\int_{B_j}\int_{B_j} c(\mathbf{u} - \mathbf{v})d\mathbf{u}d\mathbf{v}. \tag{2.7}$$

The previous description of block kriging assumes a constant mean, no measure-

ment error for $\mathrm{w}(\mathbf{s}_i)$, and known covariance function parameters. For scenarios when the previous assumptions are not met, block kriging can easily accommodate additional model requirements using the same approaches discussed in Sections 2.2.2 and 2.2.3.

The integrals in Equations (2.6) and (2.7) are approximated using the covariance function or its estimated version, say $\hat{c}(\cdot,\cdot)$. In particular, the block B_j is first discretized into a set of regular points, say $\mathbf{u}_j,\ j = 1,\ldots,n_{B_j}$. The approximations to Equations (2.6) and (2.7) are then computed as follows:

$$\begin{aligned}\mathrm{Cov}\big[\mathrm{w}(B_j),\mathrm{w}(\mathbf{s}_i)\big] &\approx \frac{1}{n_{B_j}}\sum_{j=1}^{n_{B_j}} c(\mathbf{u}_j-\mathbf{s}_i)\\ \mathrm{Cov}\big[\mathrm{w}(B_j),\mathrm{w}(B_j)\big] &\approx \frac{1}{\left(n_{B_j}\right)^2}\sum_{i=1}^{n_{B_j}}\sum_{j=1}^{n_{B_j}} c(\mathbf{u}_i-\mathbf{u}_j)\end{aligned} \tag{2.8}$$

The above quantities can also be computed using the semivariogram (Cressie, 2015).

With a Gaussian assumption on $\mathrm{w}(\mathbf{s}_i)$, things become simpler. Suppose the data follows $\mathbf{w} \sim \mathcal{N}\Big(\boldsymbol{\mu}, \boldsymbol{\Sigma}(\boldsymbol{\theta})\Big)$, where the $(i,j)^{\text{th}}$ element of $\boldsymbol{\Sigma}(\boldsymbol{\theta})$ is $c(\mathbf{s}_i,\mathbf{s}_j)$. Suppose the desired prediction on blocks is given by $\mathbf{w}_B^\intercal = \begin{pmatrix}\mathrm{w}(B_1) & \mathrm{w}(B_2) & \cdots & \mathrm{w}(B_m)\end{pmatrix}$. It follows that $\mathbf{w}_B \sim \mathcal{N}\Big(\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_B(\boldsymbol{\theta})\Big)$ since the Gaussian distribution is closed under affine transformations. Here, the j^{th} element of $\boldsymbol{\mu}_B$ is $\mu_{B_j} = \dfrac{1}{B_j}\displaystyle\int_{B_j}\mathrm{w}(\mathbf{s})d\mathbf{s}$, the $(j,j)^{\text{th}}$ element of $\boldsymbol{\Sigma}_B(\boldsymbol{\theta})$ is given in Equation (2.7), and the $(j,k)^{\text{th}}$ element of $\boldsymbol{\Sigma}_B(\boldsymbol{\theta})$ is $\mathrm{Cov}\big[\mathrm{w}(B_j),\mathrm{w}(B_k)\big] = \dfrac{1}{|B_j|\times|B_k|}\displaystyle\int_{B_j}\int_{B_k} c(\mathbf{u}-\mathbf{v})d\mathbf{u}d\mathbf{v}$. It then follows that the joint distribution of $\mathbf{w}$ and $\mathbf{w}_B$ is given by

$$\begin{pmatrix}\mathbf{w}\\ \mathbf{w}_B\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\boldsymbol{\mu}\\ \boldsymbol{\mu}_B\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Sigma}(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_{\mathrm{s},B}(\boldsymbol{\theta})\\ \boldsymbol{\Sigma}_{\mathrm{s},B}^\intercal(\boldsymbol{\theta}) & \boldsymbol{\Sigma}_B(\boldsymbol{\theta})\end{pmatrix}\right), \tag{2.9}$$

where $\boldsymbol{\Sigma}_{\mathrm{s},B}(\boldsymbol{\theta})$ is the matrix of cross-covariances between the elements of $\mathbf{w}$ and $\mathbf{w}_B$, whose elements are given in Equation (2.6). The required integration to obtain the elements of $\boldsymbol{\mu}_B, \boldsymbol{\Sigma}_{\mathrm{s},B}(\boldsymbol{\theta})$, and $\boldsymbol{\Sigma}_B(\boldsymbol{\theta})$ can be done using Monte Carlo simulation, such as in Equations (2.8). Given Equation (2.9), the conditional distribution of the blocks given the observed points can then be derived using the properties of the multivariate normal distribution.

2.3 Classical models for areal data

This section presents classical ideas in modelling spatial dependence for areal data. In such datasets, the spatial dependence is specified via a neighborhood structure of the blocks. This section discusses two classical models: the simultaneous autoregressive (SAR) model and the conditional autoregressive model (CAR).

Suppose $\mathrm{y}(B)$ denotes the areal spatial process. In the presentation below, we assume that $\mathrm{y}(B)$ is Gaussian, although the motivating examples of this PhD thesis specify $\mathrm{y}(B)$ as a Poisson random variable.

The first model, called the SAR model, was introduced by Whittle (1954). It is given by

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{B}\boldsymbol{\epsilon} + \boldsymbol{\nu}, \tag{2.10}$$

where $\mathbf{y} = \begin{pmatrix} \mathrm{y}(B_1) & \cdots & \mathrm{y}(B_{n_\mathrm{y}}) \end{pmatrix}^\intercal$, $\boldsymbol{\epsilon} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $\mathbf{B}$ is a matrix of spatial dependence parameters with diagonal elements equal to zero, $\boldsymbol{\nu}$ is a random vector with $\mathbb{E}[\boldsymbol{\nu}] = \mathbf{0}$, and $\mathbb{V}[\boldsymbol{\nu}]$ is a diagonal matrix with elements $\sigma_1^2, \ldots, \sigma_{n_\mathrm{y}}^2$. Equation (2.10) can be written as

$$(\mathbb{I} - \mathbf{B})(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\nu}, \quad \mathbb{V}(\mathbf{y}) = (\mathbb{I} - \mathbf{B})^{-1}\boldsymbol{\Sigma}_\nu(\mathbb{I} - \mathbf{B}^\intercal)^{-1}, \tag{2.11}$$

assuming that $(\mathbf{I} - \mathbf{B})$ is non-singular. A parametric form is assumed for $\mathbf{B}$ given by $\mathbf{B} = \rho\mathbf{W}$. The matrix $\mathbf{W}$ is a matrix whose values, denoted by w_{jk}, are computed using spatial proximity measures. A basic measure is $w_{jk} = 1$ if areas j and k are adjacent or share a common boundary, and $w_{jk} = 0$ otherwise. The SAR model can now be written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} = \rho\mathbf{W}\boldsymbol{\epsilon} + \boldsymbol{\nu}.$$

This model can be straightforwardly estimated using least squares theory and maximum likelihood methods. In order for the model to be well-defined, $\mathbb{I} - \rho\mathbf{W}$ needs to be non-singular.

The development of the SAR model starts by specifying a joint model for the data. Another approach is to start with conditional models instead, i.e., the conditional distribution of $\mathrm{y}(B_j)$ given its neighbours, say, δ_j. This led to the so-called conditional autoregressive model (CAR), which was introduced by Besag (1974). The set of conditional distributions $\left\{\pi\big(\mathrm{y}(B_j)|\delta_j\big)\right\}$ have a so-called *local* specification, i.e., the outcome on the jth areal unit depends only on its set of neighbors. Under this assumption, the outcomes $\mathbf{y}$ is also referred to as a *Markov random field*, the Markov property being that the conditional distribution of $\mathrm{y}(B_j)$ given the observed data is the same as $\pi\Big(\mathrm{y}(B_j)|\delta_j\Big)$. *Brook's Lemma* provides the theoretical results to obtain a proper joint distribution given only information on the full conditionals (Brook, 1964).

Banerjee et al. (2014) provides a useful theoretical discussion of the CAR model. The form of the joint distribution of $\mathbf{y}$ is given by

$$\pi\Big(\mathrm{y}(B_1),...,\mathrm{y}(B_{n_\mathrm{y}})\Big) \propto \exp\Big\{-\frac{1}{2\sigma^2}\mathbf{y}^\intercal(\mathbf{D}_w-\mathbf{W})\mathbf{y}\Big\}, \tag{2.12}$$

where $\mathbf{W}$ is the spatial proximity matrix, $\mathbf{D}_w$ is a diagonal matrix with jth diagonal element as the number of neighbours of area B_j, and σ^2 is such that $\mathbb{V}\Big[\mathrm{y}(B_j)|\mathrm{y}(B_k), k\neq j\Big]=\sigma^2$. Equation (2.12) takes the form of a *Gibbs distribution*. Geman and Geman (1984) has shown that if $\pi\Big(\mathrm{y}(B_1),...,\mathrm{y}(B_{n_\mathrm{y}})\Big)$ follows the Gibbs distribution, then $\mathbf{y}$ is a Markov random field. This is essentially the converse of the *Hammersley-Clifford Theorem*, which shows that the unique joint distribution, if it exists, given the conditional models $\pi\Big(\mathrm{y}(B_j)|\delta_j\Big)$, takes the form of a Gibbs distribution (Besag, 1974).

Equation (2.12) is an improper distribution due to the singularity of $\mathbf{D}_w-\mathbf{W}$. Nonetheless, even if the joint distribution is improper, all the conditional distributions are proper. This specification is also called the *intrinsically autoregressive* (IAR) model or intrinsic CAR model. A solution to the impropriety of the joint distribution in (2.12) is to incorporate a new parameter ρ and define $\mathbf{\Sigma}^{-1}=\mathbf{D}_w-\rho\mathbf{W}$. Positive definiteness of $\mathbf{\Sigma}^{-1}$ is assured when $|\rho|<1$. Given this specification, the full

conditionals are now given by

$$\mathrm{y}(B_j)|\mathrm{y}(B_k), k \neq j \sim \mathcal{N}\left(\rho \sum_k \frac{w_{jk}\mathrm{y}(B_k)}{w_{j+}}, \frac{\sigma^2}{w_{j+}}\right),$$

where w_{j+} denotes the number of neighbours of the j^{th} area. Here, the conditional mean of $\mathrm{y}(B_j)$ can be interpreted as a proportion on the average of its neighbors. Moreover, if $\rho = 0$, then the $\mathrm{y}(B_j)$'s are independent.

In practice, a CAR specification is incorporated in a model as random effects. In other words, the use of the CAR model is relegated to the prior distributional specification of the random effects in a Bayesian hierarchical framework. Hence, working with a proper joint distribution is not necessary. In fact, an improper specification of the joint distribution enables a wide breadth of spatial patterns in the data (Banerjee et al., 2014).

2.4 GLM and GLMM specification

This section aims to discuss classical spatial models for non-Gaussian data. This is relevant to this PhD thesis as the health models (second-stage model) involve count data as the response variable. To simplify the exposition, we assume that the response variable $\mathrm{y}(\cdot)$ is a point-referenced spatial process. Extending this to the case of an areal process is straightforward. The models and ideas in this section are also relevant for the first-stage model.

For non-Gaussian data, the spatial model for $\mathrm{y}(\mathbf{s})$ is appropriately specified as a generalized linear model (GLM), which is written as

$$\mathrm{y}(\mathbf{s}_i) \sim \mathcal{F}(\mathrm{y}; \cdot), \quad \mathrm{g}\Big(\mu_{\mathrm{y}}(\mathbf{s}_i)\Big) \equiv \mathrm{g}\Big(\mathbb{E}\big[\mathrm{y}(\mathbf{s}_i)\big]\Big) = \mathbf{z}(\mathbf{s}_i)^{\intercal}\boldsymbol{\beta}, \tag{2.13}$$

where $\mathcal{F}$ is a member of the exponential family of distributions, $\mathrm{g}(\cdot)$ is the link function which is a monotonic function of the mean $\mu_{\mathrm{y}}(\mathbf{s}_i)$, and $\mathbf{z}(\mathbf{s}_i)$ is a vector of known covariates with coefficients $\boldsymbol{\beta}$. An excellent discussion on the theory of generalized linear models is in McCullagh (2019). An important element of GLM is that the variance of the data is a function of its mean. This relationship is encoded in the

so-called *variance function*.

Under the assumption of independence of the data in Equation (2.13), the variance of $\mathbf{y} = \begin{pmatrix} \mathrm{y}(\mathbf{s}_1) & \cdots & \mathrm{y}(\mathbf{s}_n) \end{pmatrix}$ is specified as $\mathbb{V}[\mathbf{y}] = \psi \mathbf{V}_{\boldsymbol{\mu}}$, where $\mathbf{V}_{\boldsymbol{\mu}}$ is a diagonal matrix with elements equal to the variance function evaluated at $\boldsymbol{\mu}_{\mathrm{y}}$, and ψ is a scaling parameter of $\mathcal{F}(\cdot)$. Extending Equation (2.13) in order to account for spatial correlations, the form of $\mathbb{V}[\mathbf{y}]$ is modified as follows:

$$\mathbb{V}[\mathbf{y}] = \sigma^2 \mathbf{V}_{\boldsymbol{\mu}}^{1/2} \mathbf{R}(\boldsymbol{\theta}) \mathbf{V}_{\boldsymbol{\mu}}^{1/2}, \tag{2.14}$$

where $\mathbf{R}(\boldsymbol{\theta})$ is a correlation matrix whose elements are determined from an assumed covariance function (Section 2.2.1), under the assumption of weak stationarity in the spatial process. The constant σ^2 is the so-called *dispersion parameter* under the case when $\psi = 1$, which is true for commonly used distributions such as the binomial and Poisson family. The above specification is also referred to as a *marginal specification*, since the marginal mean $\mathbb{E}\big[\mathrm{y}(\mathbf{s}_i)\big]$ is viewed as a function of fixed, non-random parameters (Schabenberger and Gotway, 2017). Note that the model specified in Equation (2.3) also follows a marginal specification.

2.4.1 Conditional specification

The primary difficulty when working with non-Gaussian spatial data in the GLM framework is that the joint distribution for data with mean given in Equation (2.13), variance given by Equation (2.14), and marginal distributions of the form $\mathcal{F}(\cdot)$, may not exist. In many cases, no valid joint distribution exists that satisfies the previous three requirements. An alternative model specification that eases this difficulty is called the *conditional specification*, which assumes a spatially-varying latent process and then specifies the moments conditional on this process. The conditional specification takes the following form:

$$\begin{aligned} &\mathrm{y}(\mathbf{s}_i) | \boldsymbol{\beta}, \xi(\mathbf{s}_i) \sim \mathcal{F}(\mathrm{y}; \cdot) \\ &\mathrm{g}\Big(\mathbb{E}\big[\mathrm{y}(\mathbf{s}_i) | \boldsymbol{\beta}, \xi(\mathbf{s}_i)\big]\Big) = \mathbf{z}(\mathbf{s}_i)^{\intercal} \boldsymbol{\beta} + \xi(\mathbf{s}_i) \\ &\mathbb{V}\Big[\mathrm{y}(\mathbf{s}_i) | \boldsymbol{\beta}, \xi(\mathbf{s}_i)\Big] = \sigma^2 \nu\big(\mu(\mathbf{s}_i)\big). \end{aligned} \tag{2.15}$$

The conditional formulation leads to a generalized linear mixed model (GLMM), since $\xi(\mathbf{s}_i)$ is viewed as a random effect and is part of the linear predictor expression. Equation (2.15) is also referred to as a *hierarchical model*. The first level of the hierarchy describes how the data depend on the latent process; while the second level specifies the model of the latent process. In a Bayesian framework, there is a third level which specifies the prior distribution of all model parameters. In Equation (2.15), $\nu(\cdot)$ is the variance function while σ^2 is a dispersion parameter.

With (Gaussian) linear models, the marginal and conditional specifications give the same inference. However, in a GLM, the two specifications have different interpretations. Under the conditional model, we have $\mathbb{E}[\mathrm{y}(\mathbf{s}_i)] = \mathbb{E}_\xi\left[\mathrm{g}^{-1}\left(\mathbf{z}(\mathbf{s}_i)^\intercal\boldsymbol{\beta} + \xi(\mathbf{s}_i)\right)\right]$, where $\mathbb{E}_\xi[\cdot]$ is taken with respect to the distribution of $\xi(\mathbf{s}_i)$. On the other hand, for the marginal model, $\mathbb{E}[\mathrm{y}(\mathbf{s}_i)] = \mathrm{g}^{-1}\left(\mathbf{z}(\mathbf{s}_i)^\intercal\boldsymbol{\beta}\right)$. The advantage of using the conditional specification is the assumption that the data are independent conditional on $\xi(\mathbf{s})$ or any other random effects for that matter.

The use of a conditional specification for the spatial non-Gaussian data provides computational advantage via the assumption of conditional independence in the data. However, a disadvantage is that doing inference for some parameters, e.g. the fixed effects $\boldsymbol{\beta}$, requires integrating out $\xi(\mathbf{s}_i)$ or any other random effects, which can be numerically difficult (Breslow and Clayton, 1993).

2.4.2 Frequentist estimation approaches

A classical estimation method for GLMs is the *quasi-likelihood* approach (Wedderburn, 1974). With this approach, inference is performed based only on the first two moments of the data, not on a full likelihood. The parameter estimates, which are derived from the *score functions*, have nice asymptotic properties. While quasi-likelihood approaches were first developed for independent data, McCullagh (2019) extended the approach to dependent data, and then Gotway and Stroup (1997) extended the approach particularly for spatial GLMs. The quasi-likelihood function is a function only of the mean structure, such as in Equation (2.13), and the variance structure, as in Equation (2.14). Its derivative with respect to $\boldsymbol{\beta}$ yields the quasi-likelihood score functions, which are also known as *generalized estimating equations*

(GEE). Asymptotic properties of estimators based on the GEE in the context of spatial models are discussed in McShane et al. (1997). Gotway and Stroup (1997) proposed to iteratively estimate $\boldsymbol{\beta}$ and the variance parameters, say $\boldsymbol{\theta}$, using a reweighted generalized least squares approach in a spatial context.

Another classical method is the so-called *pseudo-likelihood* (PL) approach, which is an efficient and flexible approach for fitting GLMM (Wolfinger and O'connell, 1993). Essentially, the PL approach linearizes the problem by obtaining the first-order Taylor series expansion of the link function. This yields the so-called *pseudo data*, which also has a form for its conditional mean and conditional variance-covariance structure, as well as the marginal form. This simplifies the problem to a general linear regression model with spatially-correlated errors for the marginal case, and to a linear mixed model in the conditional case. Thus, the obtained estimators take the GLS form and are also the BLUP. The PL approach is only one of the methods in the class of so-called *linearization methods* which essentially derive a pseudo-model for the data whose properties depend only on the first two moments.

Another commonly used estimation approach is the so-called *penalized quasi-likelihood* (PQL) approach proposed by Breslow and Clayton (1993). It uses a Laplace approximation on the log-likelihood, and then uses the Fisher scoring algorithm to obtain the parameter estimates. The derived solutions from the PQL approach coincide with that of the pseudo-likelihood approach.

A popular alternative is the use of Bayesian methods. The Bayesian framework is briefly presented in Section 2.4.3. A thorough discussion of the integrated nested Laplace approximation (INLA) method, which this PhD thesis is mainly using for Bayesian model inference, is presented in Section 2.5.

2.4.3 Bayesian framework

As discussed in Section 2.4.1, the conditionally-specified model in Equation (2.15) is a hierarchical model, where the first level specifies the model for the data $\mathrm{y}(\mathbf{s}_i)$ conditional on the fixed effects $\boldsymbol{\beta}$ and the random effects $\xi(\mathbf{s}_i)$. The second level then specifies the model for the random effects $\xi(\mathbf{s}_i)$, which depends on a set of parameters,

say $\boldsymbol{\theta}_\xi$. The third level specifies the prior distribution of $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_\xi & \sigma^2 \end{pmatrix}^\intercal$, where σ^2 is a scaling parameter of the likelihood (first stage) model.

The desired posterior distribution is as follows:

$$\pi(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\theta}|\mathbf{y}) = \frac{\pi(\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\theta}, \mathbf{y})}{\pi(\mathbf{y})} = \frac{\pi(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\theta})\pi(\boldsymbol{\xi}|\boldsymbol{\theta})\pi(\boldsymbol{\beta}, \boldsymbol{\theta})}{\int\int\int \pi(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\xi}, \boldsymbol{\theta})\pi(\boldsymbol{\xi}|\boldsymbol{\theta})\pi(\boldsymbol{\beta}, \boldsymbol{\theta}) d\boldsymbol{\xi} d\boldsymbol{\beta} d\boldsymbol{\theta}}, \tag{2.16}$$

with $\pi(\cdot)$ denoting the relevant distribution, either the likelihood model, the random effects model, the prior distributions, etc.

The Bayesian approach has several advantages over competing approaches, such as the frequentist approaches discussed in Section 2.4.2. It has a more philosophically sound foundation, has a unified approach to data analysis, and can formally incorporate domain expertise or external empirical evidence into the analysis (Banerjee et al., 2014). In particular, a unified approach for data analysis means that there are no separate theories of estimation, hypothesis testing, multiple comparisons, etc., since Bayesian analyses follow directly from the posterior distribution. Other advantages include the fact that all inferences are conditional on the observed data, rather than hypothetical or unobserved datasets (in contrast to frequentist approaches, which rely on repeated sampling); Bayesian answers are more interpretable to nonspecialists; and Bayes procedures possess a range of desirable optimality properties (Carlin and Louis, 2008).

The main bottleneck with Bayesian inference is that the integrals in Equation (2.16) are not tractable in closed form for most realistic problems. Even with conjugate priors in a Bayesian hierarchical framework, which allows partial analytic evaluation of the integrals in Equation (2.16), the presence of some (nuisance) parameters, such as spatial random effects, causes some integrations to be intractable.

There have been enormous developments in the computational tools available for doing Bayesian inference. A survey of these methods is nicely summarized in Green et al. (2015). A classical and staple method for doing simulation-based inference is via Markov chain Monte Carlo (MCMC) integration methods. MCMC techniques include the Metropolis-Hastings algorithm (Hastings, 1970; Metropolis et al., 1953) and the Gibbs sampler (Gelfand and Smith, 1990; Geman and Geman, 1984). These

classical MCMC techniques have evolved in the last several years. These include the use of Langevin drift instead of random walk proposals, which provide a substantial speed-up in convergence at the extra price of computing the gradient. This approach is referred to as the Metropolis-adjusted Langevin algorithms (MALA) (Roberts and Tweedie, 1996). Another important computational tool is the Hamiltonian Monte Carlo (HMC) method, which introduces an auxiliary variable called the *momentum* and then simulates from the augmented target distribution according to Hamilton's equations (Neal, 2012).

The main issue with MCMC approaches is that they are inefficient in the context of highly complex models (Green et al., 2015). This motivated the development of approximate methods such as the ABC method (Wilkinson, 2013), variational Bayes (VB) (Jaakkola and Jordan, 2000), empirical likelihood (Owen, 2001), and integrated nested Laplace approximation (INLA) (Rue et al., 2009).

As discussed in Section 1.5, this PhD thesis will mainly use INLA to perform Bayesian inference. The INLA method is a deterministic approach for doing Bayesian inference, and is particularly well-suited for latent Gaussian models, which are formally presented in Section 2.5.1. Performing MCMC with latent Gaussian models can be painfully slow (Rue et al., 2009). For spatial applications in Bayesian hierarchical models, where commonly the size of the second-stage (latent) parameters is large, INLA method generally computes accurate approximations of the posterior marginals, which take hours for MCMC algorithms to compute.

There are several advantages to the use of INLA over other approximate methods. The posterior variance using the VB method has been shown to be significantly smaller than the true value in some applications of latent Gaussian models (Rue et al., 2009). It is not unusual for the VB methodology to underestimate the posterior variance (Wang and Titterington, 2005). There are several remedies for the limitations of the VB methodology, but the solutions are case-specific (Rue et al., 2009). Another approximation method, which is an exotic example of variational approximation, is the so-called expectation-propagation (EP) method (Minka, 2013). While the VB method tends to underestimates the posterior variance, the EP method tends to overestimates the posterior variance (Bishop, 1995). Both the VB and EP

methods are faster than the MCMC approach, but are generally slower than the approximation provided by the INLA method. Although the developments in approximation methods keep on growing due to their efficient programming, benefits from avoiding simulations, and the ability to deal with complex models, a general criticism against these methods is the overall incapacity to diagnose and quantify the amount of approximation errors involved (Green et al., 2015).

2.5 Integrated nested Laplace approximation

This section discusses the details of the INLA method. Section 2.5.1 starts with a discussion of latent Gaussian models, which is a class of Bayesian hierarchical models for which the INLA methodology is well-suited. Section 2.5.2 discusses the original INLA approach, while Section 2.5.3 discusses estimation algorithms that have superseded the classical approach in the current implementation in the `R-INLA` maintained by the original developers of INLA. This new methodology is more efficient than the classical INLA. Section 2.5.4 then discusses the iterated INLA approach, which is a recent development that allows the fitting of models with nonlinear components in the predictor expressions. Since the INLA approach has been developed as a general model fitting approach not limited to spatial modelling, exposition in the following sections is general, and not in a spatial context.

2.5.1 Latent Gaussian models

Latent Gaussian models are a subset of the class of Bayesian additive models with a structured additive predictor (Rue et al., 2009). Suppose y_i is the response variable with distribution function which is a member of the exponential family, and mean $\mathbb{E}[\mathrm{y}_i]$ which is linked to an additive predictor η_i through a link function g(.), i.e., $\mathrm{g}\big(\mathbb{E}[\mathrm{y}_i]\big) = \eta_i$. The general form of η_i is

$$\eta_i = \alpha + \sum_{k=1}^{n_\beta} \beta_k \mathrm{z}_{ki} + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \epsilon_i. \tag{2.17}$$

In Equation (2.17), α is the intercept, $\{\beta_k\}$ are the coefficients of known covariates z_{ki}, $\left\{f^{(j)}(\cdot)\right\}$ are unknown functions of covariates u_{ji}, and ϵ_i is an error term. The covariate u_{ji} can be the spatial location or time index for an observation i, so that the $f^{(j)}(u_{ji})$ terms take care of the spatial, temporal, or spatio-temporal dynamics in the model. The functions $f^{(j)}(\cdot)$ also could account for non-linear relationship between a covariate and the response. It can be modelled using parametric nonlinear terms, nonparametric forms such as a random walk model, or Gaussian processes. All unknown quantities in Equation (2.17), namely $\alpha, \{\beta_k\}, \left\{f^{(j)}(\cdot)\right\}$, and $\{\epsilon_i\}$ are assumed to be Gaussian; hence, this model specification is termed *latent Gaussian model*. The aforementioned parameters are referred to as the *latent parameters*, which could depend on another set of parameters called *hyperparameters*.

Just like any Bayesian hierarchical model, latent Gaussian models are conveniently specified as a three-stage hierarchical model. In the first stage, the y_i's are assumed to be conditionally independent given the latent parameters, say $\boldsymbol{x}$, and the model hyperparameters, say $\boldsymbol{\theta}$. With the assumption of conditional independence, the joint distribution of $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \cdots & y_n \end{pmatrix}^\intercal$ given $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is given by

$$\mathbf{y}|\boldsymbol{x}, \boldsymbol{\theta} \sim \prod_{i=1}^{n} \pi(y_i|\boldsymbol{x}, \boldsymbol{\theta}),$$

where x_i denotes the set of latent parameters linked to y_i. In the second stage, the model for the latent parameters, particularly $\{\eta_i\}, \alpha, \{\beta_k\}, \left\{f^{(j)}(\cdot)\right\}$ which are the elements of $\boldsymbol{x}$ is specified, i.e., $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ is multivariate normal with mean zero and precision matrix $\mathbf{Q}(\boldsymbol{\theta})$. Finally, in the third stage, the distribution of the vector of hyperparameters $\boldsymbol{\theta}$ is specified. The joint posterior distribution of $\boldsymbol{x}$ and $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \prod_{i=1}^{n} \pi(y_i|x_i, \boldsymbol{\theta}) \tag{2.18}$$

$$\propto \pi(\boldsymbol{\theta})|\mathbf{Q}(\boldsymbol{\theta})|^{-1/2} \exp\Big\{ -\frac{1}{2}\boldsymbol{x}^\intercal \mathbf{Q}(\boldsymbol{\theta})\boldsymbol{x} + \sum_{i=1}^{n} \log\left\{\pi(y_i|x_i, \boldsymbol{\theta})\right\}\Big\}. \tag{2.19}$$

If it is further assumed that the latent Gaussian vector $\boldsymbol{x}$ has Markov properties, i.e. for some $j \neq k$, x_j and x_k are conditionally independent given $\boldsymbol{x}_{-jk}$, then $\boldsymbol{x}$ is referred to as a Gaussian Markov random field (GMRF). A useful consequence of the

assumption is that the $(j,k)^{\text{th}}$ element of $\mathbf{Q}(\boldsymbol{\theta})$, denoted by Q_{jk}, is equal to 0 if and only if x_j and x_k are conditionally independent given $\boldsymbol{x}_{-jk}$. This makes inference computationally efficient, as working with a sparse matrix $\mathbf{Q}(\boldsymbol{\theta})$ is less costly than working with a dense matrix. The cost of factorizing a dense $n \times n$ matrix is $\mathcal{O}(n^3)$, whereas the computational cost when working with GMRFs decreases to $\mathcal{O}(n^{3/2})$ (Rue and Held, 2005).

It is usually useful to distinguish the hyperparameters which are directly linked to $\mathbf{y}$ (like shape or precision parameters), say $\boldsymbol{\theta}_1$, and those which are linked to the latent field $\boldsymbol{x}$, say $\boldsymbol{\theta}_2$. The latent Gaussian model can now be written as follows:

$$\begin{aligned} \mathbf{y}|\boldsymbol{x},\boldsymbol{\theta}_1 &\sim \prod_{i=1}^{n} \pi(\mathrm{y}_i|\boldsymbol{x},\boldsymbol{\theta}_1) \\ \boldsymbol{x}|\boldsymbol{\theta}_2 &\sim \mathcal{N}\Big(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta}_2)\Big) \\ \boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2\} &\sim \pi(\boldsymbol{\theta}). \end{aligned}$$

The primary goal is then to obtain the following marginal posteriors: $\pi(x_j|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$, where x_j and θ_j are the j^{th} elements of $\boldsymbol{x}$ and $\boldsymbol{\theta}$, respectively.

Running an MCMC algorithm for latent Gaussian models can exhibit poor performance since the components of the latent field $\boldsymbol{x}$ have strong dependence with each other, and since $\boldsymbol{\theta}$ and $\boldsymbol{x}$ are also strongly dependent, especially if the dimension of the latent parameters is large. There are several approaches to overcome these difficulties but MCMC sampling still remains a challenge under these scenarios (Rue and Held, 2005). These challenges are overcome with the use of an efficient Bayesian computational approach called *Integrated Nested Laplace Approximation* (INLA).

2.5.2 Classical INLA

INLA aims to estimate the marginal posterior distributions via the following nested integrals:

$$\pi(x_i|\mathbf{y}) = \int \pi(x_i|\boldsymbol{\theta},\boldsymbol{y})\pi(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \tag{2.20}$$

$$\pi(\theta_i|\mathbf{y}) = \int \pi(\boldsymbol{\theta}|y)d\boldsymbol{\theta}_{-j} \tag{2.21}$$

Approximating $\pi(x_i|\mathbf{y})$ and $\pi(\theta_i|\mathbf{y})$ first requires approximations of $\pi(\boldsymbol{\theta}|\mathbf{y})$ and $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$, denoted by $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ and $\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$, respectively. These are discussed in Sections 2.5.2.1 and 2.5.2.2, respectively. Equation (2.20) is computed via numerical integration, which is discussed in Section 2.5.2.4.

2.5.2.1 Approximating $\pi(\boldsymbol{\theta}|\mathbf{y})$

The approximation for $\pi(\boldsymbol{\theta}|\mathbf{y})$ is given by

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \left.\frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y})}\right|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}, \tag{2.22}$$

where $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y})$ is the Gaussian approximation to the full conditional of $\boldsymbol{x}$, and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the approximation for a given $\boldsymbol{\theta}$. The approximation is done by performing a Taylor series approximation on $\log \pi(\mathrm{y}_i|x_i,\boldsymbol{\theta})$ (see Equation (2.19)) about x_{0i} up to a second order for $i = 1,\ldots,n$. This gives an approximation of the form

$$\log \pi(\mathrm{y}_i|x_i,\boldsymbol{\theta}) \approx b_i x_i - \frac{1}{2}c_i x_i^2, \tag{2.23}$$

where both b_i and c_i depends on x_{0i}. This approximation is most accurate when the expansion is done around the modal value x_i^*, which can be searched using iterative optimization methods such as the Newton-Raphson method. Note that this Gaussian approximation is performed conditional on $\boldsymbol{\theta}$; thus, the mode of the approximation is written as $\boldsymbol{x}^*(\boldsymbol{\theta})$.

From Equation (2.23), $\pi(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y})$ can then be written as follows:

$$\begin{aligned}
\pi(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y}) &\propto \pi(\boldsymbol{x}|\boldsymbol{\theta})\pi(\mathbf{y}|\boldsymbol{x},\boldsymbol{\theta}) \\
&\approx \exp\Big\{-\frac{1}{2}\boldsymbol{x}^\intercal \mathbf{Q}(\boldsymbol{\theta})\boldsymbol{x}\Big\} \exp\Big\{\sum_{i=1}^{n}\Big(b_i x_i - \frac{1}{2}c_i x_i^2\Big)\Big\} \\
&\approx \exp\Big\{-\frac{1}{2}\boldsymbol{x}^\intercal \mathbf{Q}(\boldsymbol{\theta})\boldsymbol{x} - \frac{1}{2}\sum_{i=1}^{n} c_i x_i^2\Big\} \exp\Big\{\sum_{i=1}^{n} b_i x_i\Big\} \\
&= \exp\Big\{-\frac{1}{2}\boldsymbol{x}^\intercal \big[\mathbf{Q}(\boldsymbol{\theta}) + \mathrm{diag}(\boldsymbol{c})\big]\boldsymbol{x} + \boldsymbol{b}^\intercal \boldsymbol{x}\Big\}.
\end{aligned} \tag{2.24}$$

Since Equation (2.24) is the canonical form of the multivariate Gaussian distribution, then $\pi(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y})$ is approximated as Gaussian with mean $\big[\mathbf{Q} + \mathrm{diag}(\boldsymbol{c})\big]^{-1}\boldsymbol{b}$ and

variance matrix $\big[\mathbf{Q}+\text{diag}(\boldsymbol{c})\big]^{-1}$. The Gaussian approximation to the full conditional of $\boldsymbol{x}$ is then given by

$$\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y}) \propto \exp\Big\{-\frac{1}{2}\big(\boldsymbol{x}-\boldsymbol{x}^*(\boldsymbol{\theta})\big)^{\intercal}\big[\mathbf{Q}(\boldsymbol{\theta})+\text{diag}(\boldsymbol{c})\big]\big(\boldsymbol{x}-\boldsymbol{x}^*(\boldsymbol{\theta})\big)\Big\}. \tag{2.25}$$

An important property of the approximation in Equation (2.25) is that it inherits the Markov property of the GMRF $\boldsymbol{x}$ since we are simply adding constants on the diagonal terms and the off-diagonal terms remain unchanged. Also, the approximation is exact if $\pi(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y})$ is Gaussian. Rue and Martino (2007) showed that the approximation in Equation (2.22) is accurate for a wide variety of models in the latent Gaussian family.

2.5.2.2 Approximating $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$.

In the classical INLA framework, there are three ways to approximate $\pi(x_i|\boldsymbol{\theta},\mathbf{y})$: using Gaussian approximations, another Laplace approximation, and a simplified Laplace approximation.

The Gaussian approximation is the fastest and cheapest computationally, but it suffers from errors in the location and/or errors due to lack of skewness (Rue and Martino, 2007). Given Equation (2.25), $\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y})$ is obtained by marginalizing the approximate joint distribution. The extra difficulty in marginalizing $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\mathbf{y})$ is the computation of the marginal variances from the sparse precision matrix $\big[\mathbf{Q}(\boldsymbol{\theta})+\text{diag}(\boldsymbol{c})\big]$. Rue et al. (2009) provides a recursive process to obtain the marginal variances from the precision matrix.

The second approach provides more accurate approximations via another Laplace approximation given by

$$\tilde{\pi}(x_i|\boldsymbol{\theta},\mathbf{y}) \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{\pi}_{GG}(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})}\bigg|_{\boldsymbol{x}_{-i}=\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})},$$

where $\tilde{\pi}_{GG}(\cdot|\cdot)$ is the Gaussian approximation of $\pi(\boldsymbol{x}_{-i}|x_i,\boldsymbol{\theta},\mathbf{y})$ and $\boldsymbol{x}^*_{-i}(x_i,\boldsymbol{\theta})$ is the modal configuration. This method is computationally expensive since this has to be done for each x_i, and which requires the location of the mode and the factorization of a $(n_x-1)\times(n_x-1)$ matrix, where n_x is the dimension of $\boldsymbol{x}$. One way to ease computation

is by avoiding the optimization by approximating the modal configuration as

$$\boldsymbol{x}^*_{-i}(x_i, \boldsymbol{\theta}) \approx E_{\tilde{\pi}_G}(\boldsymbol{x}_{-i}|x_i), \tag{2.26}$$

where the expectation is computed with respect to the conditional density derived from $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y})$ in Equation (2.25). Other modifications can be found in Rue et al. (2009).

The third approach, called the simplified Laplace approximation, is less computationally expensive than the Laplace approximation and more accurate than the Gaussian approximation. It performs a Taylor expansion on the numerator and denominator of Equation (2.22) up to the third order, thus correcting for location and skewness error of the Gaussian approximation. The details of this approach can be found in Rue et al. (2009).

2.5.2.3 Approximating $\pi(\theta_j|\mathbf{y})$

Instead of providing a parametric form for $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, the INLA framework explores the approximate density to identify the evaluation points that would later be used for numerical integration when approximating the posterior marginal of x_i, which is discussed in Section 2.5.2.4. The key thing is that once $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is obtained, $\tilde{\pi}(\theta_i|\mathbf{y})$ can be determined by numerical integration, i.e., by summing out the other components. The strategy used to explore $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is as follows:

Step 1: Locate the mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ with respect to $\boldsymbol{\theta}$.

Step 2: Suppose $\boldsymbol{\theta}^*$ is the mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$. Compute the negative Hessian matrix $\mathbf{H}$ at $\boldsymbol{\theta}^*$. If the density were Gaussian, then $\mathbf{H}^-$ is the covariance matrix of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$.

Step 3: Reparametrize $\boldsymbol{\theta}$ in terms of $\boldsymbol{z}$ via $\boldsymbol{\theta}(\boldsymbol{z}) = \boldsymbol{\theta}^* + \mathbf{V}\boldsymbol{\Lambda}^{1/2}\boldsymbol{z}$, where $\mathbf{H}^{-1} = \mathbf{V}\boldsymbol{\Lambda}\mathbf{V}^\intercal$, the eigendecomposition of $\mathbf{H}^{-1}$.

Step 4: Explore $\log \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ using the $\boldsymbol{z}$-parameterization. The goal of this step is to identify values in the $\boldsymbol{\theta}$-space with high probability mass. These points will be used for doing numerical integration when approximating $\pi(x_i|\mathbf{y})$ and $\pi(\theta_i|\mathbf{y})$.

If the dimension of $\boldsymbol{\theta}$ is small, it is feasible to build a grid of $\boldsymbol{\theta}$-values where the density $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ is high. However, the computational cost of the grid approach grows exponentially with the size of $\boldsymbol{\theta}$. Another strategy, proposed in Rue and Martino (2007), is to approach the integration problem as a design problem. The goal is to identify a smaller number of integration points, using $\boldsymbol{\theta}^*$ and $\mathbf{H}$ as guide. This approach is referred to as *CCD strategy*.

The integration points identified using the grid exploration of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ can be used to implement better algorithms to approximate $\pi(\theta_j|\mathbf{y})$. For instance, the integration points can be used to construct an interpolant, and then perform numerical integration from this interpolant.

Martins et al. (2013a) proposed to approximate $\pi(\boldsymbol{\theta}|\mathbf{y})$ by a multivariate normal distribution by matching the mode and the curvature at the mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, and then obtaining $\tilde{\pi}(\theta_j|\mathbf{y})$ from it. The computation is straightforward since $\boldsymbol{\theta}^*$ and $\mathbf{H}$ are already available. To correct for skewness, the joint multivariate normal distribution is approximated as a sum of mixture of normals with possibly different scaling parameters. However, this approach becomes unstable when the number of hyperparameters is large. To remedy this, Martins et al. (2013a) proposed a method to directly approximate $\tilde{\pi}(\theta_j|\mathbf{y})$ without using integration algorithms. They proposed the following structure:

$$\tilde{\pi}(\theta_j|\mathbf{y}) = \begin{cases} \mathcal{N}(0, \sigma^2_{j+}), & \theta_j > 0 \\ \mathcal{N}(0, \sigma^2_{j-}), & \theta_j \leq 0 \end{cases}. \tag{2.27}$$

The parameters σ^2_{j+} and σ^2_{j-} are estimated using the result that if $\boldsymbol{\theta}$ is a multivariate Gaussian, then the marginal of θ_i can be viewed as a function of θ_i and $\boldsymbol{\theta}_{-i}$ evaluated at the conditional mean $\mathbb{E}(\boldsymbol{\theta}_{-i}|\theta_i)$. Hence, for each θ_j, the conditional mean $\mathbb{E}(\boldsymbol{\theta}_{-i}|\theta_i)$ will be computed, which depends only on $\boldsymbol{\theta}^*$ and $\mathbf{H}^{-1}$, both already available.

2.5.2.4 Approximating $\pi(x_i|\mathbf{y})$

Finally, Equation (2.20) can be integrated numerically via

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_{k=1}^{K} \tilde{\pi}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k, \tag{2.28}$$

where $\{\boldsymbol{\theta}_k\}_{k=1,\ldots,K}$ are the K integration points obtained from the grid exploration of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, and $\{\Delta_k\}_{k=1,\ldots,K}$ are the corresponding integration weights. A special case is to evaluate Equation (2.28) at the mode of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$, which is referred to as *empirical Bayes strategy*.

2.5.2.5 Linear combination of the latent field

It may be of interest to obtain posterior marginals of specific linear combinations of the latent parameters $\boldsymbol{x}$. Suppose that it is of interest to obtain $\boldsymbol{v} = \mathbf{A}\boldsymbol{x}$, where $\mathbf{A}$ is a $k \times n_x$ matrix, k is the number of linear combinations, and n_x is the dimension of $\boldsymbol{x}$.

There are two approaches to obtain the posterior marginals of $\boldsymbol{v}$. The first is to create an enlarged latent field $\tilde{\boldsymbol{x}} = (\boldsymbol{x}, \boldsymbol{v})$, and then use INLA to fit the enlarged model. If there are several linear combinations of interest, this can make the precision matrix less sparse, hence making the computations costly. The second approach is to perform post-processing on the results of the original model, which does not include $\boldsymbol{v}$ in the latent field. The joint density $\pi(\boldsymbol{v}|\boldsymbol{\theta}, \mathbf{y})$ can be approximated as Gaussian with mean and variance given by

$$\mathbb{E}[\boldsymbol{v}|\boldsymbol{\theta}, \mathbf{y}] = \mathbf{A}\mathbb{E}[\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y}]$$
$$\mathbb{V}[\boldsymbol{v}|\boldsymbol{\theta}, \mathbf{y}] = \mathbf{A}\mathbb{V}[\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y}]\mathbf{A}^\intercal$$

where $\mathbb{E}[\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y}]$ and $\mathbb{V}[\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y}]$ are the mean and variance of the Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y})$ given in Equation (2.25). Afterwards, $\boldsymbol{\theta}$ is integrated out from $\tilde{\pi}(\boldsymbol{v}|\boldsymbol{\theta}, \mathbf{y})$ using the same numerical integration as Equation (2.28). This approach is faster, but less accurate than the first approach.

2.5.3 Modern INLA

The modern formulation of INLA provides faster computation and more accurate inference (Van Niekerk et al., 2023). In the classical formulation, the latent field consists of $\boldsymbol{x} = \Big\{\{\eta_i\}, \alpha, \{\beta_k\}, \{f^{(j)}(\cdot)\}\Big\}$, i.e., the linear predictors $\eta_i = \mathrm{g}\Big(\mathbb{E}\big[y_i|\boldsymbol{x}, \boldsymbol{\theta}\big]\Big)$ are considered latent parameters. In the new specification, the latent parameters exclude $\{\eta_i\}$. This implies that the dimension of the new latent field is significantly smaller. Suppose we denote by $\boldsymbol{\chi} = \Big\{\alpha, \{\beta_k\}, \{f^{(j)}(\cdot)\}\Big\}$. The n linear predictors $\boldsymbol{\eta} = \begin{pmatrix}\eta_1 & \eta_2 & \cdots & \eta_n\end{pmatrix}^\intercal$ are linked to $\boldsymbol{\chi}$ via $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\chi}$, where $\mathbf{A}$ is a known sparse matrix and, as before, $\boldsymbol{\chi} \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta})^{-1}\Big)$. Thus, the new specification still falls under the class of latent Gaussian models.

This also implies some modifications in the implementation of the approximations to the posterior marginals $\pi(\chi_j|\mathbf{y})$ and $\pi(\theta_j|\mathbf{y})$. In particular, Equation (2.18) is now expressed as

$$\pi(\boldsymbol{\chi}, \boldsymbol{\theta}|\mathbf{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{\chi}|\boldsymbol{\theta}) \prod_{i=1}^{n} \pi\Big(\mathrm{y}_i|(\mathbf{A}\boldsymbol{\chi})_i, \boldsymbol{\theta}\Big),$$

where $(\mathbf{A}\boldsymbol{\chi})_i$ refers to the i^{th} row of $\mathbf{A}\boldsymbol{\chi}$.

Marginal posterior $\pi(\theta_j|\mathbf{y})$

The marginal posterior $\pi(\boldsymbol{\theta}|\mathbf{y})$ is still approximated via Equation (2.22), except that we are now working with $\boldsymbol{\chi}$ instead of $\boldsymbol{x}$. As opposed to Equation (2.24), the Gaussian approximation of $\pi(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})$ is now expressed as:

$$\begin{aligned}\pi(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y}) &\propto \pi(\boldsymbol{\chi}|\boldsymbol{\theta})\pi(\mathbf{y}|\boldsymbol{\chi}, \boldsymbol{\theta}) \\ &\approx \exp\Big\{-\frac{1}{2}\boldsymbol{\chi}^\intercal \mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta})\boldsymbol{\chi}\Big\} \exp\Big\{\sum_{i=1}^{n}\Big(b_i(\mathbf{A}\boldsymbol{\chi})_i - \frac{1}{2}c_i(\mathbf{A}\boldsymbol{\chi})_i^2\Big)\Big\} \\ &= \exp\Big\{-\frac{1}{2}\boldsymbol{\chi}^\intercal\big[\mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta}) + \mathbf{A}^\intercal\mathbf{D}\mathbf{A}\big]\boldsymbol{\chi} + \boldsymbol{b}^\intercal\mathbf{A}\boldsymbol{\chi}\Big\},\end{aligned} \tag{2.29}$$

where $\mathbf{D}$ is a diagonal matrix whose elements are $c_1, \ldots, c_n$, and that the elements of $\boldsymbol{b}$ and $\boldsymbol{c}$ are determined similar to Section 2.5.2.1. Equation (2.29) is the canonical form of a multivariate Gaussian so that we have

$$\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y} \sim \mathcal{N}\Big(\boldsymbol{\mu}^*(\boldsymbol{\theta}), \big[\mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta}) + \mathbf{A}^\intercal\mathbf{D}\mathbf{A}\big]^{-1}\Big), \tag{2.30}$$

where $\boldsymbol{\mu}^*(\boldsymbol{\theta})$ is the mode of the Gaussian approximation, which depends on $\boldsymbol{b}$ and $\mathbf{D}$. The precision matrix $\mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta}) + \mathbf{A}^\intercal\mathbf{D}\mathbf{A}$ is sparse. It depends on the Markov properties of the latent field $\boldsymbol{\chi}$ through $\mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta})$, and the non-zero entries of $\mathbf{A}^\intercal\mathbf{A}$.

The conditional marginal posterior of χ_j are calculated from Equation (2.30). Similar to the classical INLA framework, a search strategy is performed on $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ to identify the integration points. Similar approaches in Section 2.5.2.3 are used to obtain the marginal posteriors $\tilde{\pi}(\theta_j|\mathbf{y})$.

Marginal posterior $\pi(\chi_j|\mathbf{y})$

A numerical integration is performed, similar to the classical INLA framework in Section 2.5.2.4, to obtain the marginal posteriors of χ_j:

$$\tilde{\pi}(\chi_i|\mathbf{y}) = \sum_{k=1}^{K} \tilde{\pi}(\chi_i|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k,$$

where $\{\boldsymbol{\theta}_k\}_{k=1,\ldots,K}$ are the K integration points obtained from the grid exploration of $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ and $\{\Delta_k\}_{k=1,\ldots,K}$ are the corresponding integration weights.

Marginal posterior $\pi(\eta_j|\mathbf{y})$

In the classical INLA framework, the marginal posteriors $\pi(\eta_j|\mathbf{y})$ are automatically approximated since the $\{\eta_j\}$ are part of the latent field. However, in the new INLA framework, these need to be calculated in a different step.

The approximation is based on the fact that $\boldsymbol{\eta} = \mathbf{A}\boldsymbol{\chi}$, i.e., $\boldsymbol{\eta}$ is a linear function of $\boldsymbol{\chi}$, which is approximately Gaussian conditional on $\boldsymbol{\theta}$ based on Equation (2.30). This implies that we have

$$\boldsymbol{\eta}|\boldsymbol{\theta}, \mathbf{y} \approx \mathcal{N}\Big(\mathbf{A}\mathbb{E}(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y}), \mathbf{A}\mathbb{V}(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})\mathbf{A}^\intercal\Big), \tag{2.31}$$

where the quantities $\mathbb{E}(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})$ and $\mathbb{V}(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})$ are available from Equation (2.30). A bottleneck of the above approximation is that in order to obtain $\mathbb{V}(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})$, it requires the inversion of $\mathbf{Q}_{\boldsymbol{\chi}}(\boldsymbol{\theta}) + \mathbf{A}^\intercal\mathbf{D}\mathbf{A}$, which can be expensive to calculate and store. Van Niekerk et al. (2023) proposed to store only the selected values in $\mathbb{V}(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})$ which are needed to compute the marginal posteriors of η_j. They also proposed an efficient approach to calculate the variance of the conditional marginal posterior of

η_j, conditional on $\boldsymbol{\theta}$.

From Equation (2.31), it is straightforward to calculate the conditional marginal posteriors $\tilde{\pi}(\eta_j|\boldsymbol{\theta}, \mathbf{y})$, which are also Gaussian. Finally, the approximated marginal posteriors of η_j are computed by integrating out the uncertainty in $\boldsymbol{\theta}$ via numerical integration, i.e.,

$$\tilde{\pi}(\eta_j|\mathbf{y}) \approx= \sum_{k=1}^{K} \tilde{\pi}(\eta_j|\boldsymbol{\theta}_k, \mathbf{y})\tilde{\pi}(\boldsymbol{\theta}_k|\mathbf{y})\Delta_k,$$

where $\{\boldsymbol{\theta}_k\}_{k=1,\ldots,K}$ are the integration points, and $\{\Delta_k\}_{k=1,\ldots,K}$ are the corresponding integration weights.

The posterior means of $\boldsymbol{\chi}$ and $\boldsymbol{\eta}$ may be inaccurate because they are based on the Gaussian approximation of $\pi(\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y})$. A correction to the mean of $\boldsymbol{\chi}|\boldsymbol{\theta}, \mathbf{y}$ is currently implemented using variational Bayes (Van Niekerk et al., 2023).

2.5.4 Iterated linearised INLA

Lindgren et al. (2024) extends the types of models that can be fitted with INLA, from models with an additive linear predictor expression, as shown in Equation (2.17), to models whose predictor expressions involve non-linear function of some latent parameters.

In the original INLA methodology, $\boldsymbol{\eta}(\boldsymbol{x}) = \begin{pmatrix} \eta_1(\boldsymbol{x}) & \eta_2(\boldsymbol{x}) & \cdots & \eta_n(\boldsymbol{x}) \end{pmatrix}^\intercal$ is additive in the latent Gaussian components. However, suppose that the predictor expression is non-linear, denoted by $\tilde{\boldsymbol{\eta}}(\boldsymbol{x})$. Lindgren et al. (2024) proposed to linearize $\tilde{\boldsymbol{\eta}}(\boldsymbol{x})$ via a 1st order Taylor series expansion around $\boldsymbol{x}_0$, i.e.,

$$\overline{\boldsymbol{\eta}}(\boldsymbol{x}) = \tilde{\boldsymbol{\eta}}(\boldsymbol{x}_0) + \mathbf{B}(\boldsymbol{x} - \boldsymbol{x}_0), \tag{2.32}$$

where $\overline{\boldsymbol{\eta}}(\boldsymbol{x})$ is the linearized expression and $\mathbf{B}$ is the matrix of derivatives. An important aspect of the algorithm is the determination of the linearization point, say, $\boldsymbol{x}_*$. An iterative approach is used to determine $\boldsymbol{x}_*$, given in Algorithm 2.1.

All details concerning posterior non-linearity checks and the evaluation of the accuracy of the approximation is discussed in Lindgren et al. (2024). This approach is implemented in the `inlabru` library in `R`.

Algorithm 2.1 Iterative method to determine $\boldsymbol{x}_*$

Step 1: Initialize $\boldsymbol{x} = \boldsymbol{x}_0$. Compute the linearized predictor $\overline{\boldsymbol{\eta}}(\boldsymbol{x}_0)$.

Step 2: Obtain the posterior $\overline{\pi}_{\boldsymbol{x}_0}(\boldsymbol{\theta}|\mathbf{y})$, where $\overline{\pi}(\cdot)$ denotes the obtained posterior using the linearized predictor.

Step 3: Let $(\boldsymbol{\theta}_1, \boldsymbol{x}_1) = (\hat{\boldsymbol{\theta}}_{\boldsymbol{x}_0}, \hat{\boldsymbol{x}}_0)$ be the initial candidate linearization point, where $\hat{\boldsymbol{\theta}}_{\boldsymbol{x}_0}$ is the mode of $\overline{\pi}_{\boldsymbol{x}_0}(\boldsymbol{\theta}|\mathbf{y})$, while $\hat{\boldsymbol{x}}_0$ is the mode of $\overline{\pi}(\boldsymbol{x}|\mathbf{y}, \hat{\boldsymbol{\theta}}_{\boldsymbol{x}_0})$.

Step 4: Let $\boldsymbol{x}_\alpha = (1-\alpha)\boldsymbol{x}_1 + \alpha\boldsymbol{x}_0$, where α is such that $\|\overline{\boldsymbol{\eta}}(\boldsymbol{x}_1) - \tilde{\boldsymbol{\eta}}(\boldsymbol{x}_\alpha)\|$ is minimized.

Step 5: Set $\boldsymbol{x}_0 = \boldsymbol{x}_\alpha$ and go back to Step 1. The iteration is terminated once convergence is achieved at a given tolerance.

2.5.5 Posterior sampling with INLA

INLA generates posterior samples in two steps, similar to the nested approach via the nested integrals to obtain the marginal posterior distributions (see Equations (2.20) and (2.21)).

The first step is to sample hyperparameter values $\boldsymbol{\theta}$ from the approximate posterior, as shown in Equation (2.22) and discussed in Section 2.5.2.3. The approximate posterior, which is obtained either via grid exploration or interpolation, gives a set of plausible values or evaluation points for $\boldsymbol{\theta}$ and their corresponding log-posterior value or weights. The weights are first normalized across evaluation points, and then the sampling is done randomly proportional to the weight. The hyperparameter samples are viewed as samples from a weighted mixture approximation to the true posterior.

The next step is to sample the values of the latent parameters $\boldsymbol{x}$ from the approximate conditional posterior $\tilde{\pi}(\boldsymbol{x}|\boldsymbol{\theta}, \mathbf{y})$. This is straightforward to implement since $\boldsymbol{x}$ is approximately Gaussian conditional on $\boldsymbol{\theta}$ and $\mathbf{y}$, as shown in Equation (2.25) and Equation (2.30) for the classical INLA and modern INLA, respectively. The posterior sampling strategy is efficient since the latent field is a GMRF, which yields sparse precision matrices for the approximate distributions. Moreover, it uses sparse Cholesky factorization, which works efficiently even for high-dimensional latent parameters, and it restricts numerical integration to a low-dimensional hyperparameter space.

2.6 SPDE Approach

Lindgren et al. (2011) proposed an efficient computational approach to estimate Gaussian fields of the Matérn class. The approach is based on the fact that Matérn fields, which have covariance function of the form

$$c(\mathbf{s}_1, \mathbf{s}_2) = \frac{\sigma^2}{2^{\nu-1}\Gamma(\nu)} \big(\kappa \left\|\mathbf{s}_1 - \mathbf{s}_2\right\| \big)^{\nu} K_{\nu}\big(\kappa \left\|\mathbf{s}_1 - \mathbf{s}_2\right\| \big) \tag{2.33}$$

are the stationary solutions to the following linear fractional stochastic partial differential equation (SPDE) (Whittle, 1954):

$$(\kappa^2 - \Delta)^{\alpha/2} \tau x(\mathbf{s}) = \mathcal{W}(\mathbf{s}), \quad \mathbf{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \quad \kappa > 0, \quad \nu > 0 \tag{2.34}$$

where $\nu > 0$ defines the mean-square differentiability or smoothness of the field, $\kappa > 0$ is a scaling parameter, σ^2 is the marginal variance, and $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^d$. $K_{\nu}(\cdot)$ is the modified Bessel function of the second kind and of order ν, $(\kappa^2 - \Delta)^{\alpha/2}$ is a pseudodifferential operator, $\mathcal{W}(\boldsymbol{s})$ is Gaussian white noise with unit variance, and Δ is the Laplace operator defined by $\Delta = \sum_{i=1}^{d} \dfrac{\partial^2}{\partial x_i^2}$.

The smoothness parameter ν is usually fixed at some value since this is poorly identified in many applications. The scaling parameter, κ, is related to the range parameter ρ, which is the distance at which the correlation is around 0.1. The empirically derived relationship is $\rho \approx \dfrac{\sqrt{8\nu}}{\kappa}$. The variance σ^2 is given by

$$\sigma^2 = \frac{\Gamma(\nu)}{\Gamma(\nu + d/2)(4\pi)^{d/2}\kappa^{2\nu}\tau^2}.$$

According to Whittle (1954), the random field is Markovian when $\alpha = \nu + d/2$ is an integer. For instance, when d is even, we get Matérn models with $\nu \in \mathbb{N}$. When $\nu = 1$, we get the thin plate spline model. On the other hand, when d is odd, we get Matérn models with $\nu \in \dfrac{1}{2}\mathbb{N}$. This includes Gaussian fields with the exponential covariance when $\nu = 1/2$.

2.6.1 Finite element representation of the SPDE

The solution to the SPDE in Equation (2.34) is approximated via a finite element representation given by

$$x(\mathbf{s}) = \sum_{k=1}^{K} \psi_k(\mathbf{s}) w_k, \tag{2.35}$$

where $\{\psi_k\}_{k=1,\ldots,K}$ are some basis functions which are chosen to be piecewise linear functions, while $\{w_k\}_{k=1,\ldots,K}$ are Gaussian-distributed weights. Equation (2.35) is defined on a triangulation in $\mathbb{R}^d$. This method for solving the SPDE is referred to as *finite element method* (FEM) (Ciarlet, 2002). The three corners of a triangle are called the *vertices*. The observed locations are usually used as the initial vertices, and then additional vertices are added given some constraints such as maximum edge length and minimum interior angles. A special triangulation is the so-called *Delaunay* triangulation which maximizes the minimum interior angles. The triangulation is also referred to as the *mesh*, while the vertices are also referred to as the *nodes*. In Equation (2.35), K is the number of mesh nodes, while the basis function $\psi_k(\mathbf{s})$ takes value equal to 1 at the node k and 0 at the other nodes. The weights $\{w_k\}$ are interpreted as the values of the field at the mesh nodes, while the values at the interior of the triangles are determined by linear interpolation. Thus, Equation (2.35) provides a continuously-indexed but finite-dimensional solution to the SPDE. The Gaussian weights $\{w_k\}$ characterize the full distribution of the solution.

2.6.2 Finite-dimensional solutions of the SPDE

The solution to the SPDE in Equation (2.34) via the basis function representation in Equation (2.35) is obtained by satisfying the *stochastic weak formulation* of the SPDE given by

$$\int \phi_j(\mathbf{s})(\kappa^2 - \Delta)^{\alpha/2} x(\mathbf{s}) d\mathbf{s} \overset{d}{=} \int \phi_j(\mathbf{s}) d\mathcal{W}(\mathbf{s}), \tag{2.36}$$

for suitable *test functions* $\{\phi_j(\mathbf{s})\}$ and where $\overset{d}{=}$ denotes 'equal in distribution'. For $\alpha = 1$, the test functions are chosen as $\phi_k = (\kappa^2 - \Delta)^{1/2}\psi_k$, while for $\alpha = 2$, the test functions are $\phi_k = \psi_k$. The solutions from the previously defined test functions are referred to as the *least squares* and *Galerkin* solution, respectively.

Substituting Equation (2.35) on Equation (2.36) and assuming that $\alpha = 1$ or $\alpha = 2$ yields the following system of linear equations:

$$\int \psi_j(\mathbf{s})(\kappa^2 - \Delta)\Big(\sum_k \psi_k(\mathbf{s})w_k\Big)d\mathbf{s} \overset{d}{=} \int \psi_j(\mathbf{s})d\mathcal{W}(\mathbf{s}). \tag{2.37}$$

The right-hand side of Equation (2.37) has the following property:

$$\int \psi_j(\mathbf{s})d\mathcal{W}(\mathbf{s}) \sim \mathcal{N}\Big(0, \int \psi_j(\mathbf{s})d\mathbf{s}\Big),$$

while the left-hand side of Equation (2.37) can be computed.

Suppose $\boldsymbol{w} = \begin{pmatrix} w_1 & \dots, w_K \end{pmatrix}^\intercal$, and the matrices $\mathbf{C}$, $\mathbf{G}$, and $\mathbf{K}$ are defined as follows:

$$\begin{aligned} C_{ij} &= \int \psi_i(\mathbf{s})\psi_j(\mathbf{s})d\mathbf{s} \\ G_{ij} &= \int \nabla\psi_i(\mathbf{s}), \nabla\psi_j(\mathbf{s})d(\mathbf{s}) \\ (\mathbf{K}_{\kappa^2})_{ij} &= \kappa^2 C_{ij} + G_{ij}. \end{aligned}$$

When $\alpha = 2$, and $\mathbf{s} \in \mathbb{R}^1$ or $\mathbf{s} \in \mathbb{R}^2$, Equation (2.37) becomes the (normal) equation $\mathbf{K}_{\kappa^2}\boldsymbol{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$; which then implies that

$$\boldsymbol{w} \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{K}_{\kappa^2}^{-1}\mathbf{C}\mathbf{K}_{\kappa^2}^{-1}\Big).$$

Thus, the Galerkin solution for $\boldsymbol{w}$ has precision matrix $\mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\mathbf{K}_{\kappa^2}$.

An important result in Lindgren et al. (2011) is the following: Suppose $\mathbf{s} \in \mathbb{R}^1$ or $\mathbf{s} \in \mathbb{R}^2$. The solution to the SPDE in Equation (2.34) is fully specified by the weights $\boldsymbol{w} \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{Q}_{\alpha,\kappa^2}^{-1}\Big)$, where the precision matrix $\mathbf{Q}_{\alpha,\kappa^2}$ depends on α and is a function of κ^2. The form of $\mathbf{Q}_{\alpha,\kappa^2}$ is as follows:

$$\begin{aligned} \mathbf{Q}_{1,\kappa^2} &= \mathbf{K}_{\kappa^2} \\ \mathbf{Q}_{2,\kappa^2} &= \mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\boldsymbol{K}_{\kappa^2} \\ \mathbf{Q}_{\alpha,\kappa^2} &= \mathbf{K}_{\kappa^2}\mathbf{C}^{-1}\mathbf{Q}_{\alpha-2,\kappa^2}\mathbf{C}^{-1}\mathbf{K}_{\kappa^2}, \quad \text{for } \alpha = 3, 4, \dots \end{aligned} \tag{2.38}$$

The above result holds only in $\mathbb{R}^1$ and $\mathbb{R}^2$. Also, the finite-dimensional representation of the result holds only for $\alpha = 1, 2, 3, \ldots,$, which corresponds to $\nu = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}, \ldots$ in $\mathbb{R}^1$ and $\nu = 0, 1, 2, \ldots$ in $\mathbb{R}^2$. The reason for restricting $\alpha \in \mathbb{N}$ is that in such scenarios, the power spectrum of the stationary solution of the SPDE in Equation (2.34) is the reciprocal of a positive symmetric polynomial, which makes the random field Markov.

The matrices $\mathbf{C}$ and $\mathbf{G}$ are sparse since C_{ij} and G_{ij} are non-zero only when ψ_i and ψ_j share common triangles. However, $\mathbf{C}^{-1}$ is not sparse, which makes $\mathbf{Q}_{\alpha,\kappa^2}$ dense. To make the precision matrix sparse, $\mathbf{C}$ is replaced by the diagonal matrix $\tilde{\mathbf{C}}$, where $\tilde{C}_{ii} = \int \psi_i(\mathbf{s}) d\mathbf{s}$. Due to the sparse structure of $\mathbf{Q}_{\alpha,\kappa^2}$, the finite-dimensional solution of the SPDE is a GMRF. The sparsity in $\mathbf{Q}_{\alpha,\kappa^2}$ provides great computational benefits.

The solution given in Equations (2.38) is only an approximation since these are solved using a subset of test functions. Nonetheless, for a given triangulation, this approximation converges (weakly) to the full SPDE solutions. Other asymptotic results of this approximate GMRF representation of the Matèrn field are in Lindgren et al. (2011). Moreover, since the solution to the SPDE is a GMRF, then incorporating the Matérn field as part of the latent parameters in the LGM in Equation (2.17) fits in the INLA framework.

Important extensions of the classical SPDE approach, as discussed above, are as follows: solutions of the SPDE on manifolds such as spheres, construction of non-stationary locally isotropic Gaussian fields by allowing spatially-varying SPDE parameters, a more complex version of the SPDE in Equation (2.34) in order to construct oscillating fields, generalizing non-stationary SPDE to non-isotropic fields, and to generalize the SPDE to non-separable space-time models.

2.7 Data fusion

This final section presents existing data fusion models, which are relevant for the materials in Chapter 4. As seen in the motivating examples in Sections 1.2.1 and 1.2.2, data used to conduct exposure assessment in an environmental health study usually come from various sources. Most studies use data from a network of monitoring stations. However, since the network is typically sparse due to high maintenance

costs of the stations, it will be difficult to capture spatial heterogeneity of the true exposure surface. In particular, for air pollution applications, networks are typically located in urban areas where the level of pollution is typically high. This leads to biases when fitting models to assess the impact of exposure on health as the exposure surface will be over estimated (Lawson et al., 2016).

A solution to this problem is the use of additional sources of information which provide more detailed spatial and temporal information. Two specific sources are satellite images and simulated outcomes from numerical models. These two data sources are also referred to as *proxy data*. Satellite images are remotely sensed data and provide global coverage. Although both aforementioned data sources provide rich spatial and temporal information, they are biased. Remotely sensed data, such as satellite images, are subject to retrieval errors; while numerical models are sensitive to model misspecification of the underlying process, the input data, model initialization, and the discretization of the continuous field. Combining data from monitoring stations and proxy data is referred to as *data fusion* or *data assimilation* (Lawson et al., 2016).

2.7.1 Bayesian melding

An approach to data fusion, called *Bayesian melding*, assumes that both the point-referenced data and the proxy data have a common latent spatial process (Fuentes and Raftery, 2005). Suppose that $\mathrm{w}(\mathbf{s})$ is the observed outcome from the station at location $\mathbf{s}$, $\tilde{x}(B)$ is the outcome of the proxy data in grid cell B, and $x(\mathbf{s})$ is the true latent process, e.g., the true temperature field. The Bayesian melding model in a purely spatial context is based on the following equations:

$$x(\mathbf{s}) = \mu(\mathbf{s}) + \xi(\mathbf{s}) \tag{2.39}$$

$$\mathrm{w}(\mathbf{s}) = x(\mathbf{s}) + e(\mathbf{s}) \tag{2.40}$$

$$\tilde{x}(\mathbf{s}) = \alpha_0(\mathbf{s}) + \alpha_1(\mathbf{s})x(\mathbf{s}) + \delta(\mathbf{s}) \tag{2.41}$$

$$\tilde{x}(B) = \frac{1}{|B|}\int_B \tilde{x}(\mathbf{s})d\mathbf{s}. \tag{2.42}$$

Equation (2.39) postulates that the latent spatial stochastic process is decomposed into a mean process $\mu(\mathbf{s})$, which is typically a function of fixed covariates, and a residual process $\xi(\mathbf{s})$ which is spatially correlated. Equation (2.40) states that the observed value at location $\mathbf{s}$ follows the classical error model, i.e., the observed value is a sum of the true values $x(\mathbf{s})$ and a random error term $e(\mathbf{s})$, which is assumed as $e(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma_e^2\big)$. Equation (2.41) postulates a conceptual point-referenced model for the proxy data. The underlying point-level outcome $\tilde{x}(\mathbf{s})$ is a linear function of $x(\mathbf{s})$, where $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ are additive and multiplicative biases, respectively, and $\delta(\mathbf{s})$ is a random noise assumed as $\delta(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma_\delta^2\big)$. The model specification implies that the proxy data are biased and more noisy compared to the observed values from the stations. The biases $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ are assumed to vary in space. Moreover, both are usually parametrized as fixed effects to avoid identifiability issues when estimating $x(\mathbf{s})$ (Lawson et al., 2016). Finally, equation (2.42) defines the observed value of $\tilde{x}(B)$ as a spatial average of $\tilde{x}(\mathbf{s})$ over $|B|$, where $|B|$ denotes the size of block B. The idea of assuming a common latent spatial process for the observed data at different spatial scales was also used in Wikle and Berliner (2005) McMillan et al. (2010), and Sahu et al. (2010).

2.7.2 Calibration technique

Another statistical approach for data fusion fits a regression model where the outcomes of the numerical forecast model are used as predictors, while the observational data are the response variable. This class of approaches is referred to as (statistical) calibration technique (Lawson et al., 2016). For example, Chen et al. (2021) used this approach to estimate chlorophyll-A concentration over eutrophic lakes, Lee et al. (2017) to model air quality, and Berrocal et al. (2010b) to model ozone concentration.

One calibration technique model proposed in Berrocal et al. (2010b) specifies the following calibration model:

$$\mathrm{w}_1(\mathbf{s}) = \alpha_0(\mathbf{s}) + \alpha_1(\mathbf{s})\mathrm{w}_2(B_\mathbf{s}) + e(\mathbf{s}), \quad e(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2). \tag{2.43}$$

Here, $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ are spatially-varying additive and multiplicative biases, respec-

tively, while $\mathrm{w}_2(B_{\mathbf{s}})$ is the value of the numerical forecast model at the grid B which contains the point location $\mathbf{s}$. This is also the same model proposed in Berrocal et al. (2010a) and Berrocal et al. (2012). In particular, they specify a bivariate continuous spatial process between $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$ via a linear model of coregionalization (Gelfand et al., 2004).

Obviously, there is a computational advantage with the regression calibration approach, since it only uses the values $\mathrm{w}_2(B_{\mathbf{s}})$ linked to the stations data $\mathrm{w}_1(\mathbf{s})$ to fit the model, and then uses the full data $\mathrm{w}_2(B)$ for predictions. This reduces the computational cost, especially if there are only few stations. However, a limitation of the approach is that it assumes that the observed measurements $\mathrm{w}_1(\mathbf{s})$ are the gold standard, even though they are very likely to have instrumental errors. Also, since $\mathrm{w}_2(B)$ is used as a predictor, then it cannot contain missing values (Lawson et al., 2016). However, remote-sensed data can be missing due to cloud cover and highly reflective surfaces. In addition, the resolution of the predicted latent surface for the quantity of interest depends on the resolution of $\mathrm{w}_2(B)$. The statistical calibration approach is also referred to as a *downscaling* approach since it allows point-level predictions even if $\mathrm{w}_2(B)$ represents an areal average.

2.7.3 Other approaches

A similar idea to the Bayesian melding model was proposed in Moraga et al. (2017). Given a zero-mean process $\xi(\mathbf{s})$ with a stationary covariance function, the model for the data outcomes $\mathrm{w}_1(\mathbf{s})$ and $\mathrm{w}_2(B)$ is given by

$$\begin{aligned} \mathrm{w}_1(\mathbf{s})|\xi(\mathbf{s}) &\sim \mathcal{N}\Big(\mu(\mathbf{s}) + \xi(\mathbf{s}), \sigma_e^2\Big) \\ \mathrm{w}_2(B) &= \frac{1}{|B|}\int_B \Big(\mu(\mathbf{s}) + \xi(\mathbf{s})\Big) d\boldsymbol{s}. \end{aligned} \tag{2.44}$$

However, the model in Equation (2.44) does not account for the measurement error in $\mathrm{w}_2(B)$. The same model specification is used in Zhong and Moraga (2023). Although it uses the same idea as the melding model in Equations (2.41) to (2.42), where both $\mathrm{w}_1(\mathbf{s})$ and $\mathrm{w}_2(B)$ have a common latent process, it does not incorporate bias parameters such as $\alpha_0(\mathbf{s})$ and $\alpha_1(\mathbf{s})$. In a joint modelling framework, the

data coming from $\mathrm{w}_2(B)$ can dominate the parameter estimation since there are considerably more outcomes from this data source compared to $\mathrm{w}_1(\mathbf{s})$ (Lawson et al., 2016). The calibration parameters in Equation (2.41) impose a restriction on this by accounting for a higher measurement error from this data source.

Forlani et al. (2020) proposed another data fusion model, but instead of assuming a single latent process for the observed outcomes, they assumed several latent processes, which are shared across all the data sources. For instance, suppose $\mathrm{w}_1(\mathbf{s})$, $\mathrm{w}_2(\mathbf{s})$, and $\mathrm{w}_3(\mathbf{s})$ are three data sources with mean $\mu_1(\mathbf{s})$, $\mu_2(\mathbf{s})$, and $\mu_3(\mathbf{s})$, respectively. Then the proposed model is as follows:

$$\begin{aligned}
\mu_1(\mathbf{s}) &= \beta_1 + \xi_1(\mathbf{s}) \\
\mu_2(\mathbf{s}) &= \beta_2 + \lambda_2\xi_1(\mathbf{s}) + \xi_2(\mathbf{s}) \\
\mu_3(\mathbf{s}) &= \beta_3 + \lambda_3\xi_1(\mathbf{s}) + \lambda_4\xi_2(\mathbf{s}) + \xi_3(\mathbf{s})
\end{aligned} \tag{2.45}$$

The parameters β_1, β_2, and β_3 are fixed effects while $\xi_i(\mathbf{s}), i = 1, 2, 3$ are spatial effects which are shared among the three data sources, and with $\lambda_j, j = 2, 3, 4$ as unknown scaling parameters. This approach is also closely related to the so-called coregionalization model (Schmidt and Gelfand, 2003). The melding model in Equations (2.39) to (2.42) assumes a single latent process which has a clear interpretation as the true process, and that the different data sources are error-prone realizations of the true process with varying levels of accuracy. However, the model in Equations (2.45) deviates from this general principle.

One bottleneck with the melding model is that it requires considerable computational effort because of the change-of-support integral in Equation (2.42). A model which tries to overcome this difficulty was proposed in Sahu et al. (2010). The model is specified as follows:

$$\begin{aligned}
x(\mathbf{s}) &= \tilde{x}(B) + \nu(\mathbf{s}) \\
\mathrm{w}_1(\mathbf{s}) &= x(\mathbf{s}) + e(\mathbf{s}) \\
\mathrm{w}_2(B) &= \alpha_0 + \alpha_1\tilde{x}(B) + \psi(B),
\end{aligned} \tag{2.46}$$

where $\tilde{x}(B)$ is considered as the true areal process and is defined on the same grid

resolution as $w_2(B)$, $\nu(\mathbf{s})$ is a Gaussian error, and $\psi(B)$ is a discrete spatial effect which is commonly estimated using a conditionally autoregressive model (Besag, 1974). McMillan et al. (2010) added further simplification by eliminating $x(\mathbf{s})$ in Equation (2.46). A limitation of both models is that they consider the underlying true process as discrete and that they can only provide gridded predictions of the true process. This approach is also referred to as *upscaling* since the point-referenced data are coarsened to areal level (Lawson et al., 2016).

Chapter 3

Data fusion with INLA-SPDE: an initial exploration

This chapter proposes a framework for doing data fusion in a two-stage spatio-temporal model. As discussed in Section 1.2, the first stage involves modelling a continuously-indexed spatial process, such as temperature and air pollution concentration; while the second stage fits the health model where the first-stage model predictions are inputs in the model. Moreover, as explained in Section 1.3.1.2, there are usually multiple data sources available to estimate the latent process of interest. It is advantageous to combine the different data sources in order to improve the predictions; a process which is referred to as data fusion (Lawson et al., 2016). Estimating the first-stage latent field using only data from the stations may be inferior to a model which considers, additionally, outcomes from proxy data, since these data provide wider spatial coverage, although biased (Lawson et al., 2016).

The main goal of this chapter is to provide an initial exploration of a data fusion model based on the Bayesian melding model (Section 2.7.1), and estimated using INLA (Section 2.5.2) and the SPDE method (Section 2.6). The reason for using the INLA-SPDE method is that both provide fast and accurate inference for spatio-temporal models (Lindgren et al., 2011; Rue et al., 2009). For this chapter, I start with constant calibration biases for the proxy data and investigate the performance of the INLA-SPDE approach under this scenario using an extensive simulation study.

I defer to Chapter 4 the discussion of a flexible model specification and a comparison with benchmark approaches. Although this chapter only presents an initial exploration of a data fusion model, this chapter also looks at the broader two-stage modelling framework shown in Figure 1.6. The specific gap in the literature this chapter addresses is the incorporation of a data fusion model into the approach proposed by Cameletti et al. (2019) for linking spatially misaligned health and air pollution data within a Bayesian framework. In this work, I employ a data augmentation strategy with the INLA-SPDE approach.

As explained in Section 1.2, the main outcome of interest, which is disease count, is measured and available in areas or blocks; while the first-stage latent process is point-referenced. A naive approach to this point-to-area COSP is to ignore the biased proxy data, and then use the simple average of the outcomes from the stations inside a block, possibly with the use of distance-based or population-based weights (Bruno et al., 2016). However, this approach does not work when there are blocks without stations, or if the values exhibit strong spatial heterogeneity (Krall et al., 2015; Lee et al., 2015). Another classical approach is to perform block kriging (see Section 2.2.5) in the first stage using the stations data. However, this require inversion of large dense matrices, which can be computationally expensive especially for large datasets.

In this chapter, the point-to-area COSP is done by computing spatial averages of the first-stage (data fusion) model predictions over the blocks (see Equation (2.1) of Chapter 2), and then applying standard statistical methods to regress the health outcomes against the estimated block-level values (Bruno et al., 2016).

The issue of uncertainty propagation is not yet dealt with formally in this chapter. In order to account for the uncertainty in the first-stage model, I used the posterior sampling approach (Cameletti et al., 2019). Essentially, this requires sampling several times from the estimated posterior distributions of the first-stage model. The spatial averages are then computed for each sample, and are used as inputs in the second-stage model. Thus, it requires fitting the second-stage model several times. The final posterior estimates of the second-stage model parameters are then obtained by Bayesian model averaging. The issue of uncertainty propagation is discussed more

formally in Chapter 6.

This chapter is structured as follows. Section 3.1 discusses the proposed data fusion model. The model assumptions and the estimation approach are discussed in Sections 3.1.1 and 3.1.2, respectively. The discussion of the uncertainty propagation approach is in Section 3.2.3. The extensive simulation study is presented in Section 3.3. Finally, I end this chapter with a summary of important results and conclusions in Section 3.4.

3.1 Proposed data fusion model

3.1.1 Model assumptions

I extend the assumed latent process in Equation (1.1) in Chapter 1 in a spatio-temporal context. The latent process of interest, say, true pollution field or climate field, is given by

$$x(\mathbf{s},t) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s},t) + \xi(\mathbf{s},t)$$
$$\xi(\mathbf{s},t) = \varsigma\xi(\mathbf{s},t-1) + \omega(\mathbf{s},t), \quad |\varsigma| < 1, \quad t = 2,...,T.$$

This follows the spatio-temporal model proposed in Cameletti et al. (2013) for particulate matter concentration in the North-Italian region Piemonte. Here, $x(\mathbf{s},t)$ is the true outcome of the latent process at location $\mathbf{s}$ and time t, $\mathrm{z}(\mathbf{s},t)$ is a covariate, β_0 is the intercept of the model, and β_1 is the unknown coefficient of $\mathrm{z}(\mathbf{s},t)$. The spatio-temporal dependence in the model is induced by $\xi(\mathbf{s},t)$, which evolves in time as an autoregressive (AR) process of order 1, with $|\varsigma| < 1$ as the AR parameter, and $\xi(\mathbf{s},1) \sim \mathcal{N}\Big(0, \sigma_\omega^2/(1-\varsigma^2)\Big)$, which is the stationary distribution of $\xi(\mathbf{s},t)$. The term $\omega(\mathbf{s},t)$ is a temporally-independent Gaussian random field with mean 0 and Matèrn covariance function, i.e.,

$$\mathrm{Cov}\Big(\omega(\mathbf{s}_i,t),\omega(\mathbf{s}_j,u)\Big) = \begin{cases} 0 & t \neq u \\ \Sigma_{i,j} & t = u \end{cases},$$

with $\Sigma_{i,j} = \frac{\sigma_\omega^2}{2^{\nu-1}\Gamma(\nu)}\big(\kappa \|\mathbf{s}_i - \mathbf{s}_j\|\big)^\nu K_\nu\Big(\kappa\big(\|\mathbf{s}_i - \mathbf{s}_j\|\big)\Big)$ (see also Section 2.6). Here, $\|\cdot\|$ is the Euclidean distance in $\mathbb{R}^2$, $\mathbf{s}_i$ and $\mathbf{s}_j$ are two spatial locations, σ_ω^2 is the marginal variance, ν is a mean-squared differentiability parameter, and κ is a scaling parameter. The interpretation of the spatio-temporal structure of the model is in terms of σ_ω^2 and the range parameter ρ, which is the distance at which the correlation is around 0.1. The empirically derived relationship between ρ and κ is $\rho = \frac{\sqrt{8\nu}}{\kappa}$. More details are found in Section 2.6.

Following Equation (2.40) of the Bayesian melding model, the observed values at n_M stations is given by:

$$\mathrm{w}(\mathbf{s}_i, t) = x(\mathbf{s}_i, t) + e(\mathbf{s}_i, t), \quad e(\mathbf{s}_i, t) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma_e^2\big), \quad i = 1, \ldots, n_M, \tag{3.1}$$

where $\mathrm{w}(\mathbf{s}_i, t)$ is the observed value at a station in location $\mathbf{s}_i$ at time t. This equation follows the classical error model in which the observed values are error-prone realizations of the true values of the latent process, and where the error process is a white noise, i.e., $e(\mathbf{s}_i, t) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma_e^2\big)$.

Here, I treat the proxy data as point-referenced at the centroids. This is a simplification of the Bayesian melding model, since Equation (2.41) is solely used to model the proxy data, instead of using both Equations (2.41) and (2.42). This assumption is reasonable when high-resolution proxy data are available; otherwise, performing stochastic integrals over such rasters can be computationally prohibitive.

Suppose that $\tilde{x}(\mathbf{g}_j, t)$ denotes the observed value of the proxy data at the grid cell with centroid $\mathbf{g}_j, j = 1, \ldots, n_G$, and at time t. The assumed model for $\tilde{x}(\mathbf{g}_j, t)$ is given by:

$$\tilde{x}(\mathbf{g}_j, t) = \alpha_0 + \alpha_1 x(\mathbf{g}_j, t) + \delta(\mathbf{g}_j, t), \quad \delta(\mathbf{g}_j, t) \sim \mathcal{N}\big(0, \sigma_\delta^2\big). \tag{3.2}$$

This simplification avoids the integration of the conceptual point-referenced model for the proxy data, given in Equation (2.42). In Equation (3.2), I assume a constant additive and multiplicative bias, denoted by α_0 and α_1, respectively. The assumption of a time-varying and spatially-structured additive bias, which provides more

flexibility, is presented in Chapter 4. The proposed model is not a full Bayesian melding approach, since both the proxy data and the stations data are considered point-referenced. The (statistical) regression calibration method (see Section 2.7.2) also eliminates the need to evaluate stochastic integrals, since it uses the proxy grid cell linked to a station at point location $\mathbf{s}_i$ as a predictor (Berrocal et al., 2010b; Lawson et al., 2016), or the estimated values of the proxy data obtained, say, via bilinear interpolation (Lee et al., 2017).This is also a simplifying assumption in Forlani et al. (2020). Even with the simplification, the proposed model follows the same principle as the Bayesian melding model, i.e., both data sources are a function of the same latent spatial process, and that one data source is more biased than the other.

3.1.2 Model estimation

Suppose that the data from the stations at time t is given by

$$\mathbf{w}_t^\intercal = \begin{pmatrix} \mathrm{w}(\mathbf{s}_1, t) & \mathrm{w}(\mathbf{s}_2, t) & \dots & \mathrm{w}(\mathbf{s}_{n_M}, t) \end{pmatrix}^\intercal, \quad t = 1, \dots, T.$$

Moreover, suppose that the proxy data at time t is given by

$$\tilde{\boldsymbol{x}}_t^\intercal = \begin{pmatrix} \tilde{x}(\mathbf{g}_1, t) & \tilde{x}(\mathbf{g}_2, t) & \dots & \tilde{x}(\mathbf{g}_{n_G}, t) \end{pmatrix}^\intercal, \quad t = 1, \dots, T.$$

Since both $\mathbf{w}_t$ and $\tilde{\boldsymbol{x}}_t$ are error-prone realizations of the latent process $x(\mathbf{s}, t)$, I define the vector of the latent (unknown) values for both the stations and proxy data at time t as $\boldsymbol{x}_t = \begin{pmatrix} \boldsymbol{x}_{t,S} & \boldsymbol{x}_{t,P} \end{pmatrix}^\intercal$, where $\boldsymbol{x}_{t,S}$ and $\boldsymbol{x}_{t,P}$ denote the vector of true values at the stations and at the grid centroids of the proxy data at time t, respectively. The vector of true exposures for all $t = 1, \dots, T$ is then denoted by $\boldsymbol{x} = \begin{pmatrix} \boldsymbol{x}_1^\intercal & \boldsymbol{x}_2^\intercal & \dots & \boldsymbol{x}_T^\intercal \end{pmatrix}^\intercal$. Similarly, I define $\boldsymbol{\xi} = \begin{pmatrix} \boldsymbol{\xi}_1^\intercal & \boldsymbol{\xi}_2^\intercal & \dots & \boldsymbol{\xi}_T^\intercal \end{pmatrix}^\intercal$, where $\boldsymbol{\xi}_t^\intercal = \begin{pmatrix} \boldsymbol{\xi}_{t,S}^\intercal & \boldsymbol{\xi}_{t,P}^\intercal \end{pmatrix}$, with $\boldsymbol{\xi}_{t,S}$ and $\boldsymbol{\xi}_{t,P}$ as the vector of spatio-temporal random effects at the stations and at the grid centroids of the proxy data at time t, respectively. Finally, $\boldsymbol{\omega} = \begin{pmatrix} \boldsymbol{\omega}_1^\intercal & \boldsymbol{\omega}_2^\intercal & \dots & \boldsymbol{\omega}_T^\intercal \end{pmatrix}^\intercal$ is the vector of values of the random field $\omega(\mathbf{s}, t)$, where $\boldsymbol{\omega}_t^\intercal = \begin{pmatrix} \boldsymbol{\omega}_{t,S}^\intercal & \boldsymbol{\omega}_{t,P}^\intercal \end{pmatrix}$; and $\mathbf{z} = \begin{pmatrix} \mathbf{z}_1^\intercal & \mathbf{z}_2^\intercal & \dots & \mathbf{z}_T^\intercal \end{pmatrix}^\intercal$ the vector of a single covariate,

where $\mathbf{z}_t^\intercal = \begin{pmatrix} \mathbf{z}_{t,S}^\intercal & \mathbf{z}_{t,P}^\intercal \end{pmatrix}$. This can easily be generalized to the case of more than one covariate.

The proposed data fusion model is as follows:

$$\mathbf{w}_t = \boldsymbol{x}_{t,S} + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathcal{N}\big(\mathbf{0}, \sigma_e^2 \mathbb{I}_{n_M}\big), \quad t = 1, \ldots, T \tag{3.3}$$

$$\tilde{\boldsymbol{x}}_t = \alpha_0 \mathbf{1}_{n_G} + \alpha_1 \boldsymbol{x}_{t,P} + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim \mathcal{N}\big(\mathbf{0}, \sigma_\delta^2 \mathbb{I}_{n_G}\big), \tag{3.4}$$

$$\boldsymbol{x}_t = \beta_0 \mathbf{1}_{n_M+n_G} + \beta_1 \mathbf{z}_t + \boldsymbol{\xi}_t, \tag{3.5}$$

$$\boldsymbol{\xi}_t = \varsigma \boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \ \boldsymbol{\omega}_t \sim \text{Matèrn field}, \quad t = 2, \ldots, T \tag{3.6}$$

where $\mathbb{I}_{n_M}$ and $\mathbb{I}_{n_G}$ are identity matrices of dimension $n_M \times n_M$ and $n_G \times n_G$, respectively, and $\mathbf{1}_{n_G}$ and $\mathbf{1}_{n_M+n_G}$ are vectors of 1's of dimension n_G and $n_M + n_G$, respectively.

As discussed in Section 3.1.1, $\boldsymbol{\omega}_t$ is a temporally-independent Gaussian vector with mean zero and covariance matrix $\boldsymbol{\Sigma}_t$, whose elements are computed using the Matérn covariance function (see Equation (2.33)).

In the system of equations above, the latent vector $\boldsymbol{x}_{t,S}$ is present in Equations (3.3) and (3.5). Also, $\boldsymbol{x}_{t,P}$ is present in Equations (3.4) and (3.5). In order to make sure that the values of $\boldsymbol{x}_{t,S}$ and $\boldsymbol{x}_{t,P}$ are equivalent for the different equations when fitting the joint model, $\boldsymbol{x}_{t,S}$ in Equation (3.3) is assumed to be an (almost) identical 'copy' of $\boldsymbol{x}_{t,S}$ in Equation (3.5). Similarly, $\boldsymbol{x}_{t,P}$ in Equation (3.4) is assumed to be an (almost) identical 'copy' of $\boldsymbol{x}_{t,P}$ in Equation (3.5), with α_1 as a scaling parameter. To create these 'copies', the latent field $\boldsymbol{x}_t$ is extended to $\boldsymbol{\chi}_t = \begin{pmatrix} \boldsymbol{x}_t^\intercal & \boldsymbol{x}_t^{*\intercal} \end{pmatrix}^\intercal, t = 1, \ldots, T$, where $\boldsymbol{x}_t^* = \begin{pmatrix} \boldsymbol{x}_{t,S}^{*\intercal} & \boldsymbol{x}_{t,P}^{*\intercal} \end{pmatrix}^\intercal$ is a copy of $\boldsymbol{x}_t$. The prior specification for the extended latent field at time t, $\pi\big(\boldsymbol{\chi}_t\big)$, will ensure that $\boldsymbol{x}_t^*$ is an identical copy of $\boldsymbol{x}_t$. In particular, $\boldsymbol{x}_{t,S}^*$ and $\boldsymbol{x}_{t,P}^*$ will be defined later in such a way that $\mathbb{E}\left(\boldsymbol{x}_{t,S}^*\right) = \boldsymbol{x}_{t,S}$ and $\mathbb{E}\left(\boldsymbol{x}_{t,P}^*\right) = \alpha_1 \boldsymbol{x}_{t,P}$. This is the same approach in Martins et al. (2013a) and Ruiz-Cárdenas et al. (2012), and is called a *data augmentation approach*. The copy trick allows us to use a latent effect in multiple model components, appropriately scaled, to propagate dependence and uncertainty consistently while avoiding duplication of parameters.

Since $\boldsymbol{x}_t$ on the left-hand side of Equation (3.5) is unknown, $\boldsymbol{x}_t$ is transposed on the right-hand side of the equation. The remaining vector of zeroes in the left-hand side are referred to as 'pseudo-zeroes'. This gives the following re-expression of the joint model:

$$\mathbf{w}_t = \boldsymbol{x}^*_{t,S} + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathcal{N}\big(\mathbf{0}, \sigma_e^2 \mathbb{I}_{n_M}\big), \quad t = 1, \ldots, T \tag{3.7}$$

$$\tilde{\boldsymbol{x}}_t = \alpha_0 \mathbf{1}_{n_G} + \boldsymbol{x}^*_{t,P} + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim \mathcal{N}\big(\mathbf{0}, \sigma_\delta^2 \mathbb{I}_{n_G}\big), \tag{3.8}$$

$$\mathbf{0}_t = - \begin{pmatrix} \boldsymbol{x}_{t,S} \\ \boldsymbol{x}_{t,P} \end{pmatrix} + \beta_0 \mathbf{1}_{n_M+n_G} + \beta_1 \mathbf{z}_t + \boldsymbol{\xi}_t, \tag{3.9}$$

$$\boldsymbol{\xi}_t = \varsigma \boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \mathcal{N}\big(\mathbf{0}, \boldsymbol{\Sigma}_t\big), \quad t = 2, \ldots, T \tag{3.10}$$

The model parameters are $\boldsymbol{\theta} = \Big\{\sigma_e^2, \sigma_\delta^2, \alpha_0, \alpha_1, \beta_0, \beta_1, \varsigma, \sigma_\omega^2, \kappa\Big\}$. The posterior distribution of interest is $\pi\Big(\boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \big| \mathbf{w}, \tilde{\boldsymbol{x}}, \mathbf{0}\Big)$, given by

$$\pi\Big(\boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \big| \mathbf{w}, \tilde{\boldsymbol{x}}, \mathbf{0}\Big) \propto \pi\Big(\mathbf{w} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) \pi\Big(\tilde{\boldsymbol{x}} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) \pi\Big(\mathbf{0} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) \pi\Big(\boldsymbol{\xi} \big| \boldsymbol{\theta}\Big) \pi\Big(\boldsymbol{\theta}\Big) \pi\Big(\boldsymbol{\chi}\Big).$$

The first two are straightforward since $\mathbf{w}_t \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \overset{\text{iid}}{\sim} \mathcal{N}\Big(\boldsymbol{x}^*_{t,S}, \sigma_e^2 \mathbb{I}_{n_M}\Big)$ and $\tilde{\boldsymbol{x}}_t \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \overset{\text{iid}}{\sim} \mathcal{N}\Big(\alpha_0 \mathbf{1}_{n_G} + \boldsymbol{x}^*_{t,P}, \sigma_\delta^2 \mathbb{I}_{n_G}\Big)$. This implies that

$$\pi\Big(\mathbf{w} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) = \prod_{t=1}^{T} \pi(\mathbf{w}_t \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}) \propto (\sigma_e^2)^{-\frac{n_M T}{2}} \exp\Big\{ - \frac{1}{2\sigma_e^2} \sum_{t=1}^{T} \Big(\mathbf{w}_t - \boldsymbol{x}^*_{t,S}\Big)^\intercal \Big(\mathbf{w}_t - \boldsymbol{x}^*_{t,S}\Big) \Big\},$$

$$\pi\Big(\tilde{\boldsymbol{x}} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) = \prod_{t=1}^{T} \pi(\tilde{\boldsymbol{x}}_t \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}) \propto (\sigma_\delta^2)^{-\frac{n_G T}{2}} \exp\Big\{ - \frac{1}{2\sigma_\delta^2} \sum_{t=1}^{T} \Big(\tilde{\boldsymbol{x}}_t - \alpha_0 \mathbf{1}_{n_G} - \boldsymbol{x}^*_{t,P}\Big)^\intercal \Big(\tilde{\boldsymbol{x}}_t - \alpha_0 \mathbf{1}_{n_G} - \boldsymbol{x}^*_{t,P}\Big) \Big\}.$$

The pseudo-zeroes follows $\mathbf{0}_t \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta} \sim \mathcal{N}\Big(- \boldsymbol{x}_t + \beta_0 \mathbf{1}_{n_M+n_G} + \beta_1 \mathbf{z}_t + \boldsymbol{\xi}_t, \frac{1}{\tau_0} \mathbb{I}_{n_M+n_G}\Big)$, where τ_0 is a precision parameter and is fixed at a large value because of the absence of measurement error in the pseudo-zeroes. Since $\pi\Big(\mathbf{0} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) = \prod_{t=1}^{T} \pi(\mathbf{0}_t \big| \boldsymbol{\chi}_t, \boldsymbol{\xi}_t, \boldsymbol{\theta}_t)$, then

$$\pi\Big(\mathbf{0} \big| \boldsymbol{\chi}, \boldsymbol{\xi}, \boldsymbol{\theta}\Big) \propto \Big(\frac{1}{\tau_0}\Big)^{-\frac{(n_M+n_G)\times T}{2}} \exp\Big\{ - \frac{\tau_0}{2} \sum_{t=1}^{T} \Big(\boldsymbol{x}_t - \beta_0 \mathbf{1}_{n_M+n_G} - \beta_1 \mathbf{z}_t - \boldsymbol{\xi}_t\Big)^\intercal \Big(\boldsymbol{x}_t - \beta_0 \mathbf{1}_{n_M+n_G} - \beta_1 \mathbf{z}_t - \boldsymbol{\xi}_t\Big) \Big\}.$$

The form of the distribution of $\boldsymbol{\xi} \big| \boldsymbol{\theta}$ uses the fact that $\boldsymbol{\xi}_t \big| \boldsymbol{\xi}_{t-1} \sim \mathcal{N}\Big(\varsigma \boldsymbol{\xi}_{t-1}, \boldsymbol{\Sigma}\Big), t = 2, \ldots, T$, and that $\boldsymbol{\xi}_1 \sim \mathcal{N}\Big(\mathbf{0}, \frac{1}{1-\varsigma^2} \boldsymbol{\Sigma}\Big)$. This means that $\pi\Big(\boldsymbol{\xi} \big| \boldsymbol{\theta}\Big) = \pi(\boldsymbol{\xi}_1 \big| \boldsymbol{\theta}) \prod_{t=2}^{T} \pi(\boldsymbol{\xi}_t \big| \boldsymbol{\xi}_{t-1}, \boldsymbol{\theta})$,

so that

$$\pi\Big(\boldsymbol{\xi}|\boldsymbol{\theta}\Big) \propto \Big|\tfrac{1}{1-\varsigma^2}\boldsymbol{\Sigma}\Big|^{-\frac{1}{2}} \exp\Big\{-\frac{1}{2}\boldsymbol{\xi}_1^{\intercal}\Big(\tfrac{1}{1-\varsigma^2}\boldsymbol{\Sigma}\Big)^{-1}\boldsymbol{\xi}_1\Big\} \times \prod_{t=2}^{T}|\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp\Big\{-\frac{1}{2}\Big(\boldsymbol{\xi}_t-\varsigma\boldsymbol{\xi}_{t-1}\Big)^{\intercal}\boldsymbol{\Sigma}^{-1}\Big(\boldsymbol{\xi}_t-\varsigma\boldsymbol{\xi}_{t-1}\Big)\Big\}.$$

The distribution of the extended latent field $\boldsymbol{\chi}$ is as follows:

$$\pi(\boldsymbol{\chi}) = \prod_{t=1}^{T}\pi(\boldsymbol{\chi}_t) = \prod_{t=1}^{T}\pi\Big(\boldsymbol{x}^*_{t,S}\big|\boldsymbol{x}_{t,S}\Big)\pi\Big(\boldsymbol{x}^*_{t,P}\big|\boldsymbol{x}_{t,P}\Big)\pi(\boldsymbol{x}_t).$$

Since $\boldsymbol{x}^*_{t,S}$ and $\boldsymbol{x}^*_{t,P}$ are independent copies of $\boldsymbol{x}_{t,S}$ and $\boldsymbol{x}_{t,P}$, respectively, then both are assumed to be Gaussian centered on $\boldsymbol{x}_{t,S}$ and $\alpha_1\boldsymbol{x}_{t,P}$ and with very high precision, i.e., $\boldsymbol{x}^*_{t,S}\big|\boldsymbol{x}_{t,S} \sim \mathcal{N}\Big(\boldsymbol{x}_{t,M}, \dfrac{1}{\tau_{x^*}}\Big)$ and $\boldsymbol{x}^*_{t,P}\big|\boldsymbol{x}_{t,P} \sim \mathcal{N}\Big(\alpha_1\boldsymbol{x}_{t,P}, \dfrac{1}{\tau_{x^*}}\Big)$, where τ_{x^*} is fixed at some large value. $\pi\big(\boldsymbol{x}_t\big)$ is then assumed to be independent Gaussian centered at zero but with fixed high value for variance (low precision), i.e., $\boldsymbol{x}_t \sim \mathcal{N}\Big(\mathbf{0}, \dfrac{1}{\tau_x}\Big)$, where τ_x is a small value. Although the precision is very low, the pseudo-zeroes have very high precision, and so the value of $\boldsymbol{x}_t$ in Equation (3.9) is forced to be close to its true value. This is necessary since we want to accurately estimate the latent field $\boldsymbol{x}_t$.

Finally, the components of $\boldsymbol{\theta}$ are assumed independent, i.e., $\pi(\boldsymbol{\theta}) = \prod_{i=1}^{H}\pi(\theta_i)$, where H is the number of parameters in $\boldsymbol{\theta}$.

The joint model specified in Equations (3.7) – (3.10) is a correct representation of Equations (3.3) – (3.6) which does not violate latent Gaussianity; hence, allowing the use of INLA for inference.

3.1.2.1 SPDE representation

The Matérn field in the data fusion model is estimated using the SPDE approach (see Section 2.6). The discretization at an arbitrary location $\mathbf{s}$ at time t is given by

$$\omega^D(\mathbf{s},t) = \sum_{k=1}^{K}\psi_k(\mathbf{s})w_{kt}, \tag{3.11}$$

where $\big\{\psi_k\big\}_{\forall k}$ are basis functions, and $\big\{w_{kt}\big\}_{\forall k,t}$ are Gaussian-distributed weights. In Equation (3.11), the basis functions are not indexed by t since the same mesh is used for all time points. The weights $\big\{w_{kt}\big\}_{\forall k,t}$, on the other hand, vary for different time points; hence, the weights are indexed by t. Equation 3.11 provides a continuously-

indexed but finite-dimensional approximation to the Matérn field $\omega(\mathbf{s}, t)$, as discussed in Section 2.6.

Suppose that $\boldsymbol{\omega}_t^D$ denotes the vector of values at the mesh nodes at time t. It follows that $\boldsymbol{\omega}_t^D \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{\mathrm{s}}^{-1}\right)$, where $\mathbf{Q}_{\mathrm{s}}$ is a sparse precision matrix of dimension $K \times K$. Equation (3.6), now defined on the nodes of the mesh, is then expressed as

$$\begin{aligned} &\boldsymbol{\xi}_t^D = \varsigma \boldsymbol{\xi}_{t-1}^D + \boldsymbol{\omega}_t^D, \quad \boldsymbol{\omega}_t^D \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{\mathrm{s}}^{-1}\right), \quad t = 2, \ldots, T, \\ &\boldsymbol{\xi}_1^D \sim \mathcal{N}\left(\mathbf{0}, \tfrac{1}{1-\varsigma^2}\mathbf{Q}_{\mathrm{s}}^{-1}\right) \end{aligned} \tag{3.12}$$

where $\boldsymbol{\xi}_t^D$ is a K-dimensional vector of spatio-temporal random effects at the K mesh nodes at time t. The joint distribution of the $(T \times K)$-dimensional vector $\boldsymbol{\xi}^D = \left(\boldsymbol{\xi}_1^{D\intercal} \quad \ldots \quad \boldsymbol{\xi}_T^{D\intercal}\right)^{\intercal}$ is $\boldsymbol{\xi}^D \sim \mathcal{N}\left(\mathbf{0}, \left(\mathbf{Q}_{\mathrm{s}}^{-1} \otimes \mathbf{Q}_T^{-1}\right)\right)$, where $\mathbf{Q}_T$ is the precision matrix for the autoregressive process of order 1, the form of which is given in Rue and Held (2005).

Since $\boldsymbol{\xi}$ and $\boldsymbol{\omega}$ are estimated in a mesh whose nodes may be different from the spatial locations of the data, there needs to be a linear mapping from the mesh nodes to the locations of the data. This is done by incorporating a projection or mapping matrix, say $\mathbf{B}$, which is a sparse $(n_M + n_G) \times K$ matrix, so that

$$\begin{aligned} &\boldsymbol{x}_t = \beta_0 \mathbf{1}_{n_G+n_M} + \beta_1 \mathbf{z}_t + \mathbf{B}\boldsymbol{\xi}_t^D, t = 1, \ldots, T, \quad \text{or} \\ &x(\mathbf{s}_i, t) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}_i, t) + \sum_{k=1}^{K} b_{ik}\xi_{tk}, \end{aligned}$$

where ξ_{tk} is the kth element of the vector $\boldsymbol{\xi}_t^D$ and b_{ik} is the $(i,k)^{\text{th}}$ element of the mapping matrix $\mathbf{B}$.

The data fusion model, in its SPDE representation, is then given by

$$\begin{aligned} &\mathbf{w}_t = \boldsymbol{x}_{t,S}^* + \boldsymbol{e}_t, \quad \boldsymbol{e}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_e^2 \mathbb{I}_{n_M}\right), \quad t = 1, \ldots, T \\ &\tilde{\boldsymbol{x}}_t = \alpha_0 \mathbf{1}_{n_G} + \boldsymbol{x}_{t,P}^* + \boldsymbol{\delta}_t, \quad \boldsymbol{\delta}_t \sim \mathcal{N}\left(\mathbf{0}, \sigma_\delta^2 \mathbb{I}_{n_G}\right) \\ &\mathbf{0}_t = - \begin{pmatrix} \boldsymbol{x}_{t,S} \\ \boldsymbol{x}_{t,P} \end{pmatrix} + \beta_0 \mathbf{1}_{n_M+n_G} + \beta_1 \mathbf{z}_t + \mathbf{B}\boldsymbol{\xi}_t^D \\ &\boldsymbol{\xi}_t^D = \varsigma \boldsymbol{\xi}_{t-1}^D + \boldsymbol{\omega}_t^D, \quad \boldsymbol{\omega}_t^D \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{\mathrm{s}}^{-1}\right), \quad t = 2, \ldots, T \end{aligned} \tag{3.13}$$

The joint model specified in Equations (3.13) is a latent Gaussian model and, therefore, can be fitted using the INLA method. The latent parameters are $\left\{\beta_0, \beta_1, \boldsymbol{\xi}^D, \boldsymbol{\omega}^D\right\}$ and the SPDE weights $\left\{w_{kt}\right\}_{\forall k,t}$. The hyperparameters are $\left\{\sigma_e^2, \sigma_\delta^2, \sigma_\omega^2, \rho, \varsigma\right\}$.

3.2 Application in spatial epidemiology

The models and estimation strategies discussed in Sections 3.1.1 and 3.1.2 are considered the initial step in a two-stage modeling framework (see Figure 1.6). There are two main model structures: one for the environmental process of interest, such as pollution concentration, climate conditions, etc., which is referred to as the *first-stage model*, and where the data fusion model presented in Section 3.1 is relevant; and another model to link the predictions from the first-stage model and observed health outcomes, which is referred to as the *second-stage model*.

3.2.1 Second-stage model

The second-stage model specifies the model for the health outcomes as the response variable and the spatial average of $x(\mathbf{s}, t)$ in area B_i at time t, denoted by $x(B_i, t)$, as input to the model. I assume that the (true) block-level value of $x(\cdot, t)$ at block B_i at time t, denoted by $x(B_i, t)$, is a spatial average over B_i, i.e.,

$$x(B_i, t) = \frac{1}{|B_i|} \int_{B_i} x(\mathbf{s}, t) d\mathbf{s}. \tag{3.14}$$

Equation (3.14) is a special case of Equation (2.1), where all points $\mathbf{s}$ in B_i are equally weighted. The observed count at a block B_i at time t, denoted by $\mathrm{y}(B_i, t)$, is assumed to be a Poisson outcome with mean $\mathbb{E}\Big[\mathrm{y}(B_i, t)\Big] = \mu_{\mathrm{y}}(B_i, t)$, i.e.,

$$\mathrm{y}(B_i, t) \sim \text{Poisson}\Big(\mu_{\mathrm{y}}(B_i, t)\Big), \quad \mu_{\mathrm{y}}(B_i, t) = \mathrm{E}(B_i, t) \times \lambda(B_i, t), \tag{3.15}$$

where $\mathrm{E}(B_i, t)$ is the expected number of cases in area B_i at time t (Shaddick et al., 2023; Waller and Carlin, 2010; Waller and Gotway, 2004). $\lambda(B_i, t)$ is the disease risk

modelled as

$$\log\Big(\lambda(B_i,t)\Big) = \gamma_0 + \gamma_1 x(B_i,t) + \varphi_{it},$$

where γ_0 is the intercept, γ_1 is the coefficient of the block-level exposure $x(B_i,t)$, and φ_{it} is a spatio-temporal random effect for block B_i and time t. The spatio-temporal random effect accounts for extra variation in the respnse variable that cannot be explained solely by the covariates, and accounts for unmeasured factors that vary across space and/or time. A conditional independence assumption (see Section 2.4.1) given the effects is assumed for the Poisson outcome, i.e.,

$$Y(B_i,t)\big|\gamma_0,\gamma_1,\varphi_{it} \overset{\text{ind}}{\sim} \text{Poisson}\Big(\mu_{\text{y}}(B_i,t) = \text{E}(B_i,t) \times \lambda(B_i,t)\Big). \tag{3.16}$$

The spatio-temporal random effect term can take several forms, following Knorr-Held (2000) and Blangiardo and Cameletti (2015). A general form is given by

$$\log\Big(\lambda(B_i,t)\Big) = \gamma_0 + \gamma_1 x(B_i,t) + \phi_i + \psi_i + \zeta_t + \nu_t + \upsilon_{it}, \tag{3.17}$$

where ϕ_i is an iid spatial random effect, i.e., $\phi_i \overset{\text{iid}}{\sim} \mathcal{N}\Big(0,\sigma^2_\phi\Big)$, ψ_i is a structured spatial random effect such as the conditional autoregressive process (Besag et al., 1991) discussed in Section 2.3, ζ_t is an iid temporal random effect, i.e, $\zeta_t \overset{\text{iid}}{\sim} \mathcal{N}\Big(0,\sigma^2_\zeta\Big)$, ν_t is a structured temporal random effect such as the random walk model, and υ_{it} is a spatio-temporal interaction effect. For the interaction effect, Knorr-Held (2000) proposed four types, depending on which of the two spatial and time effects, either the structured or the unstructured, interact. These different ways to specify the spatio-temporal structure of the model do not violate latent Gaussianity, and hence all posterior marginals of the model parameters can be estimated using the INLA method (Blangiardo and Cameletti, 2015).

Here, the latent parameters are $\Big\{\gamma_0,\gamma_1,\boldsymbol{\varphi}\Big\}$, where $\boldsymbol{\varphi}$ is the vector of all the spatial effects, temporal effects, and their interaction. The hyperparameters are all the model constants used to parameterize the components of $\boldsymbol{\varphi}$, say $\boldsymbol{\theta}_\varphi$. For instance, if a random walk model of order 1 is used for the temporal effect, then the variance

parameter for the random walk process is included in the hyperparameter vector. The posterior distribution of interest is

$$\pi\Big(\gamma_0, \gamma_1, \boldsymbol{\varphi}, \boldsymbol{\theta}_\varphi | \mathbf{y}\Big) \propto \pi\Big(\mathbf{y} | \gamma_0, \gamma_1, \boldsymbol{\varphi}, \boldsymbol{\theta}_\varphi\Big) \pi\Big(\boldsymbol{\varphi} | \boldsymbol{\theta}_\varphi\Big) \pi\Big(\gamma_0, \gamma_1, \boldsymbol{\theta}_\varphi\Big),$$

where $\mathbf{y} = \big\{\mathrm{y}(B_i, t)\big\}_{\forall i,t}$. $\pi\Big(\mathbf{y} | \gamma_0, \gamma_1, \boldsymbol{\varphi}, \boldsymbol{\theta}_\varphi\Big)$ has a simple form since the elements of $\mathbf{y}$ are independent conditional on all model unknowns. The form of $\pi\Big(\boldsymbol{\varphi} | \boldsymbol{\theta}_\varphi\Big)$ depends on the model specified for the elements of $\boldsymbol{\varphi}$ and has a straightforward structure (Blangiardo and Cameletti, 2015). Lastly, $\pi\Big(\gamma_0, \gamma_1, \boldsymbol{\theta}_\varphi\Big)$ are assumed to be a product of the individual priors.

3.2.2 Computing block-level estimates

I explore two methods to compute the spatial averages introduced in Cameletti et al. (2019). The first method considers all the prediction grid cells which overlap with block B_i. Suppose that $\hat{x}\big(\mathbf{s}_j^*, t\big)$ denotes the posterior mean of $x\big(\mathbf{s}_j^*, t\big)$, and $h\big(\mathbf{s}_j^*, B_i\big)$ is the proportion of the area of block B_i which overlaps with the grid cell with centroid $\mathbf{s}_j^*$. The first method is computed as follows:

$$\textbf{Method 1}: \qquad \hat{x}\big(B_i, t\big) = \sum_{\forall j} \hat{x}\big(\mathbf{s}_j^*, t\big) h\big(\mathbf{s}_j^*, B_i\big). \tag{3.18}$$

The first method computes the value of $\hat{x}\big(B_i, t\big)$ as a weighted mean of the predicted values in the grid cells that overlap with the block B_i, where the weights are the proportion of the block B_i that overlap with the grid cells. On the other hand, the second method uses only the grid cells whose centroids are inside block B_i. The computed value of $\hat{x}\big(B_i, t\big)$ using the second method is then a simple mean, i.e.,

$$\textbf{Method 2}: \qquad \hat{x}\big(B_i, t\big) = \frac{1}{\#(\mathbf{s}_j^* \in B_i)} \sum_{\mathbf{s}_j \in B_i} \hat{x}\big(\mathbf{s}_j^*, t\big), \tag{3.19}$$

where $\#(\mathbf{s}_j^* \in B_i)$ is the number of grid cells whose centroids are inside block B_i.

In the simulation study performed in Cameletti et al. (2019), the results show that the choice of either Method 1 or Method 2 does not have a substantial effect on

the risk parameter γ_1 in terms of the bias and RMSE. However, this would depend on the amount of local spatial variability. Method 1 averages across more points, so this method can give more stable estimates if there is high spatial variability within and between areas.

3.2.3 Uncertainty propagation

One approach to propagate uncertainty from the first-stage to the second-stage model is to sample several times from the estimated posterior predictive distribution of the latent field, $\hat{\pi}\left(x(\mathbf{s}_j^*, t)\middle|\mathbf{w}, \tilde{\boldsymbol{x}}\right)$, $\forall \mathbf{s}_j^*$ in the prediction grid and $\forall t$, and then compute the spatial averages for each sample using either Equation (3.18) or (3.19). The spatial averages are then used as an input in the second-stage model, which is fit multiple times. The final parameter estimates of the second-stage model are then computed using the combined results from all samples, via model averaging (Blangiardo et al., 2016; Cameletti et al., 2019; Lee et al., 2017; Liu et al., 2017). Algorithm 3.1 summarizes the steps to propagate uncertainty from the first-stage to the second-stage model.

Algorithm 3.1 Uncertainty propagation approach

Repeat the following J times:

Step 1: Simulate from the posterior predictive distribution of the latent field, $\hat{\pi}\left(x(\mathbf{s}_j^*, t)\middle|\mathbf{w}, \tilde{\boldsymbol{x}}\right)$, $\forall \mathbf{s}_j^*$ in the prediction grid and $\forall t$.

Step 2: Compute block-level exposures, $\hat{x}(B_i, t)$, $\forall B_i$ in the study region and $\forall t$, using the two methods in Equations (3.18) and (3.19), and using the obtained samples in Step 1.

Step 3: Fit the second-stage model using INLA as described in Section 3.2.1, using the computed block-level values in Step 2 as input.

Step 4: Store all posterior results from Step 3.

After completing all J cycles, all samples from step 4 in Algorithm 3.1 are combined or pooled together to approximate the posterior distribution of the second-stage model parameters, i.e., performing Bayesian model averaging. As an example, for the

parameter γ_1, the final posterior marginal is computed as

$$\hat{\pi}\Big(\gamma_1\big|\mathbf{w},\tilde{\boldsymbol{x}}\Big) = \frac{1}{J}\sum_{j=1}^{J}\hat{\pi}\Big(\gamma_1\big|\mathbf{w},\tilde{\boldsymbol{x}},\boldsymbol{x}_j^*\Big), \tag{3.20}$$

where $\boldsymbol{x}_j^*$ is the j^{th} sample from the estimated posterior $\hat{\pi}\Big(x(\mathbf{s}_j^*,t)\big|\mathbf{w},\tilde{\boldsymbol{x}}\Big)$ (see Step 1 of Algorithm 3.1). Equation (3.20) is an approximation to

$$\pi\Big(\gamma_1\big|\mathbf{w},\tilde{\boldsymbol{x}}\Big) = \int \pi\Big(\gamma_1\big|\mathbf{w},\tilde{\boldsymbol{x}},\boldsymbol{x}\Big)\pi\Big(\boldsymbol{x}\big|\mathbf{w},\tilde{\boldsymbol{x}}\Big)d\boldsymbol{x}, \tag{3.21}$$

which integrates out the uncertainty in the first-stage posterior distribution $\pi\Big(\boldsymbol{x}\big|\mathbf{w},\tilde{\boldsymbol{x}}\Big)$.

3.3 Simulation Study

The performance of the proposed two-stage method is investigated using a simulation study. The study region used is the Belo Horizonte region in Brazil, which is available in the `spdep` package in `R` (Bivand and Piras, 2015), and is the same study region used in Cameletti et al. (2019) on which the extensions done in this work are based. This study region is used since it has few areas (98 of them), which makes it more computationally manageable. Figure 3.1, which is a simulated spatio-temporal data, shows the study region map.

3.3.1 Simulating from the first-stage model

For the Matèrn covariance function parameters, the spatial variance σ_ω^2 is set to 1.5, while the range parameter ρ is 1.89, which corresponds to around 46% of the maximum distance in the 100×100 simulation grid. The autoregressive parameter ς is set to 0.7. The single covariate $\mathrm{z}(\mathbf{s},t)$ was generated from $\mathcal{N}(\mu=0,\sigma=1)$. The fixed effects are $\beta_0 = 0$ and $\beta_1 = 2$. The simulated values of $x(B_i,t)$, as shown in Equation (3.14), is computed as follows:

$$x(B_i,t) = \frac{1}{|B_i|}\int_{B_i} x(\mathbf{s},t)d\mathbf{s} \approx \frac{1}{\#(\mathbf{s}\in B_i)}\sum_{\forall \mathbf{s}\in B_i} x(\mathbf{s},t).$$

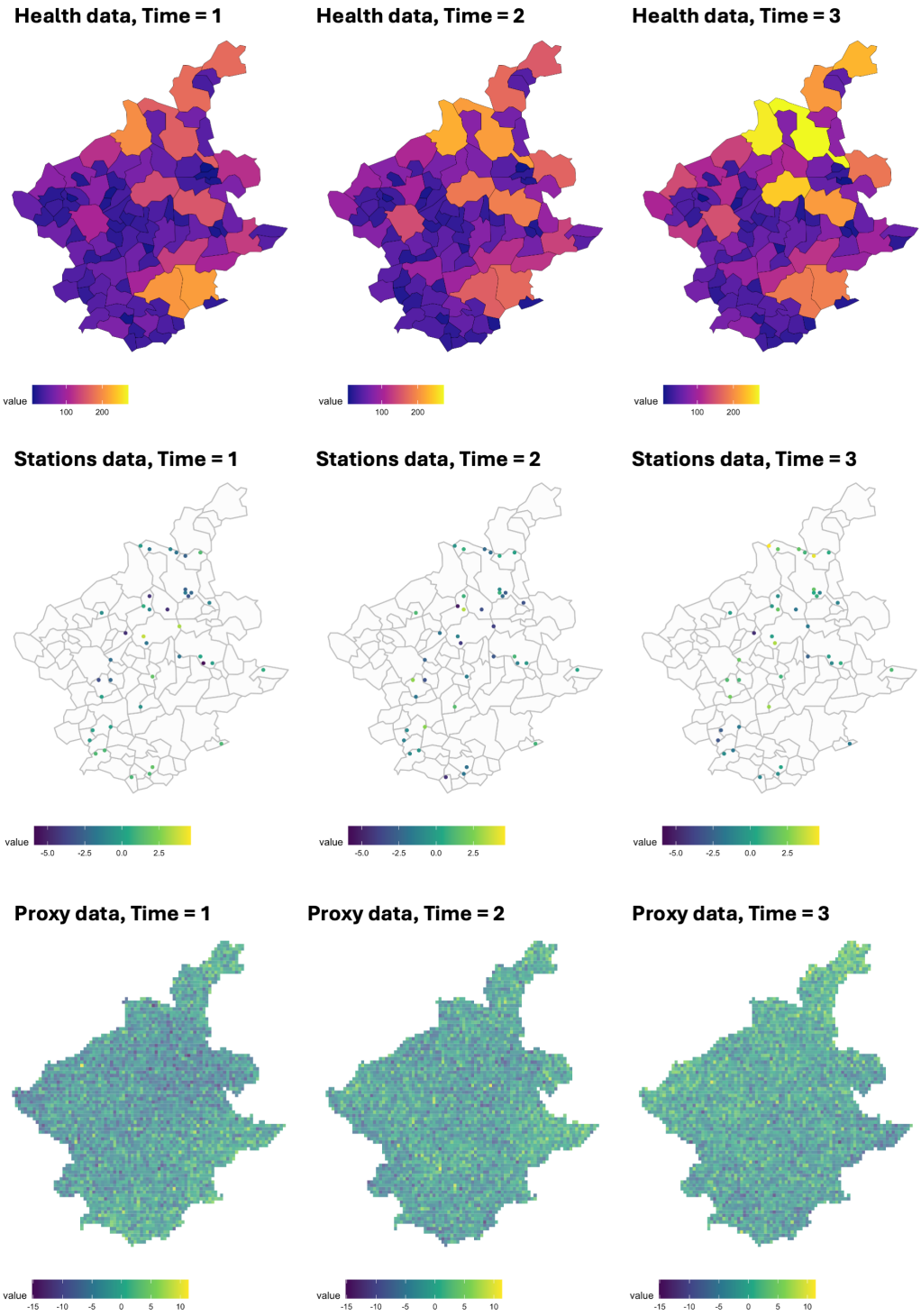


Figure 3.1: Sample data for a spatio-temporal analysis: count data (top row), stations data (middle row), proxy data (bottom)

3.3.2 Simulating from the second-stage model

I assume a simple form for the log risks, as follows:

$$\log\Big(\lambda(B_i,t)\Big) = \gamma_0 + \gamma_1 x(B_i,t) + \phi_i + \nu_t, \tag{3.22}$$

where ϕ_i is an iid random effect in space and ν_t follows a random walk model of order 1 in time, i.e., $\phi_i \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_\phi^2\right)$ and $\nu_t|\nu_{t-1} \sim \mathcal{N}\left(\nu_{t-1}, \sigma_\nu^2\right)$. Here, I assume that φ_{it} is equal to the sum of an unstructured spatial effect and a structured temporal effect. The latent parameters are $\left\{\gamma_0, \gamma_1, \boldsymbol{\phi}, \boldsymbol{\nu}\right\}$, while the hyperparameters are $\left\{\sigma_\phi^2, \sigma_\nu^2\right\}$.

The fixed effects are $\gamma_0 = -3$ and $\exp(\gamma_1) = 1.2$. The assumed value of γ_1 implies an expected increase of 20% in the relative risk for a one unit increase in $x(B_i, t)$. The assumed values of the variance parameters of the spatial effect and the temporal effect are $\sigma_\phi^2 = \sigma_\nu^2 = 0.02$. The expected number of cases $\mathrm{E}(B_i, t)$ are generated from a uniform distribution, and are made to be proportional to the size of the block, so that blocks with bigger surface areas have higher expected number of cases.

3.3.3 Simulation of the stations and proxy data

The stations are simulated by generating a random sample of points from the simulation grid. A non-sparse network was considered and investigated in Cameletti et al. (2019) - either getting 2%, 10%, or 30% of the points from the simulation grid inside each block. Their simulation results showed that block predictions are better in terms of RMSE and correlation with true block-level exposure values when there are more stations. In this simulation study, I consider two scenarios for the stations data: the first scenario is a non-sparse network, while the second scenario is a sparse network, i.e., there are few stations and several areas or blocks without stations inside, and is carefully chosen in such a way that it resembles real life sparse data. Figure 3.2 shows a simulated non-sparse network (left) and a sparse network (right) of stations. The observed values $\mathrm{w}(\mathbf{s}_i, t), i = 1, \ldots, n_M$ follow the classical error model, with the error term assumed as $e(\mathbf{s}_i, t) \overset{\text{iid}}{\sim} \mathcal{N}\left(0, \sigma_e^2 = 0.1\right)$. For the proxy data, the bias parameters are $\alpha_0 = -1, \alpha_1 = 1.5$, and $\sigma_\delta^2 = 1$.

3.3.4 Prediction grid

The effect of the resolution of the prediction grid on the block-level predictions using Equations (3.18) and (3.19) has been investigated in Cameletti et al. (2019). Their simulation results have shown that with a finer prediction grid, the block-level exposure predictions are also more accurate. Hence, in the simulation study in this paper,

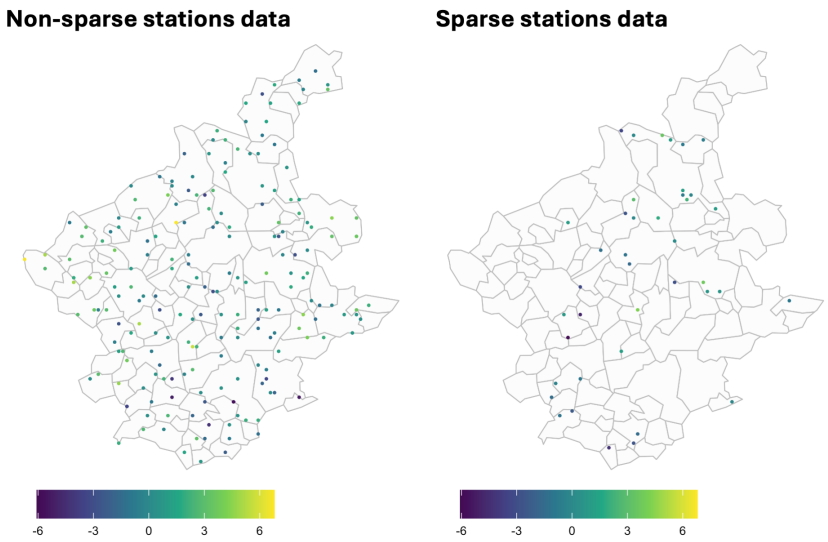


Figure 3.2: Non-sparse network of stations (left) and a sparse network of stations (right)

a prediction grid of 100×100 is used for all scenarios. This is also the same grid resolution used to simulate the true field $x(\mathbf{s}, t)$.

3.3.5 Simulation scenarios

There are three simulation settings considered in the study:

1. The sparsity of the stations data: sparse or non-sparse. This is illustrated in Figure 3.2.

2. Length of time: $T = 3$, $T = 6$, or $T = 12$. Since this simulation study is a spatio-temporal extension of that in Cameletti et al. (2019), it is important to investigate the effect of the length of time in terms of the model estimates and predictions.

3. Prior specification: use of non-informative priors or (weakly) informative priors. Since the sensitivity of the posterior estimates to the priors is an important component of Bayesian analysis, the effect of prior specification is therefore investigated in this simulation study. Only those parameters which are usually difficult to estimate are given (weakly) informative priors - these include the parameters of the Matérn field σ^2_ω, ρ, and ς; the variance parameters $\sigma^2_\nu, \sigma^2_\phi, \sigma^2_e$, and σ^2_δ; and α_1, which is the scaling parameter of $\boldsymbol{x}_{t,P}$.

There are a total of 12 simulation scenarios. For each scenario, 500 independent data replicates are generated in order to evaluate the performance of the proposed method. In fitting the second-stage model, the number of simulations from the posterior predictive distribution of the latent field $x(\mathbf{s}, t)$ is $J = 100$ (see Algorithm 3.1). For each estimated posterior marginal distribution, 200 random values were simulated to compute posterior quantities of interest, which include the posterior mean, posterior median, and 95% credible intervals. Table 3.1 shows all the simulation scenarios and how they are labelled in the figures in the succeeding sections.

Sparsity	**Priors specification**	**Length of time**		
		3	6	12
No	Informative	A	E	I
	Non-informative	B	F	J
Yes	Informative	C	G	K
	Non-informative	D	H	L

Table 3.1: Simulation scenarios in the simulation study

Table 3.2 shows the priors for the first-stage model parameters for the two cases. Only non-informative priors are used for the fixed effects β_0, β_1, and α_0. The non-informative priors of the Matèrn parameters are defined in terms of a reparameterization of the parameters of the SPDE in Equation (2.34); a detailed discussion of which is in Lindgren and Rue (2015). The (weakly) informative priors for ρ and σ_ω^2 are the so-called penalized-complexity (PC) priors (Fuglstad et al., 2019; Simpson et al., 2017). The PC prior penalizes complexity or additional flexibility in the model, i.e., the prior tends to prefer the simpler base model. It works on the principle that a model further away from the base model should be more strongly penalized. The PC prior is defined using probability statements on the model parameters in the appropriate scale. For the Matèrn parameters, the PC prior shrinks the model to the base model with infinite range and zero marginal variance. The PC priors for ρ and σ_ω^2, shown in Table 3.2, are a joint specification, where $\sigma_{\omega,0}$ and ρ_0 are the upper and lower limit for σ_ω and ρ, respectively, and α is the tail probability. In fitting the models in the simulation study, $\sigma_{\omega,0}$ and ρ_0 are set as equal to the true value, and $\alpha = 0.5$.

There are three precision parameters in the first-stage model that are fixed at an appropriate level. The first one is the precision parameter for the pseudo-zeroes, τ_0,

Parameter	Informative Prior	Non-informative Prior
β_0	-	$\mathcal{N}(0, \infty)$
β_1	-	$\mathcal{N}(0, \infty)$
α_0	-	$\mathcal{N}(0, \infty)$
α_1	$\mathcal{N}(1.5, 1)$	$\mathcal{N}(0, 1000)$
σ_e^2	Inv Gamma$(100, 10)$	Inv Gamma$(1, 5e{-}5)$
σ_δ^2	Inv Gamma$(10, 10)$	Inv Gamma$(1, 5e{-}5)$
ρ	PC$(\rho_0, \alpha)^*$	see Lindgren and Rue (2015)
σ_ω^2	PC$(\sigma_{\omega,0}, \alpha)^*$	see Lindgren and Rue (2015)
ς	$\log\left(\frac{1+\varsigma}{1-\varsigma}\right) \sim \mathcal{N}(0.7, 0.05^2)$	$\log\left(\frac{1+\varsigma}{1-\varsigma}\right) \sim \mathcal{N}(0, 0.15^2)$

* These are called **penalized-complexity priors**, which are weakly informative priors (Fuglstad et al., 2019; Simpson et al., 2017).

Table 3.2: Priors specification for first-stage model parameters

which is fixed at a large value. The second one is the precision parameter, τ_x, for the prior of the latent field $\boldsymbol{x}_t$, which is fixed at a very small value. Finally, for the conditional distribution of the copies $\boldsymbol{x}_{t,S}^*$ and $\boldsymbol{x}_{t,P}^*$, the precision parameter τ_x* is also fixed at a large value, so that both mimic $\boldsymbol{x}_{t,S}$ and $\alpha_1 \boldsymbol{x}_{t,P}$, respectively.

Parameter	Informative Prior	Non-informative Prior
γ_0	-	$\mathcal{N}(0, \infty)$
γ_1	-	$\mathcal{N}(0, 1000)$
σ_ϕ^2	PC$(\sigma_{\phi,0}, \alpha)^*$	Inv Gamma$(1, 5e{-}5)$
σ_ν^2	PC$(\sigma_{\nu,0}, \alpha)^*$	Inv Gamma$(1, 5e{-}5)$

* These are called **penalized-complexity priors**, which are weakly informative priors (Fuglstad et al., 2019; Simpson et al., 2017).

Table 3.3: Priors specification for second-stage model parameters

Table 3.3 shows the priors for the second-stage model parameters for the two cases. The parameters γ_0 and γ_1 are given only non-informative priors. The variance parameters σ_ϕ^2 and σ_ν^2 are given the PC priors for the case of (weakly) informative priors. Here, I also used the true values for $\sigma_{\phi,0}$ and $\sigma_{\nu,0}$, and $\alpha = 0.5$.

3.3.6 Model evaluation

The performance of the method is evaluated by looking at the bias and root mean square error (RMSE) of the parameter estimates. There is a slight difference in the notation for the formulas for calculating the model performance metrics between the first-stage and second-stage model parameters. For a first-stage model parameter, say θ, $\hat{\theta}_{ik}$ denotes the kth sampled value from the estimated posterior distribution of θ in the ith simulated data, $i = 1, \ldots, n_{\text{sim}}$ and $k = 1, \ldots, K$. For a second-stage model parameter θ, $\hat{\theta}_{ijk}$ denotes the kth sampled value from the estimated posterior distribution of θ using the block-level estimates $\hat{x}(B, t)$ computed using the jth simulated values from the marginal posterior distribution of the latent field $x(\mathbf{s}, t)$, $j = 1, \ldots, J$, $k = 1, \ldots, K$, $i = 1, \ldots, n_{\text{sim}}$. In the simulation study, $n_{\text{sim}} = 500, J = 100$ and $K = 200$.

1. **Bias** - The bias is computed as

$$\text{bias} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \left(\frac{1}{K} \sum_{k=1}^{K} \hat{\theta}_{ik} - \theta \right) \quad \text{or} \quad \text{bias} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \left(\frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} \hat{\theta}_{ijk} - \theta \right)$$

   for a first-stage and second-stage model parameter, respectively.

2. **Root mean square error (RMSE)** - The RMSE is computed as

$$\text{RMSE} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \sqrt{\frac{1}{K} \sum_{k=1}^{K} (\hat{\theta}_{ik} - \theta)^2} \quad \text{or}$$

$$\text{RMSE} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \sqrt{\frac{1}{JK} \sum_{j=1}^{J} \sum_{k=1}^{K} (\hat{\theta}_{ijk} - \theta)^2}$$

   for a first-stage and second-stage model parameter, respectively.

The same model evaluation metrics are used for the block-level exposure estimates. It should be noted that the true values of the block-level exposures vary for the different data replicates.

3.3.7 Simulation results

This section presents the results of the simulation study. Section 3.3.7.1 presents the results for the first-stage model parameters, Section 3.3.7.2 for the block-level estimates of the first-stage latent field, and Section 3.3.7.3 for the second-stage model parameters.

3.3.7.1 First-stage model parameters

Figures 3.3 and 3.4 show plots of the biases and RMSEs for the additive and multiplicative bias in the proxy data, α_0 and α_1, respectively. These parameters are important since both account for the biases or systematic errors in the proxy data. The obtained estimates of these bias parameters is not only used to recover the true latent process $x(\mathbf{s}, t)$, but also to calibrate or de-noise the proxy data. The results show that the biases are generally close to zero for all scenarios. However, the RMSEs are higher when the stations data is sparse (scenarios C, D, G, H, K, L); but the values are decreasing with more time points. The proposed method is able to correctly estimate α_1, which is usually difficult to estimate, since it is a scaling parameter of the latent field $x(\mathbf{s}, t)$. There is no difference in the bias and the RMSE for α_1 between the use of non-informative prior or (weakly) informative priors.

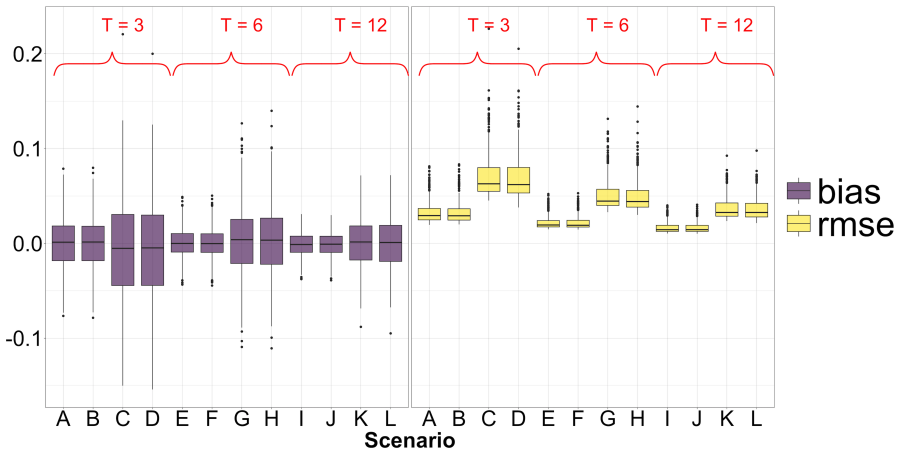


Figure 3.3: Plot of bias (purple) and RMSE (yellow) for α_0

Figures 3.5 and 3.6 show the biases and RMSEs for the measurement error variance of the stations and proxy data, σ_e^2 and σ_δ^2, respectively. Both parameters are important since they describe the amount of random noise or the precision of the

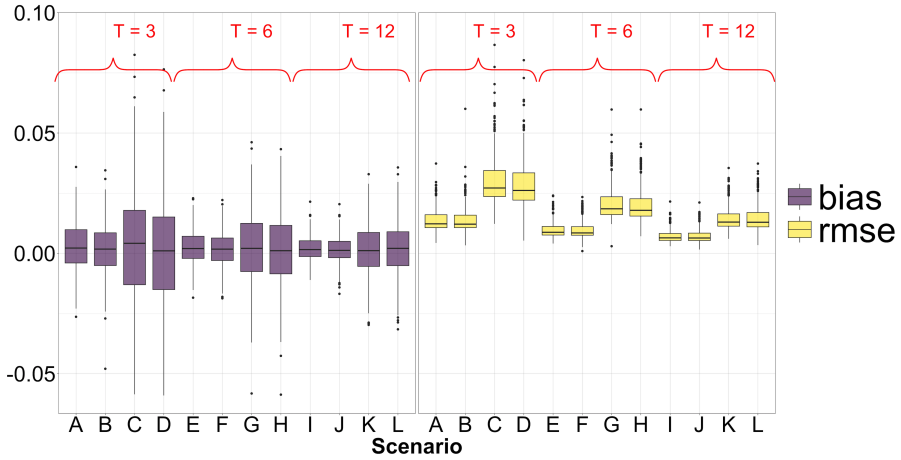


Figure 3.4: Plot of bias (purple) and RMSE (yellow) for α_1

two data sources. Accurate estimates of these parameters also imply good estimates for the latent field $x(\mathbf{s}, t)$. For σ_e^2, the biases are close to zero for all scenarios, but the RMSEs are generally higher when the stations data is sparse and when non-informative priors are used (scenarios D, H, L). This is expected since with few data from the stations, there is also little information at hand to do the estimation. But for as long as informative priors are used or there are several time points, the RMSEs are generally lower, even if the data on the stations is sparse. For σ_δ^2, the biases are also close to zero for all scenarios. The RMSEs are also generally small, and is decreasing with more time points. The sparsity in the stations data and the specification of priors does not seem to affect the bias and the RMSE for σ_δ^2.

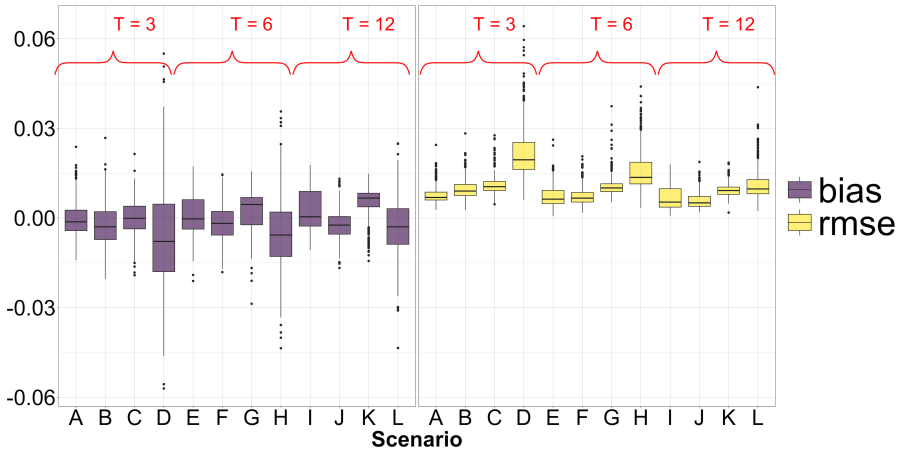


Figure 3.5: Plot of bias (purple) and RMSE (yellow) for σ_e^2

Figures 3.7a and 3.7b show the absolute bias and RMSEs for the Matèrn marginal variance σ_ω^2. The biases and RMSEs are clearly smaller when (weakly) informative

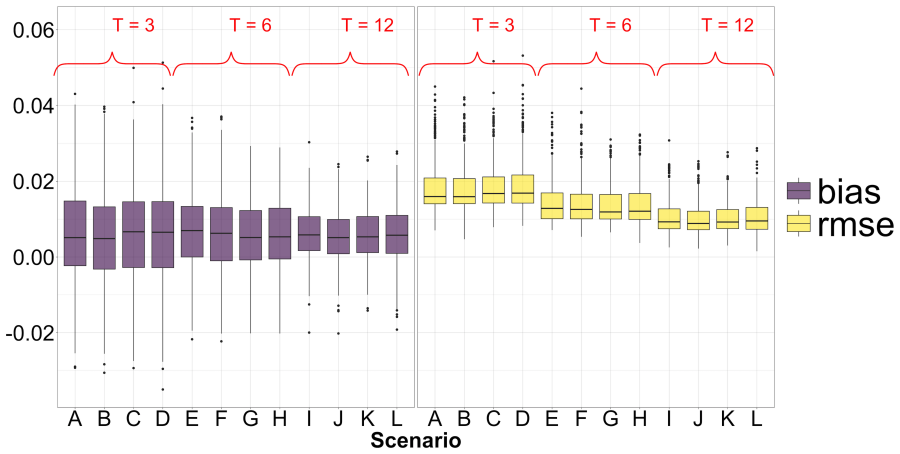
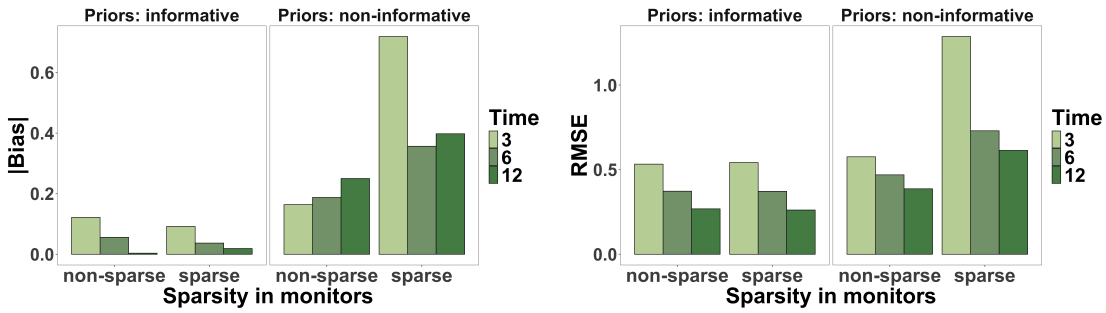


Figure 3.6: Plot of bias (purple) and RMSE (yellow) for σ_δ^2

priors are used, which is the expected result since the Matèrn parameters are typically difficult to estimate. Even if the stations data is sparse, the biases and RMSEs are still small as long as informative priors are used. This is expected since with few stations, we would rely on informative priors to correctly estimate the Matèrn parameters. When non-informative priors are used and there are very few stations, the bias and RMSEs are expected to be high.



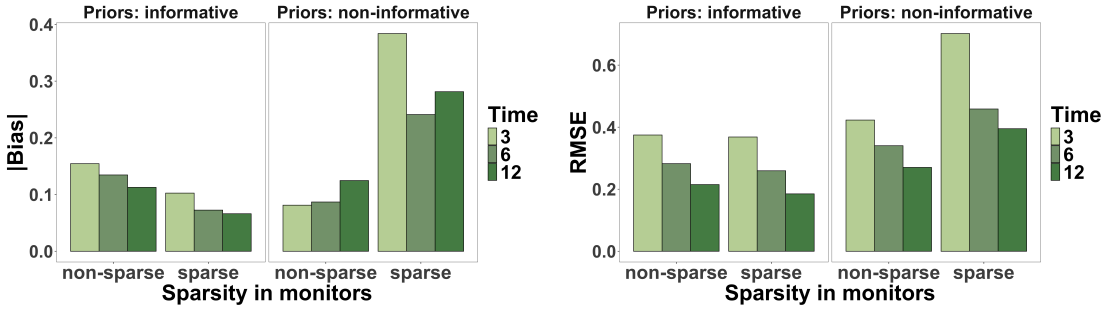
(a) Bias

(b) RMSE

Figure 3.7: Plot of bias and RMSE for the Matèrn marginal variance σ_ω^2

Figures 3.8a and 3.8b show the absolute bias and RMSE for the Matèrn range parameter ρ. The patterns here are similar to what Figures 3.7a and 3.7b show. When the stations data is sparse and the priors are non-informative, the (absolute) bias and RMSE are expected to be high.

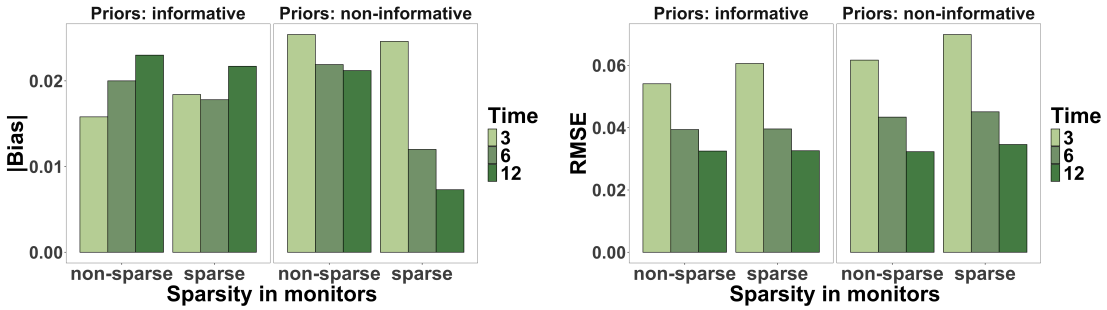
Figures 3.9a and 3.9b show the absolute bias and RMSE for the temporal parameter ς. For this parameter, the biases and RMSEs are generally close to zero for all scenarios. The RMSEs are relatively higher when there are fewer time points. This is expected since ς parametrizes the temporal evolution of the spatial field, so



(a) Bias (b) RMSE

Figure 3.8: Plot of bias and RMSE for the Matèrn range parameter ρ

that more time points means more information available for estimation. Even with non-informative priors, for as long as there are relatively more time points, the ς parameter is relatively well estimated.



(a) Bias (b) RMSE

Figure 3.9: Plot of bias and RMSE for the temporal autoregressive parameter ς

Lastly, shown in Figures 3.10 and 3.11 are the biases and RMSEs for β_0 and β_1. For both parameters, the biases are close to zero for all scenarios. However, when the data on the stations is sparse (scenarios C, D, G, H, K, L), the RMSE of β_1 is relatively larger. The parameter β_1 is the coefficient of the covariate of the latent field $x(\mathbf{s}, t)$. Although this information is also available for the high-resolution proxy data, it is noisier and less correlated with the true values of the latent field. Hence, this could be the reason for the difficulty in estimating it correctly when the stations data is sparse. For more time points, the RMSEs for both β_0 and β_1 are generally smaller, since more time points means more information available for estimation.

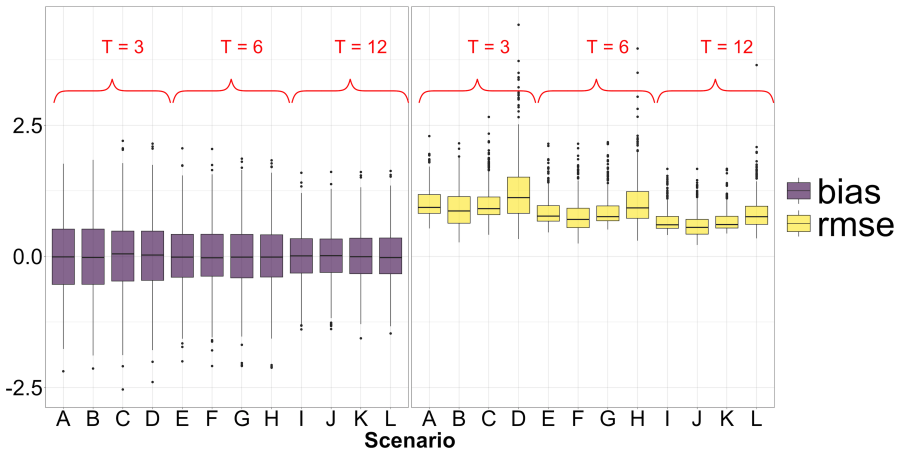


Figure 3.10: Plot of bias (purple) and RMSE (yellow) for β_0

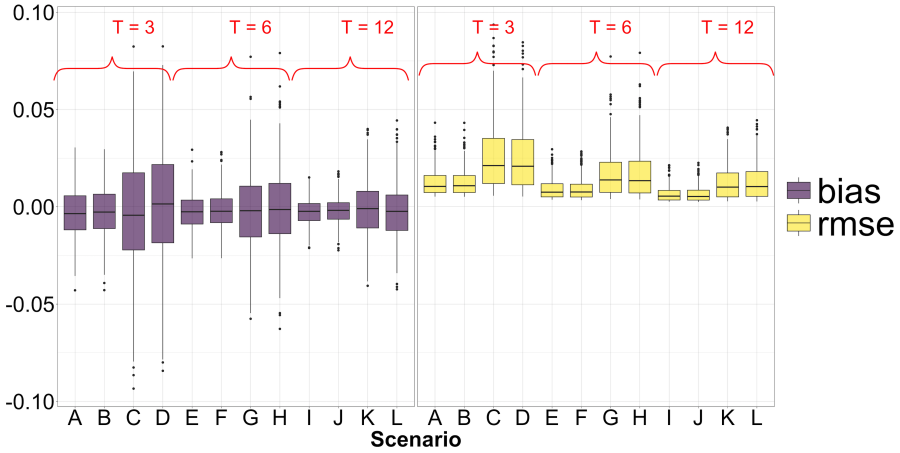


Figure 3.11: Plot of bias (purple) and RMSE (yellow) for β_1

3.3.7.2 Block-level estimates

Figure 3.12 shows the average of the correlations between the block-level estimates $\hat{x}(B_i, t)$ and the corresponding true values for all scenarios. Each point in the plot corresponds to a block B_i in the study region. Since there are T time points, the values shown in the plot are the average of the correlations from the T time points, which are further averaged from all 500 replicates. All the (average) correlations range from around 0.97 to some value close to 1.0, but using Method 2 for computing spatial averages generally gave higher correlations than Method 1 for all scenarios. Also, the correlations are generally higher when there are more time points, which is true for both methods.

Figures 3.13a and 3.13b show the biases and RMSEs, respectively, in $\hat{x}(B_i, t)$ for all scenarios. Similar to Figure 3.12, each point in the plot corresponds to a block

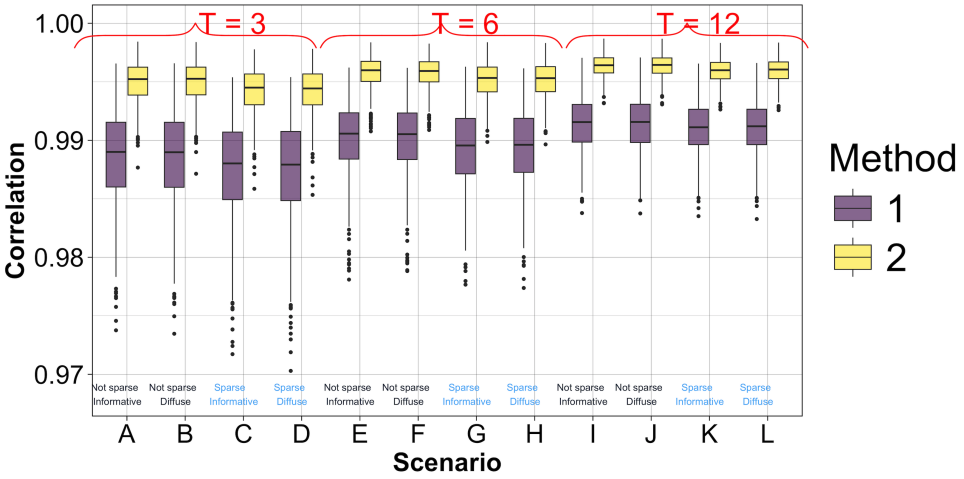
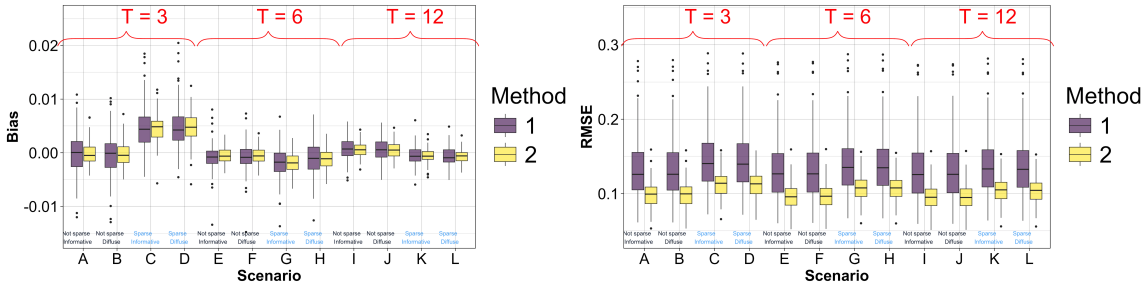


Figure 3.12: Plot of correlations of true and estimated bock-level values $x(B_i, t)$ for all scenarios

B_i. I first compute the bias and RMSE for each block at a given time point. The values shown in Figures 3.13a and 3.13b are then the average of the values for all time points T. The biases are generally close to zero; but the spread in the biases are wider for Method 1 than Method 2. Also, when the stations data is sparse and the priors are non-informative, the biases are generally larger, especially for scenarios with fewer time points. This pattern observed for the biases is also true for the RMSEs. There are a couple of areas with very high RMSEs when using Method 1, and this is consistent for all scenarios. The RMSEs from using Method 1 are generally higher compared to Method 2 for all scenarios.



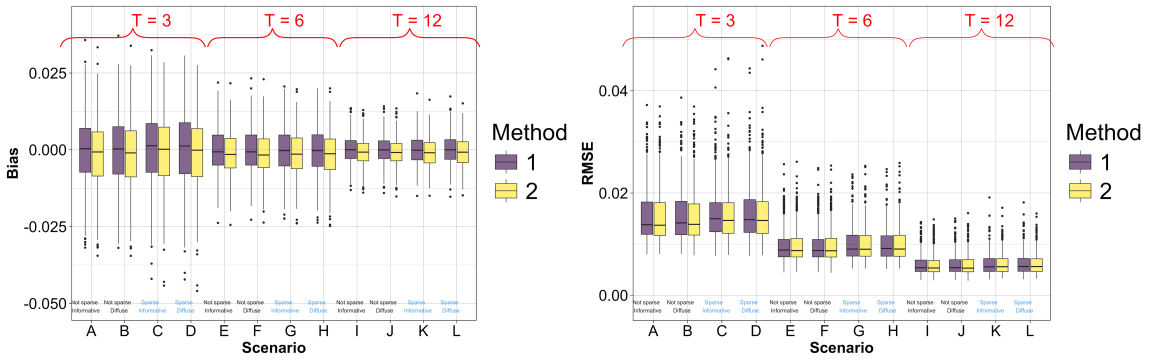
(a) Bias **(b)** RMSE

Figure 3.13: Plot of bias and RMSE for bock-level estimates $\hat{x}(B_i, t)$ for all scenarios

3.3.7.3 Second-stage model parameters

The main parameter of interest is γ_1 since it quantifies the relationship between the first-stage latent process $x(\mathbf{s}, t)$ and the health outcome. Figures 3.14a and 3.14b show the biases and the RMSEs, respectively, for γ_1 for all scenarios. There is no striking

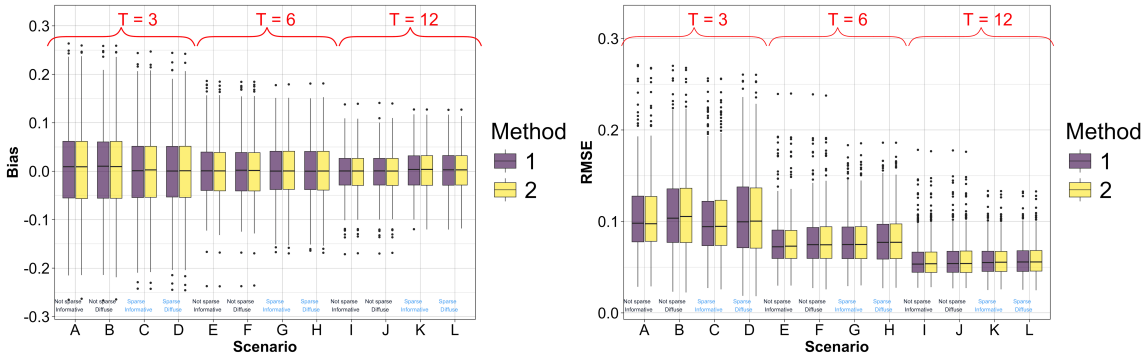
difference in the biases and RMSEs between the two methods for computing $\hat{x}(B_i, t)$. Moreover, the results show that the bias in γ_1 is close to zero for all scenarios. The RMSE tends to be high for fewer time points, but it becomes smaller for more time points. This is expected since with more time points, more information is available to estimate the parameter correctly. Moreover, the sparsity of the stations data does not affect the quality of estimates for γ_1. This is also expected since, as shown in Figures 3.12 and 3.13, the obtained block-level estimates $\hat{x}(B_i, t)$ are highly correlated and are close to the true values $x(B_i, t)$, whether the stations data is sparse or not. Hence, the obtained estimates for γ_1 will be similar for either case. Lastly, even with non-informative priors on γ_1, the bias and the RMSE are consistently small. Note that in the simulation study, the γ_1 parameter has a non-informative prior for all the scenarios, and so all the values shown in Figures 3.14a and 3.14b are computed using non-informative priors for γ_1. The insights for γ_1 also holds true for the intercept γ_0, as shown in Figures 3.15a and 3.15b.



(a) Bias **(b)** RMSE

Figure 3.14: Plot of biases and RMSEs for γ_1 for all scenarios

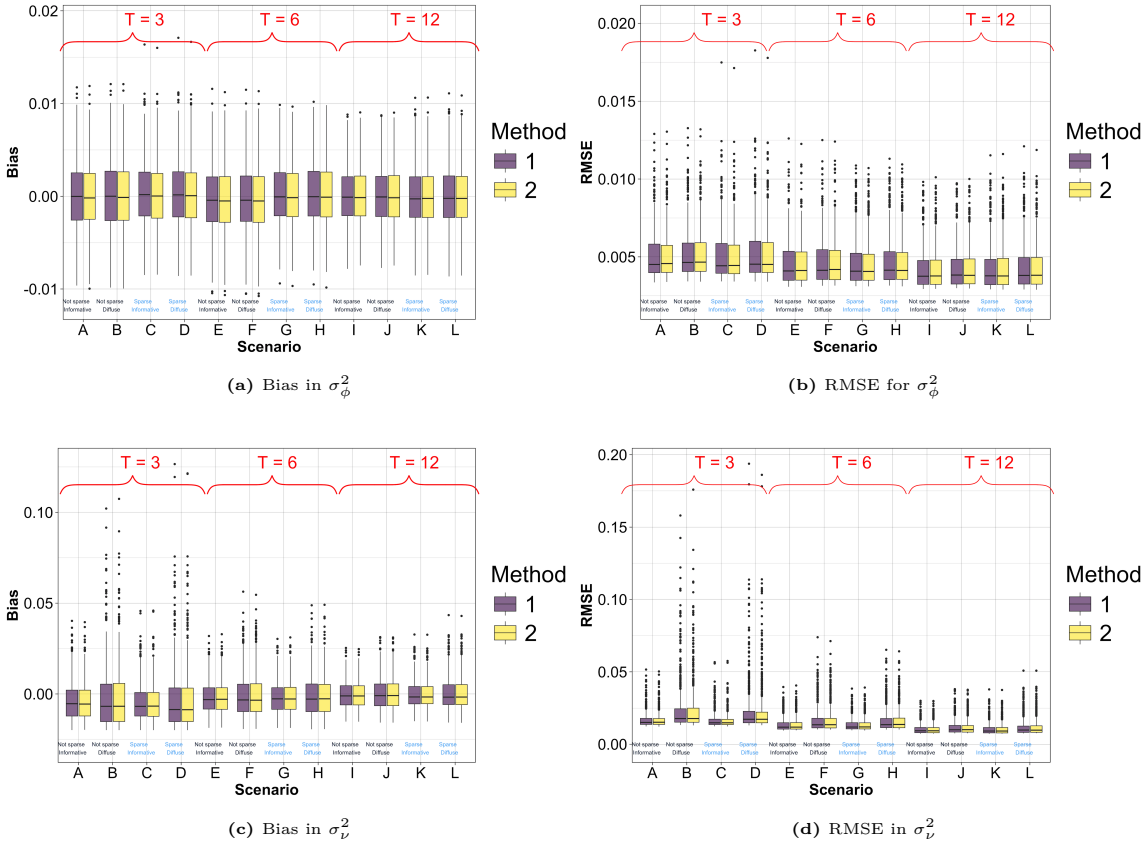
Figure 3.16 shows the biases and RMSEs for the variance parameters σ^2_ϕ and σ^2_ν. For the variance of the block-specific effect σ^2_ϕ, the biases and the RMSEs are generally close to zero for all the scenarios. The number of time points does not seem to unduly affect the biases and RMSEs. This is expected since the model in the simulation study assumes that the block-specific effect is independent with time, and so the number of time points in the data does not potentially affect the quality of the estimates for σ^2_ϕ. In addition, this parameter does not seem to be sensitive to the prior specification and the sparsity of the stations data. For σ^2_ν, the biases and RMSEs



(a) Bias **(b)** RMSE

Figure 3.15: Plot of biases and RMSEs for γ_0 for all scenarios

are generally smaller when there are more time points. This makes sense since σ_ν^2 is a parameter of the temporal random effect, and so it is more accurately estimated when the length of the time series is longer. In addition, the prior specification affects the precision of the estimates. If the prior distribution is informative, the RMSEs for σ_ν^2 are generally lower compared to the case when the prior is non-informative. The difference in the RMSEs for σ_ν^2 between the two priors diminishes with more time points.



(a) Bias in σ_ϕ^2 **(b)** RMSE for σ_ϕ^2

(c) Bias in σ_ν^2 **(d)** RMSE in σ_ν^2

Figure 3.16: Plot of biases and RMSEs for variance parameters of the second-stage model

I also compute the coverage probabilities for the second-stage model parameters, which are shown in Table 3.4. The coverage probabilities are computed as follows. Suppose that θ is a model parameter and that $\hat{\theta}_{ijk}$ is the k^{th} sampled value from the estimated posterior distribution of θ for the i^{th} data replicate and j^{th} sample from the first-stage model (using Algorithm 3.1). Suppose $\hat{\theta}_i^{(2.5)}$ and $\hat{\theta}_i^{(97.5)}$ are the 2.5^{th} and 97.5^{th} percentile, respectively, $\forall j, k$ and for a fixed i. The 95% coverage probability for θ is given by

$$\text{coverage probability} = \frac{1}{n_{\text{sim}}} \sum_{i=1}^{n_{\text{sim}}} \mathbb{I}_i(\theta), \qquad \mathbb{I}_i(\theta) = \begin{cases} 1 & \hat{\theta}_i^{(2.5)} < \theta < \hat{\theta}_i^{(97.5)} \\ 0 & \text{otherwise} \end{cases}.$$

The results show that for γ_1, all the coverage probabilities are very close to the nominal value of 95%. There is no difference between the two methods for computing $\hat{x}\big(B_i, t\big)$. For the intercept γ_0, the coverage probabilities are also reasonably close to the nominal value, except for the case when the number of time points is very small. Moreover, the coverage probabilities for the variance of the temporal effect σ_ν^2 is higher when (weakly) informative priors are used or when there are more time points. Finally, for the variance of the spatial effect σ_ϕ^2, the coverage probabilities are consistently high and close to the nominal value for all scenarios.

T	**Sparse**	**Priors**	γ_0		γ_1		σ_ϕ^2		σ_ν^2	
			M1	**M2**	**M1**	**M2**	**M1**	**M2**	**M1**	**M2**
3	No	informative	85.4	85.2	94.2	93	95.2	93.8	88.4	88.6
		non-informative	81.8	81.2	93.6	93.4	94.8	93.4	71.6	71.6
	Yes	informative	86.4	85.8	94	95	95	95.4	88	88
		non-informative	83	82.6	93.2	93.8	94.2	94.6	71.2	71
6	No	informative	91.8	91.6	93.6	94	94.2	93	89.8	89.4
		non-informative	91.4	91.2	94	94	94	92.6	84.2	83.4
	Yes	informative	91	90.6	94.4	94	94.2	94	89.6	90.6
		non-informative	90	90.4	94.2	94	94.2	93.6	86	85.4
12	No	informative	92.6	93.2	94	93.8	94.2	95.2	92.4	93
		non-informative	93.4	92.8	94	94.4	94.2	94.6	90	90
	Yes	informative	94.8	94.6	95.4	93.8	93.6	93.4	92.8	92
		non-informative	93.6	94.2	93.4	93.6	93.2	93.4	89.8	89.2

Table 3.4: Coverage probabilities (in %) of second-stage model parameters for all scenarios. M1 = Method 1, M2 = Method 2.

3.4 Discussion and conclusions

This chapter investigates a data fusion model to combine data from a network of monitoring stations, which is usually sparse but accurate, and from a proxy data, which has wide spatial coverage but biased. The data fusion model is the first-stage model in a two-stage modelling framework, which is primarily motivated by the epidemiological problem of linking health outcomes and exposure variables, such as the concentration of certain air pollutants, and where both data have different spatial supports. This chapter extends the two-stage model proposed by Cameletti et al. (2019) by incorporating a data fusion model in the first stage. The data fusion model is estimated using a data augmentation approach with the INLA-SPDE.

The proposed data fusion model is based on the Bayesian melding model, which assumes a common latent process for the different data sources. However, instead of treating the proxy data as areal, this work treats them as point-referenced at the centroids of the grid cells. This makes computation easier, since it avoids estimating stochastic integrals, especially in scenarios when the proxy data has high resolution. The data fusion model incorporates an additive and multiplicative bias for the proxy data, since it is noisier and biased compared to the measurements from the stations. The current model assumes a constant additive and multiplicative bias, which may not be flexible in some data applications, since the biases may be varying in space and time. This model specification, via the calibration biases, ensures that the proxy data does not dominate model estimation, which can be very likely when the biases are ignored, since there are more data from the proxy data than the stations data. The proposed model regards the measurements from the stations as more accurate, but at the same time allowing for measurement error.

The data fusion model is the first-stage model in the two-stage modelling framework presented in Figure 1.6. The second stage fits a health model where the spatial averages of the predicted first-stage latent field is an input variable, and the health outcome is the response variable. Both the first-stage and second-stage models, being latent Gaussian, are fitted using INLA. I explored two methods proposed in Cameletti et al. (2019) for computing the block-level estimates of the first-stage latent field. In

order to account for the uncertainty in the first-stage model, I used the posterior sampling method, which generates several samples from the estimated posterior distributions of the first-stage latent field on a fine prediction grid. For each set of sampled values, I fit the second-stage model using the corresponding block-level estimates of the field. All results are then combined using Bayesian model averaging, in order to obtain the posterior estimates of the second-stage model.

Summary of important results

A simulation study was carried out to assess the performance of the proposed method under different scenarios. The following settings are considered in the study: the sparsity of the stations data, number of time points, and the specification of the priors. It is common to work with sparse stations data, so it is interesting to look at the effect of the sparsity on the quality of the parameter estimates. Also, it is important in Bayesian analysis to assess the sensitivity of the results to the priors, more so in the context of a complex spatio-temporal process with several model parameters that need to be estimated.

All first-stage model parameters have generally small biases, but there is difficulty in estimating the Matèrn field parameters if non-informative priors are used. As long as informative priors are used, the bias and RSMEs are very small even if the data on the stations is sparse.

For the main parameter of interest, γ_1, the proposed framework provides very good estimates across all scenarios considered in the simulation study. There is no difference between the two methods of computing the spatial averages in terms of the bias, RMSEs, and coverage probabilities. Even with non-informative prior on γ_1 and sparse stations data, the estimates for γ_1 are close to the true value. Finally, with more time points, the RMSEs tend to decrease.

The simulation study showed that the sparsity of the stations data can potentially affect the quality of the parameter estimates, especially for the first-stage model. When the stations data is sparse, the RMSE of the covariate (fixed) effect is large. This is also true for the measurement error variance in the stations; but the use of informative priors can be helpful to accurately estimate the parameter. It makes sense for these two parameters to be seriously affected under scenarions of sparse

stations data, since with less stations, there is also less information at our disposal to accurately estimate them. The RMSEs of the Matèrn parameters and the bias parameters of the proxy data are also generally higher when the stations data is sparse. The parameters in the second-stage model seem to be not affected by the sparsity of the stations data since the proposed method, even with sparse data, was able to estimate well the latent field, which also means that the corresponding block-level estimates are close to the true values, at least for most of the areas, as shown in the correlations, biases, and RMSEs from the simulation results.

The use of (weakly) informative priors gave better parameter estimates, especially for the Matérn parameters, which are typically difficult to estimate. The autoregressive parameter of the latent field also benefits with the use of informative priors, giving smaller RMSEs, although there is no substantial difference in the bias between the use of non-informative and informative prior. In addition, the measurement error variance in the stations also have lower RMSEs when informative priors are used, especially for the case when the stations data is sparse. The rest of the parameters are not too sensitive to the priors.

The number of time points can also potentially affect the quality of the estimates. As already mentioned, if there are more time points, the RMSEs of the fixed effects in the second-stage model are smaller. This is also true for the variance of the time effect in the second-stage model. The bias parameters of the proxy data in the first-stage model also have better estimates with more time points.

The method for computing block-level estimates of the first-stage latent field does not show to have an impact on the parameter estimates. The biases in the block-level estimates for both methods are close to zero, although the first method seems to give relatively higher biases for certain blocks and also higher RMSEs overall. Moreover, the second method gave slightly higher correlations between the true block-level values and the estimated values; although, both methods gave fairly high correlations. In terms of the quality of the parameter estimates of the second-stage model, both methods worked equally well.

Limitations of the model

A major limitation of the proposed data fusion model is that both the additive

bias and multiplicative bias are assumed constant. This may not be sufficient or flexible enough to account for the biases in the proxy data. This is immediately addressed in Chapter 4.

The use of the full Bayesian melding model, which treats the proxy data as areal, would also potentially increase the computational costs. Under the full Bayesian melding model, it is required to evaluate stochastic integrals and to have covariate information on $\mathrm{z}(\mathbf{s}, t)$ at a resolution that is finer than the resolution of the proxy data. The use of a data augmentation approach needs an enlarged extended field, say $\boldsymbol{\chi}_t = \begin{pmatrix} \boldsymbol{x}_{t,S}^\intercal & \boldsymbol{x}_{t,P^*}^\intercal & \boldsymbol{x}_{t,S}^{*\intercal} & \boldsymbol{x}_{t,P^*}^{*\intercal} \end{pmatrix}^\intercal, t = 1, \ldots, T$; $\boldsymbol{x}_{t,P^*}^{*\intercal}$ is of larger dimension than $\boldsymbol{x}_{t,P}^{*\intercal}$ in the original model, where P^* is the resolution of the covariate data $\mathrm{z}(\mathbf{s}, t)$, which should be greater than the number of grid cells for the proxy data. Hence, when the proxy data has a very high resolution, applying the full Bayesian melding model might amplify the computation effort required. The use of a full Bayesian melding model is a problem that will be investigated in the next chapter.

Other extensions

The proposed data fusion model is not compared to benchmark models, such as the use of data solely from stations, since the goal of the current chapter is to provide an initial investigation of the capabilities of doing data fusion using the INLA-SPDE methodology, in a two-stage modelling framework. However, it is important to compare the performance of the proposed model with other existing approaches. Chapter 4 proposes a flexible data fusion model, whose performance is compared with benchmark approaches. The data application is on meteorological data from the Philippines, which is an ideal case study since the weather stations data is very sparse, while the proxy data is evidently more biased.

Although this chapter considered only two data sources for data fusion, the proposed model and the estimation approach can be easily extended to more than two data sources. In this context, we only need to introduce a new likelihood model for each additional data source, and assume that each one is a function of the same latent process, but with a different set of bias parameters. Even if the new data sources have different spatial supports, the proposed model and framework are feasible.

Another extension is to look into the problem of uncertainty propagation in two-

stage Bayesian models. This chapter used a resampling approach to account for the uncertainty in the first-stage model. Although the approach is intuitive, it needs to be formally validated. Another method for uncertainty propagation which does not require resampling is also another methodological innovation. These are explored in this thesis; the results of which are presented in Chapter 6.

Another extension of the current model is to consider other measurement error processes. For as long as the latent Gaussian assumption is satisfied, the current framework of using the INLA-SPDE approach can still be used in such extensions.

The proposed model allows one to obtain estimates of the risks at the level of the administrative units where the health outcomes are observed. However, one might be interested to obtain estimates of risks at a fine scale, since this allows one to look at fine-scale heterogeneity of the risks in space. A recent work has been done on this in a so-called fusion area-cell spatio-temporal generalized geoadditive-Gaussian Markov random feld (FGG-GMRF) (Jaya and Folmer, 2022). A new model specification which incorporates a latent risk surface or intensity field is explored in Chapter 6.

The proposed model assumes a separable covariance function, i.e., the covariance matrix of the SPDE representation of the first-stage model is a Kronecker product of the covariance matrix in space, whose elements are computed from the SPDE model, and the covariance matrix in time, whose elements are computed from the assumed temporal model. The separability assumption is convenient because it simplifies computation; but this assumption could be inadequate. An extension of the model is to assume non-separability, which is currently an active area of research. One approach of doing this is to start from an SPDE which yields a Gaussian field with a separability parameter (Lindgren et al., 2020). This parameter determines the type of non-separability of the spatio-temporal covariance function.

Finally, the current model assumes that the hyperparameters of the latent spatio-temporal process is constant through time. But these hyperparameters could evolve at some point, i.e., the Matèrn field parameters could change, or the mean structure of the model could evolve as well. This is a change-point detection problem, which is also a promising future work since relatively only few work has been done for change-point detection in spatio-temporal processes.

Chapter 4

A flexible data fusion model: application on meteorological data in the Philippines

This chapter extends the proposed data fusion model in Chapter 3. In particular, the model proposed in this chapter flexibly accounts for calibration biases in the data fusion model, and also defines a single and interpretable latent process for the different data sources. The novelty is in the specification of a random field for the additive bias $\alpha_0(\mathbf{s})$, which I call an *error field*. In a spatio-temporal context, I assume that the error field is a time-varying random field. In addition to a flexible model specification, the proposed method also allows us to gauge the quality of the different data sources. The proposed model in this chapter extends the existing data fusion models presented in Section 2.7.3 in Chapter 2.

The proposed data fusion model is motivated by a data application for meteorological data in the Philippines. As discussed in Section 1.2.1, there are two primary meteorological data sources in the country: observational data from a sparse network of weather synoptic stations and simulated outputs from a numerical weather forecast model called the *Global Spectral Model* (GSM) (PAGASA, 2023). While the latter provides broad spatial coverage, it is typically biased due to sensitivity to model initialization and parameterization. In contrast, the weather stations data,

which are likely to be less biased, provide limited spatial coverage, leaving key areas under-sampled, as shown in Figure 1.3 and Figure 4.1. To address these limitations, I propose using both data sources together through data fusion.

In this chapter, a simulation study compares the performance of the proposed data fusion model with two benchmark methods: a stations-only model and a statistical calibration model. In the data application, the predictions from the data fusion approach and the two benchmark approaches are compared using leave-group-out cross-validation (Adin et al., 2023; Liu and Rue, 2022). Although the data application only considers two data sources, the model and framework can easily extend to more than two data sources and are relevant beyond this specific context.

This chapter is structured as follows. Section 4.1 discusses the meteorological data problem, which is the motivation for this chapter. A preliminary simulation study is performed presented in Section 4.2, which informs the proposed data fusion framework and model in Section 4.3. Section 4.4 discusses the estimation approach, which is Bayesian model averaging with INLA. Section 4.5 discusses the simulation study. The results for the real-life data application are in Section 4.6. Finally, I end this chapter with some discussion and conclusion in Section 4.7.

4.1 Meteorological data from the Philippines

The Philippines is an archipelagic country, covering an area of ca. 300 thousand km^2 (see Figure 4.1), situated in tropical Southeast Asia. The eastern part and some southern parts of the country are mostly classified as tropical rainforest, and are characterized by the lack of a distinct wet or dry season and with relatively high rainfall all year round. On the other hand, most of the country's western section is classified as tropical monsoon or tropical Savannah, characterized by pronounced dry and wet seasons (Coronas, 1920; Kintanar, 1984a; PAGASA, 2023). The rainy season of the country, which also coincides with the hot episode of a year, lasts from June to November, while the rest of the year is generally considered dry. The dry season is further categorized into either a cool dry or a hot dry period, where the former lasts from December to February and the latter from March to May (PAGASA, 2023).

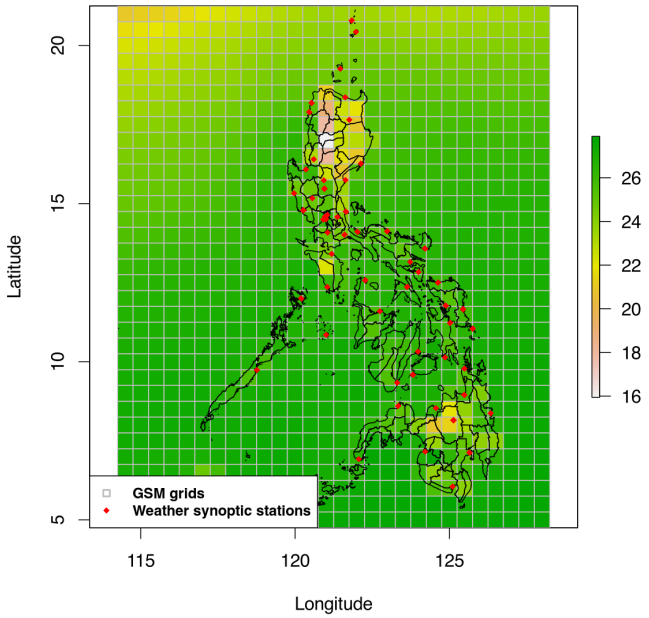


Figure 4.1: Meteorological data sources for the Philippines: a sparse network of weather synoptic stations and an outcome of a numerical weather forecast model called *Global Spectral Model*. The measurements are monthly aggregated values of temperature for August 2019.

The Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) maintains a network of 57 weather synoptic stations that regularly record several meteorological variables including temperature, relative humidity, and rainfall. The spatial distribution of the weather stations, shown in Figure 4.1, is a sparse network relative to the country's total surface area, with some regions, especially in the north, being heavily undersampled. Consequently, reconstructing meteorological variable surfaces based only on the weather stations would therefore result in high uncertainty in many parts of the country. To remedy this problem of data sparsity, PAGASA utilizes outcomes from the Global Spectral Model, a numerical weather forecast model maintained by the Japanese Meteorological Agency. The GSM provides forecast outputs of up to 132 hours four times a day (with initial times 0000, 0600, 1200, and 1800 UTC) within 4 hours of the initial time, and up to 264 hours twice a day (with initial times 0000 and 1200 UTC) within 7 hours of the initial time. As an illustration, Figure 4.1 shows a map of mean temperature from the GSM outcomes for August 2019 at a spatial resolution of 0.5 degrees (approximately 55km $\times$ 55 km) corresponding to 924 grid cells. Although the GSM outcomes are gridded, PAGASA interprets them as point-referenced at the centroids (PAGASA, 2023).

PAGASA provided the aggregated monthly data from both the weather stations and GSM for 2019 and 2020, and for the following meteorological variables: mean temperature (in °C), mean relative humidity (in %), and total rainfall (in mm). The GSM outcomes were first simulated daily, using the 0000 UTC initial time, to produce forecasts at 3-hour intervals and up to eight forecast horizons. The simulated outcomes were then aggregated at the monthly level, yielding averages for temperature and relative humidity and totals for rainfall. The use of a monthly temporal scale is motivated by its relevance to future work (Chapter 5), where the model predictions will serve as inputs to an epidemiological model for dengue, as the case counts are typically available at the monthly level (Abdullah et al., 2022; Naish et al., 2014). The goal of this chapter is, therefore, to reconstruct monthly surfaces for meteorological variables, specifically temperature, relative humidity, and rainfall. Here, the focus is on improving the accuracy of spatial predictions and mapping of these variables in space, rather than on forecasting future outcomes.

To assess the bias in the GSM outcomes, I interpolate the GSM values on the three meteorological variables of interest, and predict the values at the weather stations' locations. I do this by fitting the following geostatistical model:

$$\mathrm{w}(\mathbf{s},t) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s},t) + \xi(\mathbf{s},t) + e(\mathbf{s},t).$$

Here, $\mathrm{w}(\mathbf{s},t)$ is the simulated value of the meteorological variable from the GSM at the grid cell with centroid $\mathbf{s}$, $\mathrm{z}(\mathbf{s},t)$ is a known covariate, $\xi(\mathbf{s},t)$ is a spatio-temporal random effect, and $e(\mathbf{s},t)$ is a random noise, i.e., $e(\mathbf{s},t) \overset{\text{iid}}{\sim} \mathcal{N}\left(0,\sigma_e^2\right)$. For the single covariate $\mathrm{z}(\mathbf{s},t)$, I used elevation for temperature and relative humidity, and relative humidity for the log-transformed rainfall. I assume that $\xi(\mathbf{s},t)$ evolves in time as an autoregressive process of order 1, i.e.,

$$\xi(\mathbf{s},t) = \phi\xi(\mathbf{s},t-1) + \omega(\mathbf{s},t), \quad t = 2,\ldots,T,$$

where $|\phi| < 1$, and $\omega(\mathbf{s},t)$ is a time-independent Gaussian process with a Matérn covariance function and which follows the stationary distribution of the process at time $t = 1$. I fit this model using INLA and the SPDE method (Lindgren et al.,

2011; Rue et al., 2009) and then predict the values of $\mathbb{E}\big[\mathrm{w}(\mathbf{s},t)\big]$ at the weather stations' locations using the posterior predictive mean. The predicted values are then compared with the observed values, which are shown in Figure 4.2. The results show a general agreement between the two sets of values, but a clear bias is visible. Figure 4.2 highlights specific weather stations, where it becomes clear, especially for temperature, that there is a spatially-varying additive bias but no multiplicative bias, since the GSM outcomes at a specific weather station seem parallel to the identity line. For the other two meteorological variables, it is also clear that there is a spatially-varying additive bias, but accounting for a multiplicative bias parameter with a complex structure might be necessary.

Figure 4.2 also shows that the quality of the GSM outcomes varies among the three variables. The discrepancy in the outcomes between the GSM and weather stations for the rainfall data is bigger than for the other two meteorological variables. The proposed data fusion model, which is discussed in Section 4.3, is able to gauge the relative quality of the GSM outcomes for the three meteorological variables.

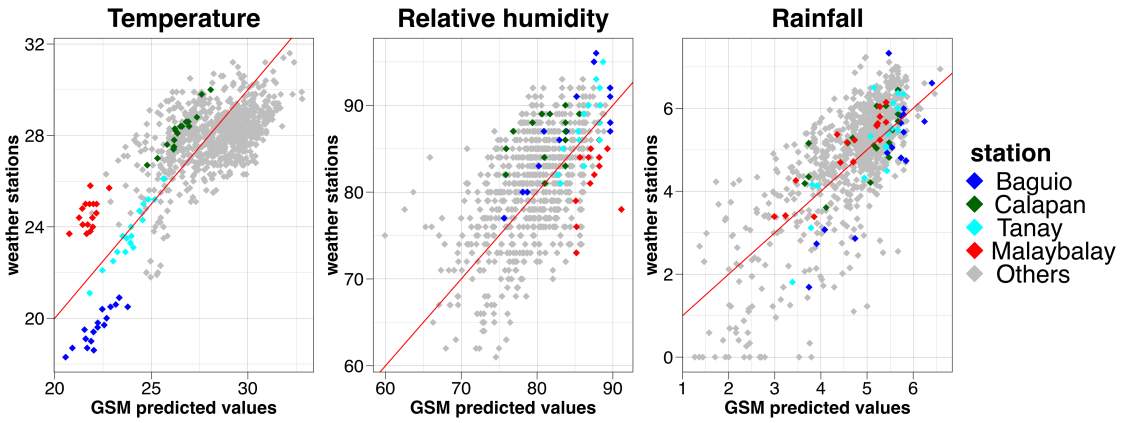


Figure 4.2: Scatterplot of the observed values at the weather stations versus interpolated outcomes of the GSM for three meteorological variables: temperature, relative humidity, and log-transformed rainfall. The plot shows the discrepancies in the values between the two data sources.

4.2 Preliminary results

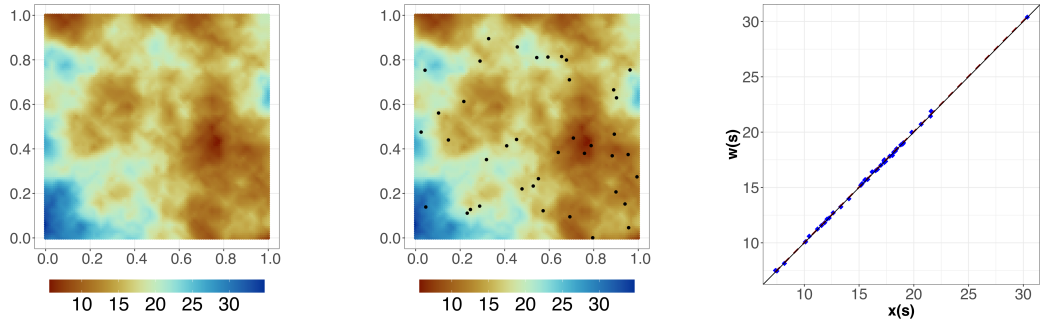
In Chapter 3, I treated the outcomes of the proxy data as point-referenced at the centroids of the grid cells. This preliminary section explores a comparison of the full Bayesian melding model (Section 2.7.1), which treats the outcomes of numerical models as gridded, and the one which treats the outcomes as point-referenced.

Suppose that the latent field follows the following model: $x(\mathbf{s}) = 10 + 3\mathrm{z}_1(\mathbf{s}) + 2\mathrm{z}_2(\mathbf{s}) + \xi(\mathbf{s})$, where $\mathrm{z}_1(\mathbf{s})$ and $\mathrm{z}_2(\mathbf{s})$ are predictor variables, and $\xi(\mathbf{s})$ is a random field which follows the Matérn process with effective range of 2, marginal variance of 4, and mean-squared differentiability parameter equal to 1. A simulated field is shown in Figure 4.3a.

Suppose that there are three data sources for the unknown field in Figure 4.3a, with the following model structures:

$$\begin{aligned}
&\text{(Data source A)} \quad \mathrm{w}_1(\mathbf{s}) = x(\mathbf{s}) + \epsilon_1(\mathbf{s}) \\
&\text{(Data source B)} \quad \mathrm{w}_2(\mathbf{s}) = \alpha_{02}(\mathbf{s}) + \alpha_{12}x(\mathbf{s}) + \epsilon_2(\mathbf{s}) \\
&\qquad\qquad\qquad\quad \mathrm{w}_2(B) = \frac{1}{|B|}\int_B \mathrm{w}_2(\mathbf{s})d\mathbf{s} \\
&\text{(Data source C)} \quad \mathrm{w}_3(\mathbf{s}) = \alpha_{03}(\mathbf{s}) + \alpha_{13}x(\mathbf{s}) + \epsilon_3(\mathbf{s}) \\
&\qquad\qquad\qquad\quad \mathrm{w}_3(B) = \frac{1}{|B|}\int_B \mathrm{w}_3(\mathbf{s})d\mathbf{s}
\end{aligned} \tag{4.1}$$

Data source A, which are point observations, follows the classical error model, where $\epsilon_1(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}(0, 0.01)$. The observed locations for $\mathrm{w}_1(\mathbf{s})$ are shown in Figure 4.3b, while a comparison of the observed values and true values are shown in Figure 4.3c.

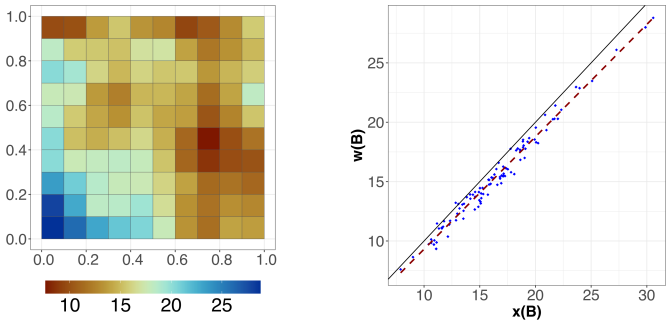


(a) Simulated true field **(b)** Point locations of $\mathrm{w}_1(\mathbf{s})$ **(c)** $\mathrm{w}_1(\mathbf{s})$ vs true values

Figure 4.3: (a) simulated true field $x(\mathbf{s})$ (b) simulated observed data at finite point locations (c) comparison of observed values $\mathrm{w}_1(\mathbf{s})$ and true values of $x(\mathbf{s})$

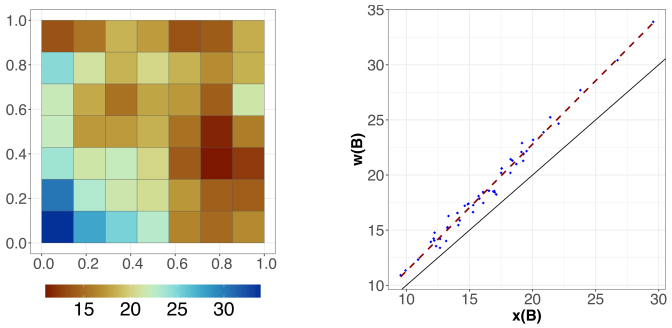
Moreover, Data source B is denoted by $\mathrm{w}_2(B)$ and has a conceptual point-referenced model denoted by $\mathrm{w}_2(\mathbf{s})$. The additive bias, $\alpha_{02}(\mathbf{s})$, is simulated from a Matérn field with effective range of 3 and marginal variance of 1, while the multiplicative bias is $\alpha_{12} = 0.9$. Figure 4.4a shows a simulated $\mathrm{w}_2(B)$, while Figure 4.4b shows a comparison of $\mathrm{w}_2(B)$ and $x(B) = \frac{1}{|B|}\int_B x(\mathbf{s})d\mathbf{s}$. It shows that the observed values are

biased and tend to be smaller than the true values due to the multiplicative bias $\alpha_{12} < 1$. Lastly, data source C, denoted by $\mathrm{w}_3(B)$, is also areal and has a conceptual point-referenced model given by $\mathrm{w}_3(\mathbf{s})$. The additive bias, $\alpha_{03}(\mathbf{s})$, is simulated from the same process as $\alpha_{02}(\mathbf{s})$, while the multiplicative bias is given by $\alpha_{12} = 1.1$. In addition, I assume that $\mathrm{w}_3(B)$ is observed at a coarser resolution than $\mathrm{w}_2(B)$. Figure 4.5a shows a simulated $\mathrm{w}_3(B)$, while Figure 4.5b shows a comparison of $\mathrm{w}_3(B)$ and $x(B) = \frac{1}{|B|}\int_B x(\mathbf{s})d\mathbf{s}$. Unlike $\mathrm{w}_3(B)$, Figure 4.5b shows that $\mathrm{w}_3(B)$ tend to be larger than the true values due to the multiplicative bias $\alpha_{12} > 1$.



(a) Observed $\mathrm{w}_2(B)$ **(b)** Observed vs true values

Figure 4.4: (a) simulated $\mathrm{w}_2(B)$ (b) comparison of observed values $\mathrm{w}_2(B)$ and $x(B) = \frac{1}{|B|}\int_B x(\mathbf{s})d\mathbf{s}$



(a) Observed $\mathrm{w}_3(B)$ **(b)** Observed vs true values

Figure 4.5: (a) simulated $\mathrm{w}_3(B)$ (b) comparison of observed values $\mathrm{w}_3(B)$ and $x(B) = \frac{1}{|B|}\int_B x(\mathbf{s})d\mathbf{s}$

4.2.1 Modelling approaches

I compare four modelling approaches: (1) a model that uses only data from $\mathrm{w}_1(\mathbf{s})$, (2) a model that uses all three data sources but assuming a classical error model for $\mathrm{w}_2(B)$ and $\mathrm{w}_3(B)$, (3) a full Bayesian melding model, (4) and a simplification of the

Bayesian melding model by treating the proxy data as point-referenced. For model estimation, I assume that only one of the two covariates, specifically $z_1(\mathbf{s})$, is known and available.

The first approach is a standard model and can be easily fitted using INLA-SPDE (Cameletti et al., 2013). The second approach is similar to the models in Moraga et al. (2017) and Zhong and Moraga (2023) (see Section 2.7.3), and which can be straightforwardly fitted using INLA-SPDE as well. On the other hand, the third approach has more involved predictor expressions in its three likelihood components. The joint model is given by:

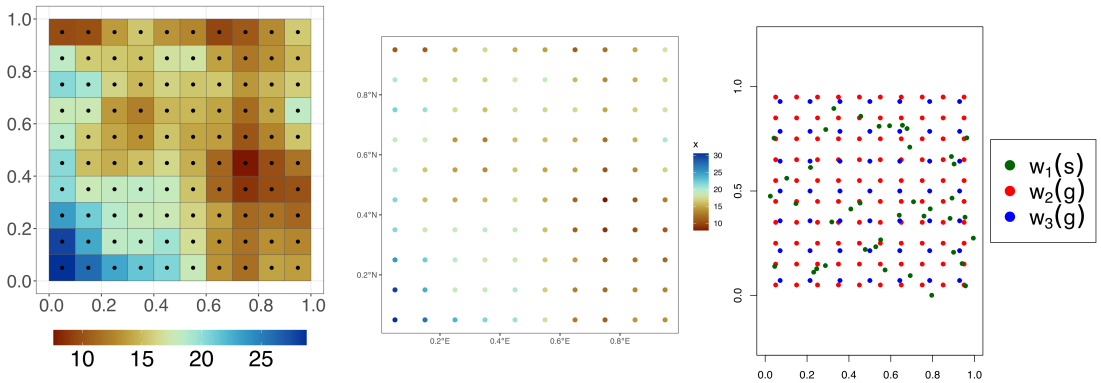
$$\begin{aligned} w_1(\mathbf{s}) &= \beta_0 + \beta_1 z_1(\mathbf{s}) + \xi(\mathbf{s}) + \epsilon_1(\mathbf{s}), \quad \epsilon_1(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{\epsilon_1}\big) \\ w_2(B) &= \frac{1}{|B|}\int_B \Big[\alpha_{02}(\mathbf{s}) + \alpha_{12} \times \Big(\beta_0 + \beta_1 z_1(\mathbf{s}) + \xi(\mathbf{s})\Big) + \epsilon_2(\mathbf{s})\Big] d\mathbf{s} \\ w_3(B) &= \frac{1}{|B|}\int_B \Big[\alpha_{03}(\mathbf{s}) + \alpha_{13} \times \Big(\beta_0 + \beta_1 z_1(\mathbf{s}) + \xi(\mathbf{s})\Big) + \epsilon_3(\mathbf{s})\Big] d\mathbf{s} \\ \epsilon_2(\mathbf{s}) &\overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{\epsilon_2}\big), \quad \epsilon_3(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{\epsilon_3}\big) \end{aligned} \tag{4.2}$$

The complication here is due to the fact that the predictor expressions for $w_2(B)$ and $w_3(B)$ in Equation (4.2) involve a product of two unknowns: a scaling parameter and a Gaussian field. Furthermore, this also involves an aggregation of the values over the blocks B. To fit this model, I used the Bayesian model averaging with INLA (Gómez-Rubio et al., 2020), which is discussed in more detail in Section 4.4. Essentially, a grid of values is defined for α_{12} and α_{13}, and then Equation (4.2) is estimated conditional on α_{12} and α_{13}. The final posterior estimates are then computed using model averaging. For both parameters, I defined a grid of values from 0.7 to 1.2, with a length step of 0.1.

The fourth approach simplifies Equation (4.2) by treating $w_2(B)$ and $w_3(B)$ as point-referenced, which yields the following predictor expressions for the joint model:

$$\begin{aligned} w_1(\mathbf{s}) &= \beta_0 + \beta_1 z_1(\mathbf{s}) + \xi(\mathbf{s}) + \epsilon_1(\mathbf{s}), \quad \epsilon_1(\mathbf{s}) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{\epsilon_1}\big) \\ w_2(\mathbf{g}) &= \alpha_{02}(\mathbf{g}) + \alpha_{12} \times \Big(\beta_0 + \beta_1 z_1(\mathbf{g}) + \xi(\mathbf{g})\Big) + \epsilon_2(\mathbf{g}) \\ w_3(\mathbf{g}) &= \alpha_{03}(\mathbf{g}) + \alpha_{13} \times \Big(\beta_0 + \beta_1 z_1(\mathbf{g}) + \xi(\mathbf{g})\Big) + \epsilon_3(\mathbf{g}) \\ \epsilon_2(\mathbf{g}) &\overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{\epsilon_2}\big), \quad \epsilon_3(\mathbf{g}) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{\epsilon_3}\big) \end{aligned} \tag{4.3}$$

Figure 4.6a shows a simulated $w_2(B)$ with the grid centroids in black points, while Figure 4.6b shows the point-referenced version of $w_2(B)$, here denoted by $w_2(\mathbf{g})$. Fitting the joint model in Equations (4.3) is simpler since it does not perform aggregation of the latent fields over the block B. Figure 4.6c shows the data points from the three data sources when using estimation approach (4). However, even with this simplification, the prediction expressions in Equation (4.3) still involve a product of a scaling parameter and an unknown Gaussian field. A Bayesian model averaging approach with INLA is also used to fit the joint model.



(a) $w_2(B)$ and grid centroids **(b)** $w_2(\mathbf{g})$ **(c)** Point-referenced data

Figure 4.6: (a) simulated $w_2(B)$ with the grid centroids, (b) point-referenced values $w_2(\mathbf{g})$, (c) data for model fitting using approach (4)

4.2.1.1 Prior specification

I assign vague priors for all fixed effects, i.e., $\beta_0 \sim \mathcal{N}(0, \infty)$ and $\beta_1 \sim \mathcal{N}(0, 1000)$. I assign PC priors for all the variance parameters and the Matérn field parameters (Fuglstad et al., 2019; Simpson et al., 2017). In defining the probability statement for the PC priors in the simulation exercise, I used the true values of the parameters and a probability value of 0.50. For the multiplicative bias parameters, I assign a uniform prior.

4.2.2 Results

4.2.2.1 (Simulated) data illustration

Figure 4.7 shows the predicted fields $\hat{x}(\mathbf{s})$, based on the mean of the posterior predictive distribution on a fine grid, for the different modelling approaches, from the

simulated data in Figures 4.3 – 4.5. The results show that the predicted fields look very similar to the truth. Figures 4.8a and 4.8b show a comparison of the bias and the posterior uncertainties in the predicted field $\hat{x}(\mathbf{s})$, respectively. The bias is defined as $x(\mathbf{s}) - \hat{x}(\mathbf{s})$, while the posterior uncertainty is the posterior standard deviation (SD) of $x(\mathbf{s})$. The results show that approaches (3) and (4) have generally smaller bias and smaller posterior uncertainty than the other two approaches. Moreover, the data fusion approach which does not account for the biases in the proxy data gave very large biases in some sections of the study region.

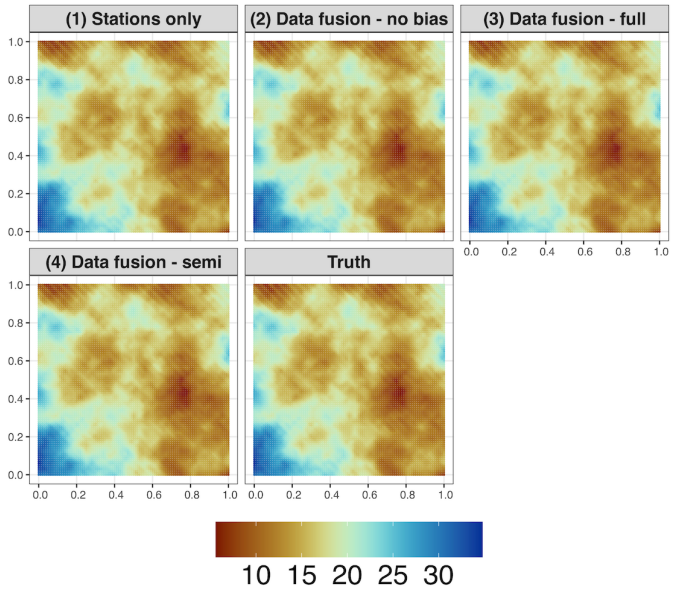
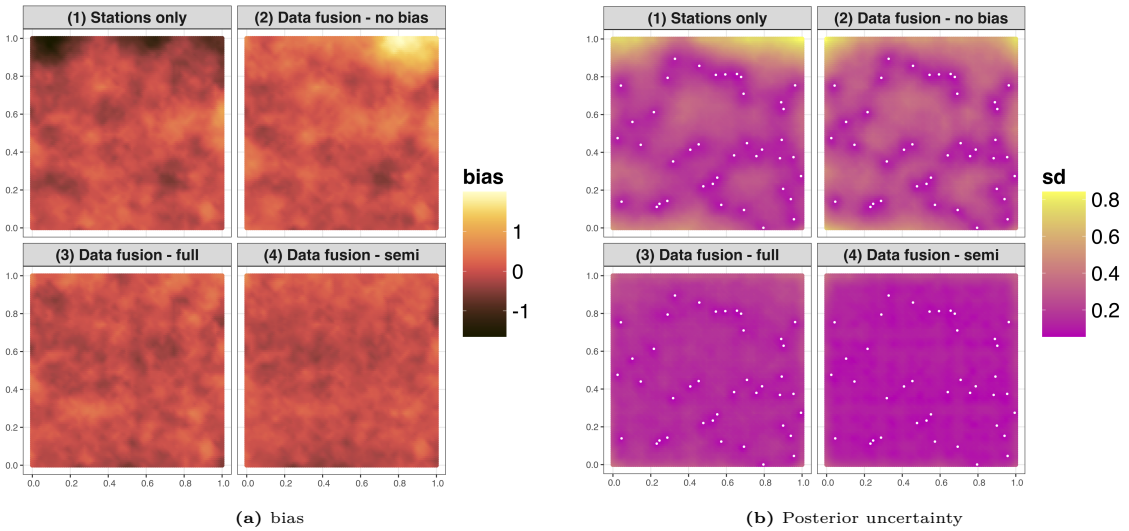


Figure 4.7: Comparison of predicted fields $\hat{x}(\mathbf{s})$ using the four different modelling approaches



(a) bias

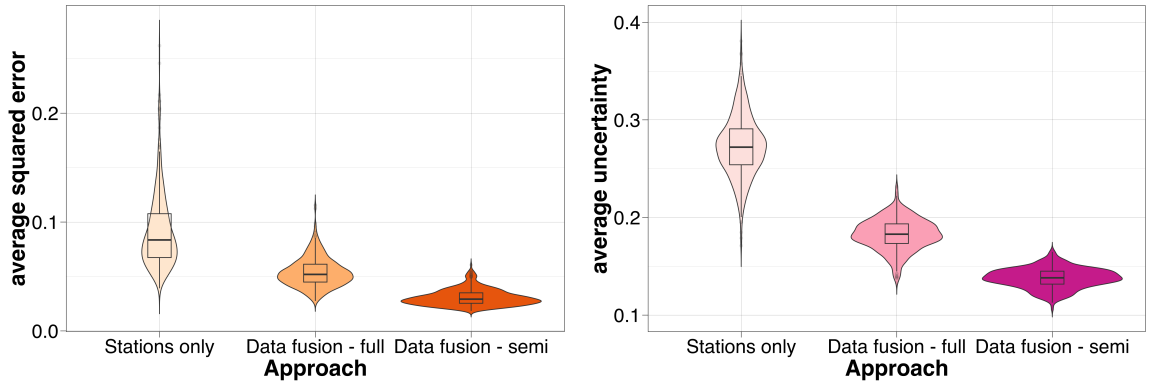
(b) Posterior uncertainty

Figure 4.8: (a) comparison of the bias in the predicted fields $\hat{x}(\mathbf{s})$ (b) comparison of the posterior uncertainties in the predicted field $\hat{x}(\mathbf{s})$

4.2.2.2 Average performance (500 replicates)

This section presents results based on 500 data replicates. Here, I only used Data source B as the proxy data for model estimation. This simplifies computation, since I only need to define a grid of values for α_{12} in fitting the conditional INLA models and in doing the model averaging.

For model comparison, I mainly looked at the average squared errors and average posterior uncertainty in the predicted fields $\hat{x}(\mathbf{s})$. The two metrics for model comparison are formalized in Section 4.5.2.



(a) Average squared error

(b) Average posterior standard deviation

Figure 4.9: (a) comparison of the bias in the predicted fields $\hat{x}(\mathbf{s})$ (b) comparison of the posterior uncertainties in the predicted field $\hat{x}(\mathbf{s})$

Figures 4.9a and 4.9b show a comparison of the average squared errors and average posterior uncertainty in the predicted fields, out of 500 data replicates, among the different modelling approaches. In the figure, I label as *Data fusion -full* the approach which performs the full Bayesian melding model, and label as *Data fusion - semi* the simplified approach which treats the proxy data as point-referenced at the centroids. Here, I exclude the second approach (data fusion but not accounting for the biases in the proxy data), since it gives very high bias, which was illustrated previously in Figure 4.8a. The results show that the stations-only approach generally has the highest average squared error and highest average posterior uncertainty; while the data fusion approach which treats the proxy data as point-referenced gave the smallest values for the two metrics. Figure 4.10 shows a comparison of the average BMA weights for the conditional INLA models, conditional on α_{12}, both for the full Bayesian melding approach and the simplified approach. The results show that

the model with $\alpha_{12} = 0.9$ has the highest average BMA weight, i.e., both estimation approaches are able to correctly estimate the true value of α_{12}.

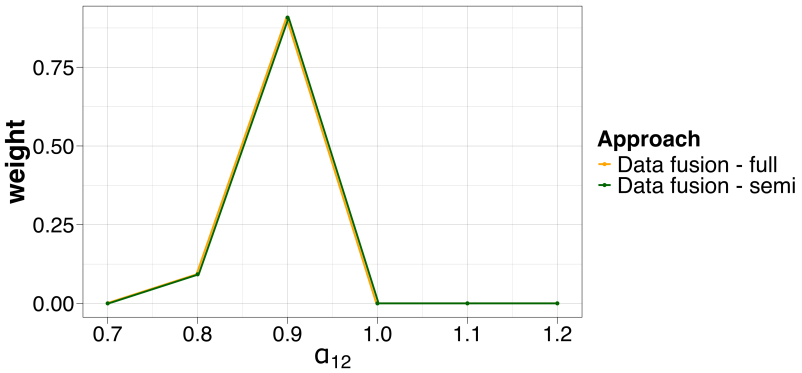


Figure 4.10: Average BMA weights for the conditional INLA models, conditional on α_{12}

Next, I looked at the potential impact of the resolution of the proxy data on the accuracy of the predicted fields. I considered three cases: 9×9, 12×12, 14×14. Figures 4.11 and 4.12 show the average squared errors and average posterior uncertainty in the predicted field with respect to the resolution of the proxy data and modelling approach. The results indicate no substantial differences in the average squared errors; but notable differences in the posterior uncertainties. In particular, the posterior uncertainty is generally smaller for the data fusion approach which treats the proxy data as point-referenced, and is true for the three levels in the resolution of the proxy data considered.

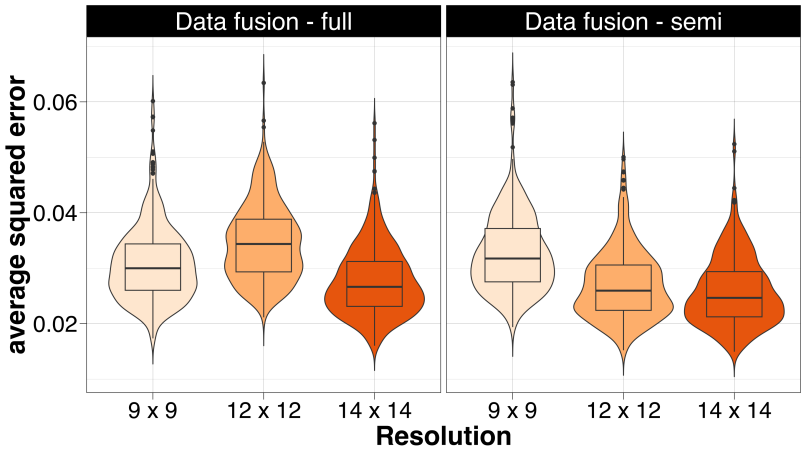


Figure 4.11: Comparison of the average squared errors with respect to the resolution of the proxy data and the data fusion approach

The main result in this preliminary investigation is that it is reasonable to treat the proxy data as point-referenced at the centroids. It has the same performance

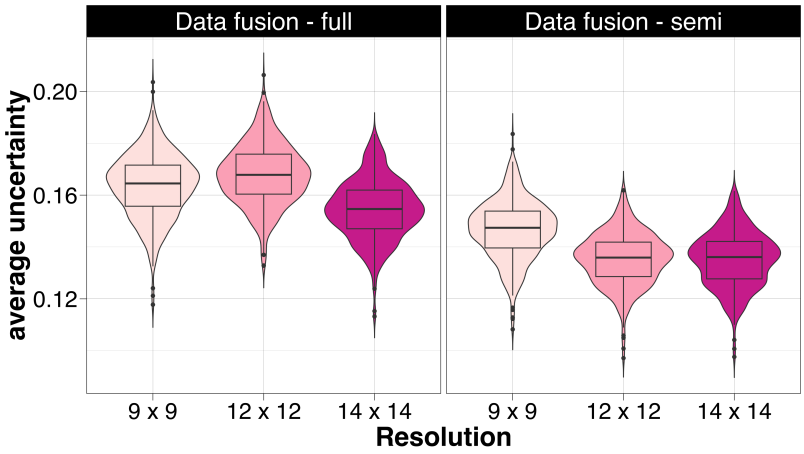


Figure 4.12: Comparison of the average posterior uncertainty respect to the resolution of the proxy data and the data fusion approach

as the full Bayesian melding model with respect to the average squared error of the predicted field, and has generally smaller posterior uncertainty. Moreover, the results also show that even if the proxy data is utilized for model estimation, but if the additive and multiplicative biases are not accounted for, it will yield model results similar to (or worse than) the stations-only model. These preliminary results serve as the basis for the data fusion framework and model proposed in this chapter, which are discussed in the next section (Section 4.3).

4.3 Data fusion framework and model

4.3.1 Framework

I assume that the latent process, at a spatial location $\mathbf{s}$ and time t, $\mathbf{s} \in \mathcal{S}$, $t = 1, \ldots, T$, is denoted by $x(\mathbf{s}, t)$. This is observed via two different data sets. The first one is

$$\mathbf{w}_{1t}^{\intercal} = \begin{pmatrix} \mathrm{w}_1(\mathbf{s}_1, t) & \mathrm{w}_1(\mathbf{s}_2, t) & \ldots & \mathrm{w}_1(\mathbf{s}_{n_M}, t) \end{pmatrix}, \quad \mathbf{s}_i \in \mathcal{S},$$

which are observations from a set of n_M stations in locations $\mathbf{s}_i$, $i = 1, \ldots, n_M$, at times $t = 1, \ldots, T$. The second one is

$$\mathbf{w}_{2t}^{\intercal} = \begin{pmatrix} \mathrm{w}_2(\mathbf{g}_1, t) & \mathrm{w}_2(\mathbf{g}_2, t) & \ldots & \mathrm{w}_2(\mathbf{g}_{n_G}, t) \end{pmatrix}, \quad \mathbf{g}_j \in \mathcal{S}$$

which denotes the proxy data, such as gridded outcomes from a numerical model – for instance the GSM (see Section 4.1). Here, $\mathrm{w}_2(\mathbf{g}_j, t)$ is the value at the grid cell with centroid $\mathbf{g}_j$, $j = 1, \ldots, n_G$, and time t. I assume that the two data sources are aligned in time. If this is not the case, it is always possible to aggregate the data with higher resolution. It is further assumed that $\mathbf{w}_{2t}$ has a much wider spatial coverage than $\mathbf{w}_{1t}$ ($n_M \ll n_G$) but more biased (Lawson et al., 2016), and that both $\mathbf{w}_{1t}$ and $\mathbf{w}_{2t}$ are error-prone realizations of the same process of interest $\boldsymbol{x}_t$. This implies that:

$$\begin{aligned} \mathbf{w}_{1t} &= f_1(\boldsymbol{x}_t, \boldsymbol{\theta}_1) + \boldsymbol{e}_{1t} \\ \mathbf{w}_{2t} &= f_2(\boldsymbol{x}_t, \boldsymbol{\theta}_2) + \boldsymbol{e}_{2t}, \end{aligned} \tag{4.4}$$

where $f_1(\cdot)$ and $f_2(\cdot)$ are some deterministic functions of the process $\boldsymbol{x}_t$ with bias parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$, respectively. The terms $\boldsymbol{e}_{1t}$ and $\boldsymbol{e}_{2t}$ are assumed independent error components. Typically, some simplifying assumptions are made on $f_1(\cdot)$ and $f_2(\cdot)$ to facilitate model inference. For instance, in the classical INLA approach (Section 2.5.2) (Rue et al., 2009), it is a requirement for the predictor to be a linear (deterministic) function of the latent Gaussian parameters; although recently, the class of models that can be fitted using INLA has been extended to those which are non-linear in the latent parameters using an iterative INLA approach (see Section 2.5.4), and which can be easily implemented using the `inlabru` package in `R` (Lindgren et al., 2024). Extending the above framework to more than two data sources is straightforward, as a new data set would be treated as yet another error-prone realization of the latent process of interest.

4.3.2 Proposed model

Using Equations (4.4) to represent the two data sources, the following data fusion model is proposed:

$$x(\mathbf{s}, t) = \boldsymbol{\beta}^\intercal \mathbf{z}(\mathbf{s}, t) + \xi(\mathbf{s}, t) \tag{4.5}$$

$$\mathrm{w}_1(\mathbf{s}_i, t) = x(\mathbf{s}_i, t) + e_1(\mathbf{s}_i, t), \quad i = 1, \ldots, n_M, \tag{4.6}$$

$$\mathrm{w}_2(\mathbf{g}_j, t) = \alpha_0(\mathbf{g}_j, t) + \alpha_1 x(\mathbf{g}_j, t) + e_2(\mathbf{g}_j, t), \quad j = 1, \ldots, n_G. \tag{4.7}$$

The latent process of interest, $x(\mathbf{s}, t)$, is modelled as a linear function of some known covariates $\mathbf{z}(\mathbf{s}, t)$ (including an intercept) and a random field $\xi(\mathbf{s}, t)$. The observed data $\mathrm{w}_1(\mathbf{s}_i, t)$ from the weather stations are assumed to be unbiased realizations of the latent process $x(\mathbf{s}, t)$ with an additive error term $e_1(\mathbf{s}_i, t) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_1}\big), i = 1, \ldots, n_M$, as in Equation (4.6). The iid assumption is justified by the fact that the weather stations are sparsely located in the spatial domain and operate independently of each other. The data $\mathrm{w}_2(\mathbf{g}_j, t)$ from the numerical weather forecast model are assumed to be biased realizations of the latent values $x(\mathbf{g}_j, t)$, $j = 1, \ldots, n_G$. I specify both an additive and a multiplicative bias for $\mathrm{w}_2(\mathbf{g}_j, t)$. I assume that the additive bias $\alpha_0(\mathbf{g}_j, t)$ varies in both space and time, and refer to it as the *error field*. On the other hand, I assume that the multiplicative bias α_1 is constant. For temperature, this is justified based on Figure 4.2. For the other two meteorological variables, a spatially-varying multiplicative bias could better fit the data. Notice, however, that having a spatially-varying multiplicative bias in the model would pose a greater computational challenge as the model would contain a product of two random fields. For this reason, in this work, I have chosen to consider a constant multiplicative bias. Finally, the model for $\mathrm{w}_2(\mathbf{g}_j, t)$ contains another unstructured error $e_2(\mathbf{g}_j, t) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_2}\big)$.

Both the spatio-temporal field $\xi(\cdot, t)$ in Equation (4.5) and the error field $\alpha_0(\cdot, t)$ in Equation (4.7) are modelled using a Matérn Gaussian space-time field. In particular, I assume that

$$\begin{aligned} \xi(\mathbf{s}, t) &= \phi_1 \, \xi(\mathbf{s}, t-1) + \omega_1(\mathbf{s}, t) \\ \alpha_0(\mathbf{s}, t) &= \phi_2 \, \alpha_0(\mathbf{s}, t-1) + \omega_2(\mathbf{s}, t), \end{aligned} \tag{4.8}$$

where $|\phi_1| < 1$ and $|\phi_2| < 1$ model the temporal dependence in a first order autoregressive (AR1) fashion, while $\omega_1(\mathbf{s}, t)$ and $\omega_2(\mathbf{s}, t)$ are time-independent Gaussian innovation processes with Matérn covariance structure, i.e.,

$$\mathrm{Cov}\Big(\omega_h(\mathbf{s}_i, t), \omega_h(\mathbf{s}_j, u)\Big) = \begin{cases} 0 & t \neq u \\ \Sigma^{(h)}_{i,j} & t = u \end{cases} \tag{4.9}$$

$$\Sigma^{(h)}_{i,j} = \frac{\sigma^2_h}{2^{\nu_h - 1}\Gamma(\nu_h)} (\kappa_h \, \|\mathbf{s}_i - \mathbf{s}_j\|)^{\nu_h} K_{\nu_h}(\kappa_h \, \|\mathbf{s}_i - \mathbf{s}_j\|), \tag{4.10}$$

for $h = 1, 2$. The model unknowns are the marginal variance σ_h^2 and scaling parameter κ_h. The smoothness parameter ν_h is fixed to 1 since this is poorly identified in many applications (Lindgren et al., 2011) (also see Section 2.6). The scaling parameter is related to the range parameter ρ_h via the empirically derived relationship $\rho_h \approx \frac{\sqrt{8\nu_h}}{\kappa_h}$. I use the SPDE approach to represent the Matérn fields in the model as Gaussian Markov random fields, which yields a sparse precision matrix, consequently making the computation efficient (Section 2.6) (Lindgren et al., 2011).

In the data problem, as presented in Section 4.1, the Philippines is an archipelagic country, consisting of numerous islands that are spatially disconnected. The models employed in this chapter do not account for these geographical discontinuities or barriers, since the physical atmospheric processes under study are largely continuous between land and sea. For example, rainfall is driven by atmospheric circulation and convection, and cloud movements do not stop at shorelines. A barrier SPDE model (Bakka et al., 2019) would be more appropriate for other contexts such as modeling disease spread, socio-economic diffusion restricted by geography, or ecological distributions where animal populations cannot cross water.

The model structure in Equations (4.8) to (4.10) implies that $\boldsymbol{\xi}_t|\boldsymbol{\xi}_{t-1} \sim \mathcal{N}\left(\phi_1\boldsymbol{\xi}_{t-1}, \boldsymbol{\Sigma}^{(1)}\right)$ and $\boldsymbol{\alpha}_{0_t}|\boldsymbol{\alpha}_{0_{t-1}} \sim \mathcal{N}\left(\phi_2\boldsymbol{\alpha}_{0_{t-1}}, \boldsymbol{\Sigma}^{(2)}\right)$ for $t = 2, \ldots, T$, where $\boldsymbol{\xi}_t = \begin{pmatrix} \xi(\mathbf{s}_1, t) & \ldots & \xi(\mathbf{s}_{n_M}, t) \end{pmatrix}^\intercal$, $\boldsymbol{\alpha}_{0_t} = \begin{pmatrix} \alpha_0(\mathbf{g}_1, t) & \ldots & \alpha_0(\mathbf{g}_{n_G}, t) \end{pmatrix}^\intercal$, and that $\boldsymbol{\Sigma}^{(1)}$ and $\boldsymbol{\Sigma}^{(2)}$ are dense covariance matrices whose elements are given in Equation (4.10). Both $\xi(\mathbf{s}, t)$ and $\alpha_0(\mathbf{s}, t)$ follow the stationary distribution at $t = 1$, i.e., $\xi(\mathbf{s}, 1) \sim \mathcal{N}\left(0, \sigma_1^2/(1-\phi_1^2)\right)$ and $\alpha_0(\mathbf{g}, 1) \sim \mathcal{N}\left(0, \sigma_2^2/(1-\phi_2^2)\right)$.

The performance of the proposed model is compared with two benchmark approaches. The first one, denoted *stations-only model*, uses only the data from the stations. This model essentially fits Equations (4.5) and (4.6) only. The second benchmark model is the regression calibration model discussed in Section 2.7.2. In the spatio-temporal scenario, I assume that the additive and multiplicative biases in the calibration model are both spatially and temporally varying. Note that a model relying exclusively on GSM data is not considered a benchmark model, since the GSM data, like any proxy data, are inherently biased. This bias has been shown in Figure 4.2, which shows the discrepancy in the values between the observed values at

the weather stations' locations and the corresponding predicted values using a model fitted solely using GSM data.

4.4 Model estimation

For estimation, I use the integrated nested Laplace approximation (INLA) (Section 2.5.2). The stations-only model and the regression calibration model are standard spatial models and are straightforward to estimate using INLA (Cameletti et al., 2013). On the other hand, the proposed model in Equations (4.5) to (4.7) can be tricky. It is useful to rewrite the model, in vector form, as follows:

$$\begin{aligned}
\mathbf{w}_{1t} &= \mathbf{Z}_t\boldsymbol{\beta} + \boldsymbol{\xi}_t + \boldsymbol{e}_{1t}, \quad \boldsymbol{e}_{1t} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2_{e_1}\mathbb{I}\big) \\
\boldsymbol{\xi}_t &= \phi_1\boldsymbol{\xi}_{t-1} + \boldsymbol{\omega}_{1t} \\
\mathbf{w}_{2t} &= \boldsymbol{\alpha}_{0_t} + \alpha_1(\mathbf{Z}_t\boldsymbol{\beta} + \boldsymbol{\xi}_t) + \boldsymbol{e}_{2t}, \quad \boldsymbol{e}_{2t} \sim \mathcal{N}\big(\mathbf{0}, \sigma^2_{e_2}\mathbb{I}\big) \\
\boldsymbol{\alpha}_{0_t} &= \phi_2\boldsymbol{\alpha}_{0_{t-1}} + \boldsymbol{\omega}_{2t}.
\end{aligned} \tag{4.11}$$

The model specification involves two likelihood components: $\mathbf{w}_{1t}$ and $\mathbf{w}_{2t}$. The latent part of the model includes the fixed effects $\boldsymbol{\beta}$, the space-time effects $\boldsymbol{\xi}_t$, and the error field $\boldsymbol{\alpha}_{0_t}$, $t = 1, \ldots, T$. The fixed effects $\boldsymbol{\beta}$ are given a non-informative Gaussian prior, while the random fields $\boldsymbol{\xi}_t$ and $\boldsymbol{\alpha}_{0_t}$ follow a Gaussian autoregressive structure as described in Equations (4.8) to (4.10). The hyperparameters include the multiplicative bias α_1, the parameters linked to $\boldsymbol{\xi}_{1t}$ (σ_1, ρ_1, and ϕ_1), the parameters linked to $\boldsymbol{\alpha}_{0_t}$ (σ_2, ρ_2, and ϕ_2), and the measurement error variance parameters $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$. The model hyperparameters, except for α_1, are given penalized complexity (PC) priors (Fuglstad et al., 2019; Simpson et al., 2017). PC priors are weakly informative priors which penalize the complexity of Gaussian random fields by shrinking the range towards infinity and the marginal variance towards zero. These are defined and expressed through probability statements of the type $\mathbb{P}(\sigma > \sigma_o) = \zeta_1$ and $\mathbb{P}(\rho < \rho_o) = \zeta_2$, where $\zeta_1, \zeta_2 \in (0, 1)$ are probability values chosen by the user, while σ_o and ρ_o are user-defined values of the standard deviation and range parameter, respectively.

The proposed data fusion model falls in the class of models that can be fitted using INLA since, given the hyperparameters, the latent field is Gaussian. However,

estimating α_1, which acts as a scaling parameter for the Gaussian field $\mathbf{Z}_t\boldsymbol{\beta} + \boldsymbol{\xi}_t$, can be difficult as the optimizer could run into numerical issues. Hence, I explore the use of a Bayesian model averaging (BMA) approach with INLA (Gómez-Rubio, 2020). This approach fits the data fusion model conditional on α_1, and then averages all the conditional INLA models to obtain the final posterior estimates. In addition, it is easy to determine a reasonable set of values for α_1: a value of 1 implies that the numerical model has no multiplicative bias, and the further the value of α_1 from 1, the more serious the multiplicative bias. In the data application, I do not expect α_1 to be very far from 1, particularly for temperature and relative humidity; thus, I defined a grid of α_1 values from 0.5 to 1.5 with a length step of 0.1.

4.4.1 Bayesian model averaging with INLA

Suppose all observed data is denoted by $\mathbf{Y}$, a latent parameter is denoted by x_j, and the hyperparameters are denoted by $\boldsymbol{\theta}$. INLA computes the posterior marginals based on the following integrals:

$$\pi(\theta_i|\mathbf{Y}) = \int \pi(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta}_{-i}$$
$$\pi(x_j|\mathbf{Y}) = \int \pi(x_j|\boldsymbol{\theta},\mathbf{Y})\pi(\boldsymbol{\theta}|\mathbf{Y})d\boldsymbol{\theta},$$

where $\boldsymbol{\theta}_{-i}$ denotes the vector of hyperparameters excluding θ_i. Suppose that $\boldsymbol{\theta} = \begin{pmatrix} \alpha_1 & \boldsymbol{\theta}_{-\alpha_1} \end{pmatrix}^\intercal$, where $\boldsymbol{\theta}_{-\alpha_1}$ includes all hyperparameters excluding α_1. The posterior marginals of $\boldsymbol{x}$ and $\boldsymbol{\theta}_{-\alpha_1}$ can then be expressed as

$$\pi(\cdot|\mathbf{Y}) = \int \pi(\cdot,\alpha_1|\mathbf{Y})d\alpha_1 = \int \pi(\cdot|\alpha_1,\mathbf{Y})\pi(\alpha_1|\mathbf{Y})d\alpha_1. \tag{4.12}$$

The probability density $\pi(\cdot|\alpha_1,\mathbf{Y})$ is a conditional marginal posterior which can be easily estimated using INLA for a fixed α_1. The density $\pi(\alpha_1|\mathbf{Y})$ in Equation (4.12) can be expressed as $\pi(\alpha_1|\mathbf{Y}) \propto \pi(\mathbf{Y}|\alpha_1)\pi(\alpha_1)$, where $\pi(\mathbf{Y}|\alpha_1)$ is the conditional marginal likelihood, and $\pi(\alpha_1)$ is the prior for α_1. The computation is done by specifying a grid of values for α_1, say $\alpha_1^{(k)}, k = 1,\ldots,K$, and then estimating the model conditional on each $\alpha_1^{(k)}$. Given this ensemble of INLA models, the weights for

model averaging are computed as:

$$w_k = \frac{\pi\Big(\mathbf{Y}|\alpha_1^{(k)}\Big)\pi\Big(\alpha_1^{(k)}\Big)}{\sum_{k=1}^{K}\pi\Big(\mathbf{Y}|\alpha_1^{(k)}\Big)\pi\Big(\alpha_1^{(k)}\Big)}. \tag{4.13}$$

The marginal posteriors given in Equation (4.12) are then computed as

$$\pi\Big(\cdot\,|\mathbf{Y}\Big) \approx \sum_{k=1}^{K}\pi\Big(\cdot\,|\alpha_1^{(k)},\mathbf{Y}\Big)w_k. \tag{4.14}$$

All other posterior quantities of interest are computed using model averaging. For example, the predicted $x(\mathbf{s},t)$ field is given by the following

$$\hat{x}(\mathbf{s},t) = \sum_{k=1}^{K}\left\{\mathbb{E}\Big[\boldsymbol{\beta}|\alpha_1^{(k)},\mathbf{Y}\Big]^{\intercal}\mathbf{z}(\mathbf{s},t) + \mathbb{E}\Big[\boldsymbol{\xi}|\alpha_1^{(k)},\mathbf{Y}\Big]\right\}w_k, \tag{4.15}$$

where $\mathbb{E}\Big[\cdot\,|\alpha_1^{(k)},\mathbf{Y}\Big]$ is evaluated with respect to the conditional marginal posteriors $\pi\Big(\boldsymbol{\beta}|\alpha_1^{(k)},\mathbf{Y}\Big)$ and $\pi\Big(\boldsymbol{\xi}|\alpha_1^{(k)},\mathbf{Y}\Big)$. Equation (4.15) is equivalent to

$$\begin{aligned}
\hat{x}(\mathbf{s},t) &= \sum_{k=1}^{K}\left\{\int\boldsymbol{\beta}\pi\Big(\boldsymbol{\beta}|\alpha_1^{(k)},\mathbf{Y}\Big)w_k d\boldsymbol{\beta}\right\}^{\intercal}\mathbf{z}(\mathbf{s},t) + \sum_{k=1}^{K}\left\{\int\boldsymbol{\xi}\pi\Big(\boldsymbol{\xi}|\alpha_1^{(k)},\mathbf{Y}\Big)w_k d\boldsymbol{\xi}\right\} \\
&= \left\{\int\boldsymbol{\beta}\Big[\sum_{k=1}^{K}\pi\Big(\boldsymbol{\beta}|\alpha_1^{(k)},\mathbf{Y}\Big)w_k\Big]d\boldsymbol{\beta}\right\}^{\intercal}\mathbf{z}(\mathbf{s},t) + \int\boldsymbol{\xi}\Big[\sum_{k=1}^{K}\pi\Big(\boldsymbol{\xi}|\alpha_1^{(k)},\mathbf{Y}\Big)w_k\Big]d\boldsymbol{\xi} \\
&= \left\{\int\boldsymbol{\beta}\pi(\boldsymbol{\beta}|\mathbf{Y})d\boldsymbol{\beta}\right\}^{\intercal}\mathbf{z}(\mathbf{s},t) + \int\boldsymbol{\xi}\pi(\boldsymbol{\xi}|\mathbf{Y})d\boldsymbol{\xi} \\
&= \mathbb{E}[\boldsymbol{\beta}|\mathbf{Y}]^{\intercal}\mathbf{z}(\mathbf{s},t) + \mathbb{E}[\boldsymbol{\xi}|\mathbf{Y}],
\end{aligned} \tag{4.16}$$

where $\mathbb{E}[\cdot|\mathbf{Y}]$ is evaluated with respect to the marginal posteriors which are approximated via Equation (4.14). Equation (4.15) implies that the predicted field $\hat{x}(\mathbf{s},t)$ can be computed by evaluating the predictor expression for $x(\mathbf{s},t)$ given in Equation (4.5) using the mean of the conditional marginal posteriors and then getting the weighted average using the weights w_k, while Equation (4.16) implies that it is equivalent to directly evaluating Equation (4.5) using the mean of the marginal posteriors in Equation (4.14). A disadvantage of the model averaging approach is that it requires fitting the models conditional on each α_1 value, which can be inefficient

especially for large spatio-temporal datasets.

4.5 Simulation Study

This section presents the results from a simulation study to assess the performance of the proposed data fusion model compared to two benchmark approaches: a stations-only model and a statistical (regression) calibration model (discussed in Section 2.7.2). Similar to Chapter 3, I use the Belo horizonte region in Brazil as the study domain, whose shapefile is available in the `R` package `spdep` (Bivand and Piras, 2015).

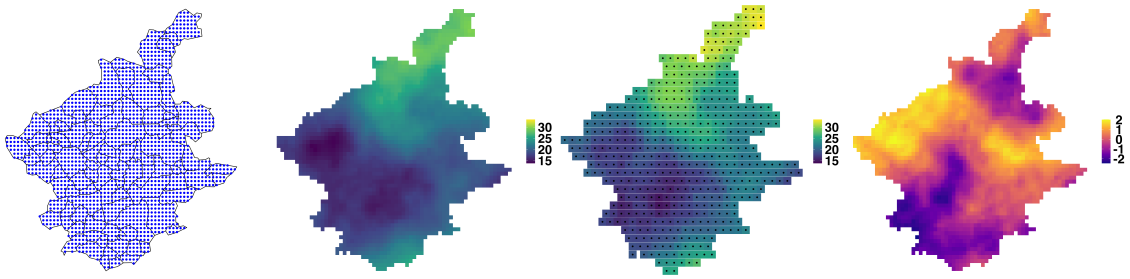
I perform the simulation study in a purely spatial context and simulate the process of interest as:

$$x(\mathbf{s}) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}) + \xi(\mathbf{s}),$$

where $\xi(\mathbf{s})$ is a Matérn random field with the following parameters: effective range $\rho_\xi = 2$ degrees, marginal standard deviation $\sigma_\xi = 3.16$, and smoothness parameter $\nu_\xi = 1$. The value of the range is such that the spatial correlation becomes negligible at a distance of ca 222 km which corresponds to half of the maximum distance in the study region. Moreover, $\mathrm{z}(\mathbf{s})$ is a known covariate and is simulated from a Matérn process with effective range of 3 degrees, a marginal variance of 1, and a smoothness parameter equal to 1. The fixed effects are $\beta_0 = 10$ and $\beta_1 = 3$.

The two observed datasets are simulated using Equations (4.6) and (4.7), respectively, but without the time component. It is assumed that $e_1(\mathbf{s}_i) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_1} = .25\big)$ and $e_2(\mathbf{g}_j) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_2} = .01\big)$. The values of $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$ are chosen based on the empirical results from the temperature model in Section 4.6.1. This implies that the noise in $\mathrm{w}_2(\mathbf{g}_j)$ is mainly attributed to the error field $\alpha_0(\mathbf{g})$, which is simulated from a Matérn process with range $\rho_{\alpha_0} = 1$, marginal standard deviation $\sigma_{\alpha_0} = 1$, and smoothness parameter $\nu_{\alpha_0} = 1$. The range parameter and marginal variance of $\xi(\mathbf{s})$ is chosen to be higher compared to $\alpha_0(\mathbf{g})$ based on the empirical results from the real data application discussed in Section 4.6. Finally, the constant multiplicative bias parameter is $\alpha_1 = 1.1$.

Figure 4.13a shows the dense simulation grid. Figure 4.13b shows a simulated $x(\mathbf{s})$ field, while Figure 4.13c shows a simulated proxy data, which is at a coarser

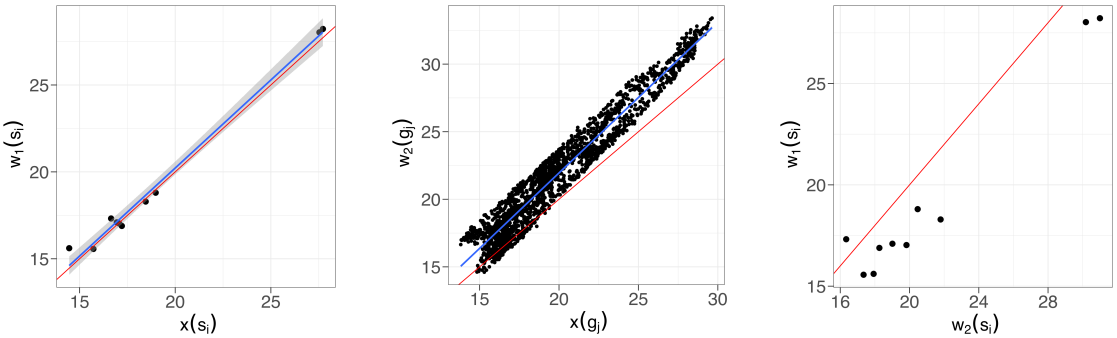


(a) Simulation grid **(b)** $x(\mathbf{s})$ **(c)** $\mathrm{w}_2(\mathbf{g}_j)$ **(d)** $\alpha_0(\mathbf{g})$

Figure 4.13: (a) dense simulation grid, (b) a simulated true field $x(\mathbf{s})$, (c) a simulated proxy data $\mathrm{w}_2(\mathbf{g}_j)$, (d) a simulated error field $\alpha_0(\mathbf{g})$

resolution than the simulation grid. In particular, the centroids of the grid cells in Figure 4.13c is a coarse subset of the points in Figure 4.13a. The corresponding simulated error field $\alpha_0(\mathbf{g})$ is shown in Figure 4.13d.

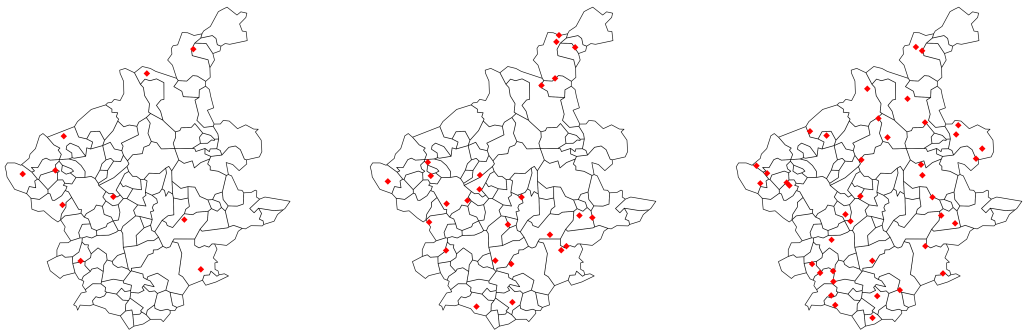
Figure 4.14c highlights a significant discrepancy between the stations data and the proxy data values at the stations' locations for a simulated data with $n_M = 10$ stations whose spatial locations are shown in Figure 4.15a. Figure 4.14c is similar to Figure 4.2 which shows discrepancies between the two data sources in the real data application. Moreover, the difference in bias severity between $\mathrm{w}_1(\mathbf{s}_i)$ and $\mathrm{w}_2(\mathbf{g}_j)$ is illustrated in Figures 4.14a and 4.14b, respectively. The data from the 10 stations closely align with the true values. On the other hand, $\mathrm{w}_2(\mathbf{g}_j)$ exhibits more bias and an overestimation of the true values due to the multiplicative bias parameter $\alpha_1 > 1$.



(a) $\mathrm{w}_1(\mathbf{s}_i)$ vs $x(\mathbf{s}_i)$ **(b)** $\mathrm{w}_2(\mathbf{g}_j)$ vs $x(\mathbf{g}_j)$ **(c)** $\mathrm{w}_1(\mathbf{s}_i)$ vs $\mathrm{w}_2(\mathbf{s}_i)$

Figure 4.14: (a) simulated observed values at 10 stations versus true values, (b) simulated proxy data versus true values, (c) simulated observed values at 10 stations versus proxy data values.

The main interest is to understand if jointly modelling the two data sources offers advantages and how these change with the sparsity of the stations data. I therefore consider three different scenarios for the number of stations. The first scenario con-



(a) $n_M = 10$ (b) $n_M = 25$ (c) $n_M = 40$

Figure 4.15: Spatial location of stations: (a) a sparse network, (b) a denser network but with an undersampled region, (c) a dense uniformly distributed network.

sists of only $n_M = 10$ stations with areas severely undersampled (Figure 4.15a). In the second case, there are $n_M = 25$ stations, and a large area that is undersampled (Figure 4.15b). The third case has $n_M = 40$ stations uniformly distributed over the study area (Figure 4.15c). The spatial locations are held constant for all the data replicates so that the configuration of the stations does not influence the results.

4.5.1 Model definition and estimation

I compare three modelling approaches: stations-only model, a regression calibration model, and the proposed data fusion model here specified as:

$$\begin{aligned} \mathrm{w}_1(\mathbf{s}_i) &= \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}_i) + \xi(\mathbf{s}_i) + e_1(\mathbf{s}_i), & e_1(\mathbf{s}_i) &\overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_1}\big), \\ \mathrm{w}_2(\mathbf{g}_j) &= \alpha_0(\mathbf{g}_j) + \alpha_1\Big(\beta_0 + \beta_1 \mathrm{z}(\mathbf{g}_j) + \xi(\mathbf{g}_j)\Big) + e_2(\mathbf{g}_j), & e_2(\mathbf{g}_j) &\overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_2}\big), \end{aligned}$$

$i = 1, \ldots, n_M; j = 1, \ldots, n_G$. The main interest here is in predicting the unknown field $x(\mathbf{s}) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}) + \xi(\mathbf{s})$. I assign β_0 and β_1 vague, zero mean, Gaussian priors. Penalized complexity (PC) priors are used for the variance parameters $\sigma^2_{e_1}$ and $\sigma^2_{e_2}$ and the parameters of the random fields $\xi(\cdot)$ and $\alpha_0(\cdot)$: ρ_ξ, σ_ξ, ρ_{α_0}, and σ_{α_0}. I define two scenarios on how to specify the PC priors. The first scenario, which I call *matching priors*, uses the actual values used to generate the data. The second scenario, which I call *non-matching priors*, uses $\sigma_{e_{1_\mathrm{o}}} = 1.5, \sigma_{e_{2_\mathrm{o}}} = 0.5, \sigma_{\xi_\mathrm{o}} = 1$, $\rho_{\xi_\mathrm{o}} = 0.5$, $\sigma_{\alpha_{0\mathrm{o}}} = .5$, $\rho_{\alpha_{0\mathrm{o}}} = .5$, which are arbitrarily chosen values. The probability value in the PC priors is set to $\zeta_1 = \zeta_2 = 0.5$ for all the parameters. As an example,

when defining the prior for $\sigma^2_{e_1}$, we have $\mathbb{P}(\sigma_{e_1} > 0.25) = 0.5$ for the matching prior scenario, and $\mathbb{P}(\sigma_{e_1} > 1.5) = 0.5$ for the non-matching prior scenario. Note that the matching priors are not necessarily more informative than the non-matching priors and that both cases are weakly informative.

I use the `inlabru` library (Lindgren et al., 2024) to fit the models. For the proposed model, the Bayesian model averaging approach with INLA is used, as discussed in Section 4.4. I define a regular grid of α_1 values centered on 1, and use a uniform prior for α_1 in computing the weights.

4.5.2 Model assessment

The performance of the three modelling approaches are compared by considering the accuracy in the predicted field $\hat{x}(\mathbf{s})$ and the estimates of model parameters. To predict the field, I consider the posterior mean $\mathbb{E}[x(\mathbf{s})|\mathbf{Y}]$ with the corresponding uncertainty given by the posterior standard deviation $\sqrt{\mathbb{V}[x(\mathbf{s})|\mathbf{Y}]}$, both evaluated over the grid shown in Figure 4.13a. The following metrics are used for model assessment:

1. Average squared error of the estimated field over $\mathcal{S}$: $\frac{1}{|\mathcal{S}|}\int_{\mathcal{S}} \Big(x(\mathbf{s}) - \mathbb{E}[x(\mathbf{s})|\mathbf{Y}]\Big)^2 d\mathbf{s}$
2. Average posterior uncertainty of the estimated field over $\mathcal{S}$: $\frac{1}{|\mathcal{S}|}\int_{\mathcal{S}} \sqrt{\mathbb{V}[x(\mathbf{s})|\mathbf{Y}]} d\mathbf{s}$
3. Average Dawid-Sebastiani (DS) score which is a measure of the closeness between an observed quantity of interest and the prediction distribution, say $\mathcal{F}$. The DS score is based on a coherent design criterion and is appropriate for predictive decision problems (Dawid and Sebastiani, 1999). Suppose $\mathbb{E}_{\mathcal{F}}[y]$ and $\mathbb{V}_{\mathcal{F}}[y]$ are the mean and variance, respectively, of the predictive distribution $\mathcal{F}(y)$. The DS score for a prediction on y is given by
$$\frac{\Big(y - \mathbb{E}_{\mathcal{F}}[y]\Big)^2}{\mathbb{V}_{\mathcal{F}}[y]} + \log\big(\mathbb{V}_{\mathcal{F}}[y]\big). \tag{4.17}$$
4. Relative error of each parameter estimate: e.g., $\left|\frac{\hat{\beta} - \beta}{\beta}\right|$, where $\hat{\beta}$ is the posterior mean of β, i.e., $\hat{\beta} = \mathbb{E}[\beta|\mathbf{Y}]$.
5. Posterior uncertainty in the parameter estimates: e.g., $\sqrt{\mathbb{V}[\beta|\mathbf{Y}]}$ for β.

The first metric is a measure of the average discrepancy between the estimated field $\hat{x}(\mathbf{s})$ and the true field $x(\mathbf{s})$, while the second metric assesses the average uncertainty in the estimated field. I approximate both integrals using the estimated values on the prediction grid. The third metric is another proper scoring rule that depends on the predictive mean and variance of the observed data (Gneiting and Raftery, 2007). As for the first two metrics, a lower value for the average DS score is preferred. Finally, the last two metrics look at the bias and uncertainty in the parameter estimates. All simulation results are computed based on 500 independent data replicates.

4.5.3 Simulation study results

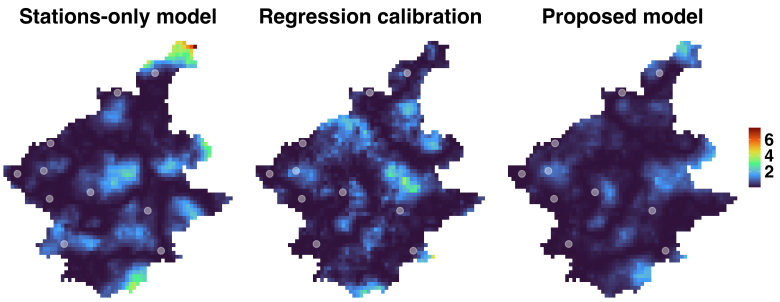


Figure 4.16: Comparison of squared errors for the simulated data in Figures 4.13 and 4.14. The errors from the proposed model are generally the smallest.

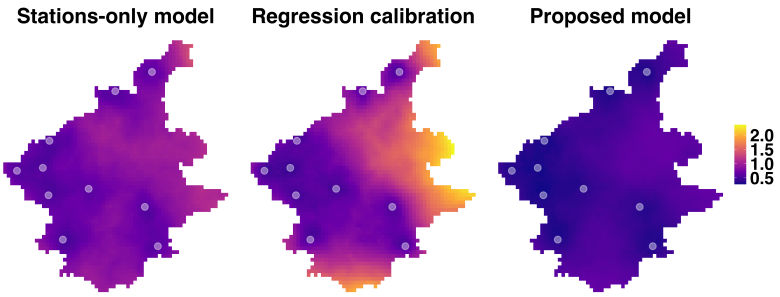
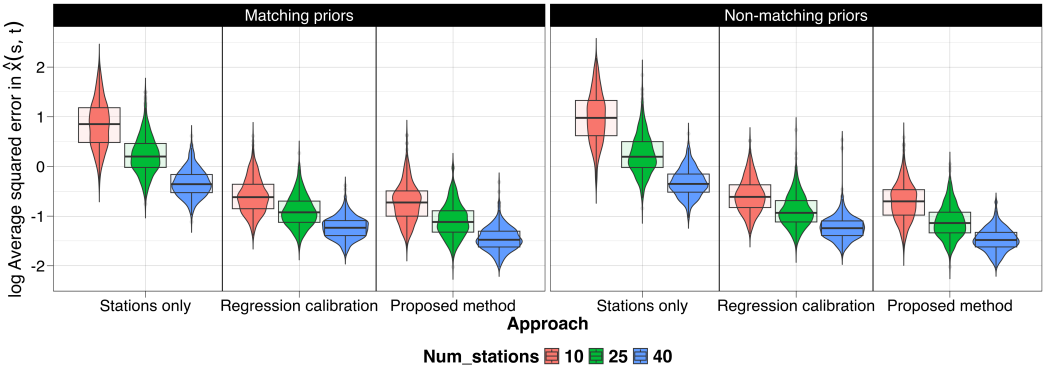


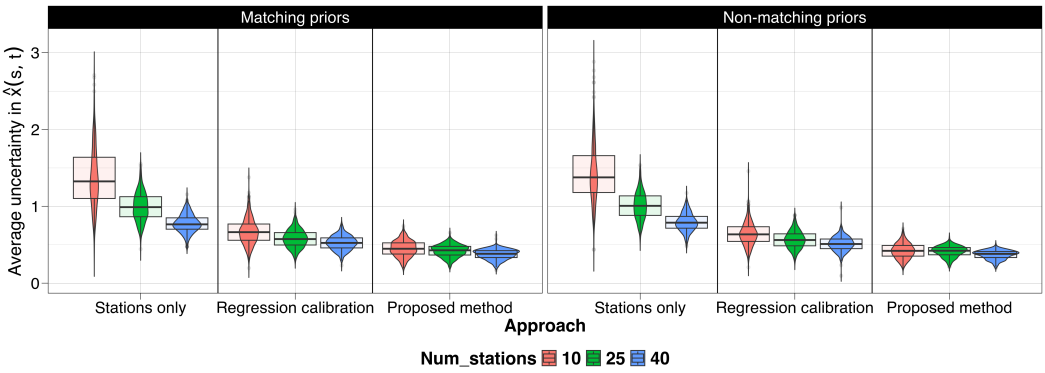
Figure 4.17: Comparison of the posterior uncertainty for the simulated data in Figures 4.13 and 4.14. The posterior uncertainty from the proposed model are the smallest.

Figures 4.16 and 4.17 show a comparison of the squared errors and the posterior standard deviation of the estimated fields, respectively, among the three different modelling approaches on the data example in Figures 4.13 and 4.14. The stations' location are shown as white points. The squared errors are largest for the stations-only model and smallest for the proposed data fusion model (see Figure 4.16). The posterior uncertainty in the estimated field is also the smallest for the proposed model (see Figure 4.17). As expected, the posterior uncertainty is smallest at the stations'

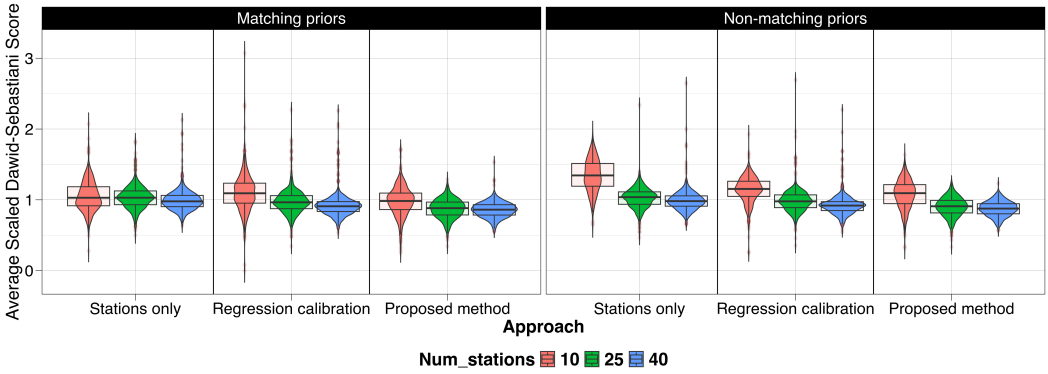
locations, which is very apparent for the stations-only model and the regression calibration model. The average squared errors for the stations-only model, regression calibration model, and the proposed model for this specific case are 0.53, 0.42, and 0.27, respectively, while the average posterior uncertainties are 0.78, 0.93, and 0.55, respectively. The average DS scores are 0.37, 0.08, and -0.23, respectively.



(a) Log average squared error of $\hat{x}(\mathbf{s})$



(b) Average posterior uncertainty of $\hat{x}(\mathbf{s})$



(c) Average DS score

Figure 4.18: Plots of the (a) log average squared errors (b) average posterior uncertainty and (c) average scaled DS scores from 500 simulated datasets with respect the number of stations, the priors used, and the modelling approach: stations-only model, regression calibration model, and proposed data fusion model. The posterior uncertainty from the proposed model is smallest. The stations-only model has the highest average squared error.

Figure 4.18a shows a plot of the log average squared errors of the estimated field $\hat{x}(\mathbf{s})$ based on 500 data replicates for the different simulation scenarios. The proposed

data fusion model generally gives smaller log average squared errors, especially when the data on the stations are very sparse. Moreover, the results show that there is no substantial difference with respect to the priors. Figure 4.18b shows the results for the average posterior uncertainty. It shows that the proposed model gives lower uncertainty estimates and that a higher number of stations is associated with lower posterior uncertainty. The same figure also shows that the specification of priors does not influence the results. The results for the DS scores are consistent with the insights from the previous two scores (see Figure 4.18c). Here, I scale the DS scores by adding the absolute value of the minimum in order to make the scores non-negative, and then applying log transformation. Furthermore, it also shows that the scores tend to decrease with the number of stations especially with the use of non-matching priors.

Figure 4.19 shows a summary of the model averaging weights of the INLA models for each α_1 value. The conditional INLA model with the highest weight corresponds to the true value $\alpha_1 = 1.1$, and with weights rapidly decreasing as α_1 goes further away from 1.1. The BMA weights do not vary much between the use of matching and non-matching priors, and the sparsity of the stations data.

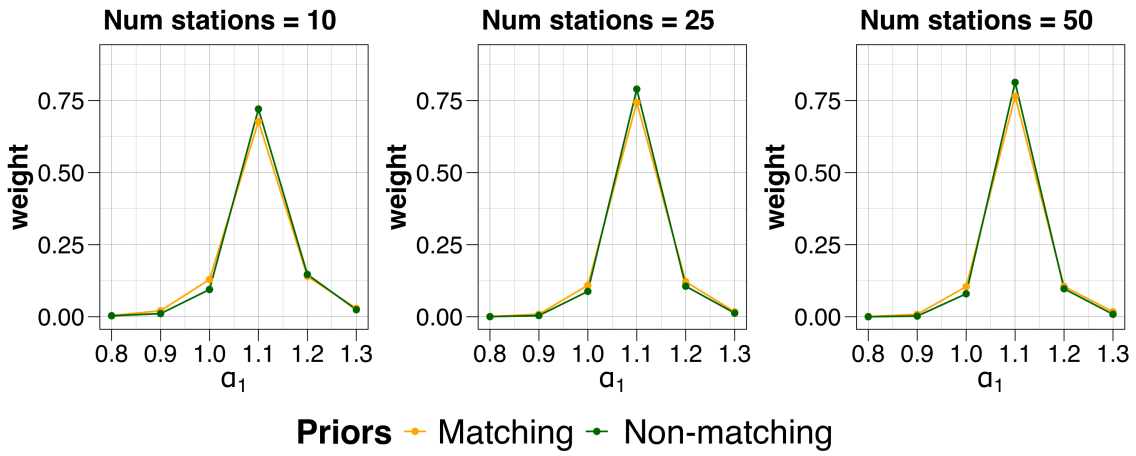


Figure 4.19: Average model averaging weights from 500 simulated datasets for different α_1 values in fitting the proposed data fusion model with respect to the sparsity of the stations data and the priors used. The correct value of α_1 has the highest weight.

Figure 4.20 compares the average relative error and average posterior uncertainty for the measurement error standard deviation σ_{e_1}. The proposed method generally outperforms the other two approaches, especially when the data from the stations are sparse.

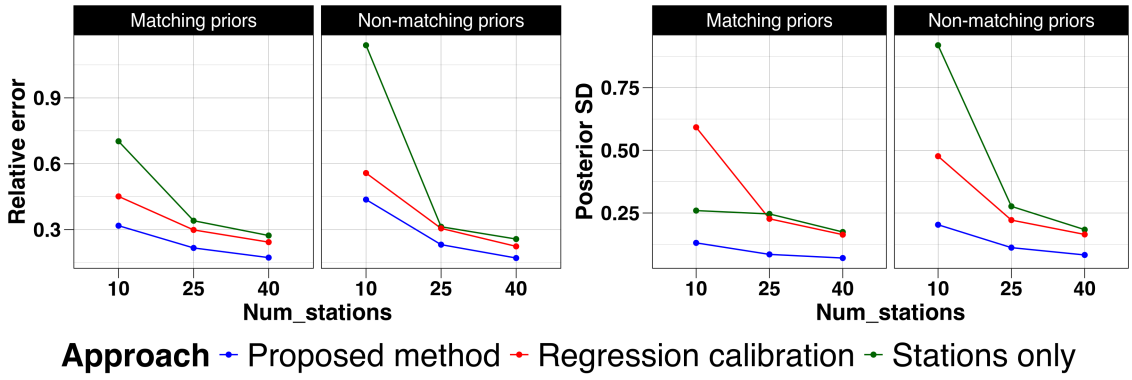


Figure 4.20: Plot of average relative errors and average posterior uncertainty from 500 simulated datasets for σ_{e_1}.

Figures A.1a and A.1b in Appendix A show the results for the marginal standard deviation and the range parameter of the spatial field $\xi(\mathbf{s})$, respectively, while Figure A.2 in Appendix A show the results for the fixed effects β_0 and β_1. Note that for the aforementioned parameters, a comparison can only be made between the stations-only model and the proposed model, since these parameters are not defined and specified in the regression calibration model, as shown in Equation (2.43) of Chapter 2. The results show that the proposed method outperforms the stations-only model in terms of the relative error and posterior uncertainty.

4.6 Results for meteorological data in the Philippines

In this section, I present the results from applying the three modelling approaches on the meteorological data in the Philippines. Sections 4.6.1, 4.6.2, and 4.6.3 discuss the results for temperature, relative humidity, and rainfall, respectively. Section 4.6.4 presents the results of the leave-group-out cross-validation.

4.6.1 Temperature

The mean (standard deviation) temperature in the stations and GSM data is 27.67°C (1.87) and 25.68°C (1.96), respectively. The percentage of missing data from the stations is only 3.8%; while the GSM data has no missing data. The following form

is assumed for the fixed effects in the latent process:

$$\text{Temperature}(\mathbf{s},t) = \beta_0 + \beta_1 \log\Big(\text{Elevation}(\mathbf{s},t)\Big) + \beta_2 \text{Cool}(\mathbf{s},t) + \beta_3 \text{ClimateType}(\mathbf{s},t). \tag{4.18}$$

The `Cool` variable in Equation (4.18) is a binary variable which takes a value of '1' for months December to February and a value of '0' for the other months. The `ClimateType` variable is also a binary variable which takes '1' for the eastern section of the country and '0' for the western section (see Section 4.1 for details).

In defining the three models (the stations-only model, the regression calibration model, and the proposed data fusion model), I used PC priors for the Matérn field parameters (Fuglstad et al., 2019; Simpson et al., 2017). The parameter values for the Matérn PC priors are as follows: $\rho_{1o} = \rho_{2o} = 300$ km, $\sigma_{1o} = 1.90$, and $\sigma_{2o} = 0.01$. The value for the range parameters is one-third the maximum distance of the spatial domain. The value for σ_{1o} is the standard deviation of the temperature values, while the value of σ_{2o} is chosen to be some value smaller than σ_{1o} based on preliminary model results. The variance parameters of $e_1(\mathbf{s}_i,t)$ and $e_2(\mathbf{g}_j,t)$ are also given PC priors, with $\sigma_{e_{1o}} = 0.2$ and $\sigma_{e_{2o}} = 0.01$. The probability value of all PC priors is set to 0.50. The rest of the model parameters are given default non-informative priors.

I defined a grid of values from 0.5 to 1.5 with a length step of 0.1 for the multiplicative bias parameter α_1, and which I assigned a uniform prior. An ensemble of INLA models were fitted for a fixed α_1, and the BMA weights are computed using Equation (4.13). The marginal log-likelihoods $\log \pi\Big(\mathbf{Y}|\alpha_1^{(k)}\Big)$ and the corresponding BMA weights w_k are shown in Table 4.1. The results show that the weight of the model with $\alpha_1 = 1$ is approximately equal to 1, while the weights for the other models are close to 0. This implies that there is no multiplicative bias which is consistent with the insights in Figure 4.2a.

Table 4.2 shows the posterior estimates of the fixed effects for the stations-only model and the proposed data fusion model. Note that these parameters are not explicitly specified and estimated using the regression calibration model, shown in Equation (2.43) of Chapter 2. The two models agree on the conclusions: it is colder at higher elevation, cooler during December to February, and areas in the western

Table 4.1: Marginal log-likelihood values conditional an α_1 and the corresponding BMA weights for the temperature data fusion model

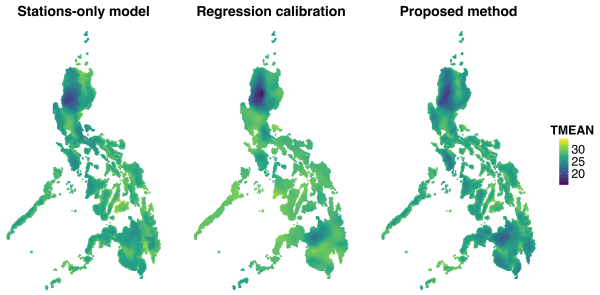
α_1	$\log \pi(\mathbf{Y}\|\alpha_1)$	w_k
0.5	-978.295	0.0000
0.6	-914.791	0.0000
0.7	-849.673	0.0000
0.8	-778.493	0.0000
0.9	-697.501	0.0001
1	-688.142	1.0000
1.1	-811.927	0.0000
1.2	-899.762	0.0000
1.3	-949.719	0.0000
1.4	-2265.074	0.0000
1.5	-2329.848	0.0000

Table 4.2: Posterior estimates of fixed effects for the temperature model – stations-only model versus proposed data fusion model

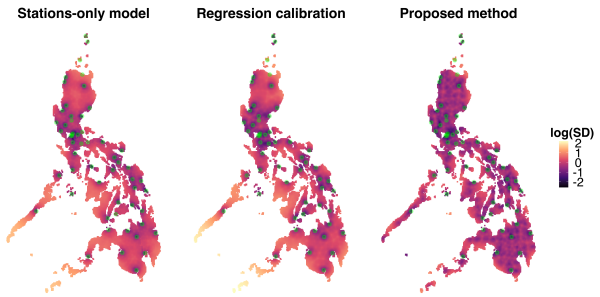
	Stations-only model				**Proposed model**			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
β_0	28.664	2.6270	23.510	33.818	28.919	4.603	19.897	37.940
β_1, log(Elevation)	-0.631	0.094	-0.815	-0.446	-0.709	0.051	-0.808	-0.609
β_2, Cool	-0.683	0.198	-1.072	-0.295	-0.6178	0.177	-0.965	-0.271
β_3, Climate Type	2.183	0.699	0.813	3.553	0.606	0.337	-0.054	1.266

section of the country are cooler. The `ClimateType` variable is not significant in the proposed data fusion model, but is significant in the stations-only model. Finally, the uncertainty in the fixed effects, expect for β_0, is smaller for the proposed model.

The posterior estimates of the hyperparameters from the three approaches are quite similar (see Table A.1 in Appendix A). In the proposed data fusion model, the estimated range of the spatial field, $\hat{\rho}_1$, is higher than the one for the error field, $\hat{\rho}_2$, indicating that $\xi(\mathbf{s}, t)$ is smoother than $\alpha_0(\mathbf{g}, t)$. The spatial correlation in the temperature spatial field and the error field becomes negligible at a distance of around 765 km and 113 km, respectively. Also, the estimated marginal standard deviation $\hat{\sigma}_1$ of the spatial field is much larger than that of the error field $\hat{\sigma}_2$. The estimated autocorrelation parameters $\hat{\phi}_1$ and $\hat{\phi}_2$ in both the spatial field and error field are close to 1 which suggests a high degree of dependence in time. Moreover, the posterior estimates of the regression calibration model are shown in Table A.2 of Appendix A. The 2.5^{th} and 97.5^{th} percentile of the multiplicative bias estimates of the regression calibration model are 0.9 and 1.1, respectively. These values do not change much in space and time, which justifies the assumption of a constant α_1 in the proposed



(a) Predicted temperature fields



(b) Posterior uncertainty (log scale)

Figure 4.21: Comparison of the estimated temperature fields and corresponding posterior uncertainties (log scale) for August 2019. The posterior uncertainties from the proposed data fusion model are smaller.

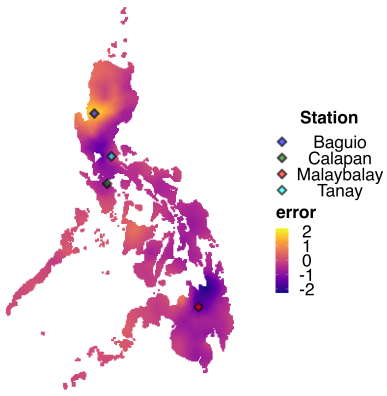
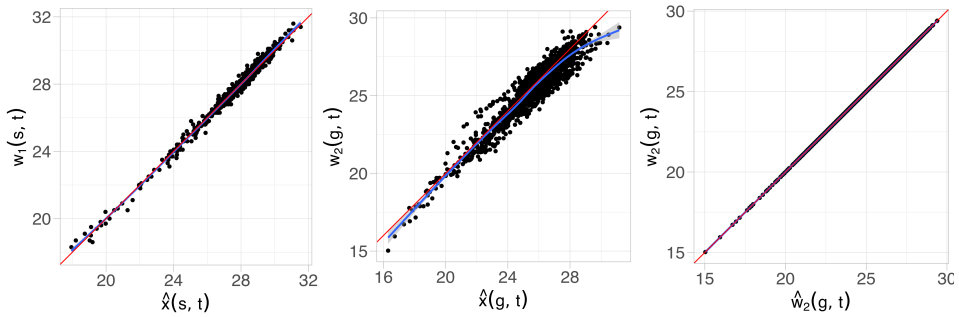


Figure 4.22: Estimated error field for the temperature model for August 2019. The estimated error fields at the specific stations correspond to the additive bias shown in Figure 4.2.

model.

Figure 4.21a shows that the estimated temperature fields (for August 2019) among the three approaches are very similar. The dark spot in the northern part of the country is primarily mountainous which makes it cooler than the other parts of the country. Moreover, Figure 4.21b shows the corresponding posterior standard deviations (SD) in log scale. The uncertainty is higher for the two benchmark approaches, especially in the islands in the lower left portion. The average posterior SD of the estimated fields from the stations-only model, regression calibration model, and the data fusion model is 1.239, 1.123, and 0.771, respectively. Moreover, the posterior SD is smaller at the stations' locations (green points) which is more apparent for the two benchmark approaches. Figure A.3 of Appendix A shows the estimated spatial fields, $\hat{\xi}(\mathbf{s}, t)$, for the three approaches and for the same month. The spatial structure looks quite similar as the estimated temperature fields in Figures 4.21a.

Figure 4.22 shows the estimated error field $\hat{\alpha}_0(\mathbf{g}_j, t)$ for the GSM data in August 2019. This plot can be compared to Figure 4.2a which shows the discrepancies in the values between the weather stations and the GSM outcomes. In particular, Figure



(a) $w_1(\mathbf{s}, t)$ vs $\hat{x}(\mathbf{s}, t)$ **(b)** $w_2(\mathbf{g}, t)$ vs $\hat{x}(\mathbf{g}, t)$ **(c)** $w_2(\mathbf{g}, t)$ vs $\hat{w}_2(\mathbf{g}, t)$

Figure 4.23: Plot of observed temperature values versus predicted values using the proposed data fusion model for (a) weather stations, (b) GSM data, and (c) calibrated GSM data. The blue line is the smooth local regression curve, while the red line is the identity line.

4.22 shows that the estimated additive bias around Baguio station (in blue) is the highest, indicating that the GSM overestimated the temperature in this region. This is consistent with Figure 4.2 which shows that the GSM values exceed the observed data at Baguio station. Similarly, the estimated additive bias around Malaybalay (in red) is negative, which aligns with the negative bias seen in the GSM outcomes for this area. For Tanay station (in cyan), the estimated additive bias is close to zero, which is also consistent with Figure 4.2 which shows little to no bias in the GSM outcomes at this location.

Figure 4.23a shows a close correspondence between the observed values at the stations $w_1(\mathbf{s}_i, t)$ and the corresponding predicted latent values $\hat{x}(\mathbf{s}_i, t)$ using the proposed data fusion model. Figure 4.23b, which shows a scatterplot between the observed GSM values $w_2(\mathbf{g}_j, t)$ and the corresponding predicted latent values $\hat{x}(\mathbf{g}_j, t)$, indicates a strong bias in the GSM values, with several points that are either overestimated or underestimated. Finally, Figure 4.23c shows a very close correspondence between the GSM data values $w_2(\mathbf{g}_j, t)$ and the predicted values $\hat{w}_2(\mathbf{g}_j, t) = \hat{\alpha}_0(\mathbf{g}_j, t) + \hat{\alpha}_1\hat{x}(\mathbf{g}_j, t)$. This implies that the error field can be effectively used to calibrate the GSM outcomes for temperature via $\hat{x}(\mathbf{g}_j, t) = \Big(w_2(\mathbf{g}_j, t) - \hat{\alpha}_0(\mathbf{g}_j, t)\Big)/\hat{\alpha}_1$.

4.6.2 Relative humidity

The mean (standard deviation) of the observed relative humidity (RH) is 81.51 (6.06) for the stations and 83.33 (4.86) for the GSM. The percentage of missing values from the stations data is only 3.74%. Although the RH variable is bounded from 0 to 100,

I assumed a Gaussian likelihood for the response variable, which is reasonable since the range of the RH values in the data is from 61 to 94.36, i.e., none of the values are equal to the bounds. Also, all the predicted values from the models are well within the bounds. A model which properly constrains the values can be considered in a future work.

The predictor expression for the fixed effects in the latent process is as follows:

$$\begin{aligned}\log\Big(\text{RH}(\mathbf{s},t)\Big) = \beta_0 + \beta_1 \log \text{Temperature}(\mathbf{s},t) + \beta_2\Big(\log \text{Temperature}(\mathbf{s},t)\Big)^2 \\ + \beta_3 \log\Big(\text{Elevation}(\mathbf{s},t)\Big) + \beta_4 \text{ClimateType}(\mathbf{s},t).\end{aligned} \tag{4.19}$$

`Elevation` and `ClimateType` are also used as predictors. In addition, I used log temperature and its quadratic term, as recommended by PAGASA (2023). Such non-linear relationship between RH and temperature is also established in the atmospheric science literature (Goody, 1995). A log transformation on temperature is possible since the Philippines is a tropical country with mean temperature ranging from 16°C to 32°C. I used the predictions generated from the temperature model in Section 4.6.1 as input in Equation (4.19). In the current results, the uncertainty from the predicted values of temperature are not accounted for, but this can be considered in a future work.

PC priors are used for the Matérn field parameters. For the range, I used the same values as in Section 4.6.1. For the marginal standard deviation, I set $\sigma_{1_0} = 0.08$ and $\sigma_{2_0} = 0.01$. The variance parameters of $e_1(\mathbf{s}_i, t)$ and $e_2(\mathbf{g}_j, t)$ are also given PC priors, with $\sigma_{e_{1_0}} = 0.01$ and $\sigma_{e_{2_0}} = 0.004$. The probability value in the PC priors are also set equal to 0.50. The rest of the model parameters are given the default non-informative priors.

I used the same grid of α_1 values as Section 4.6.1 to fit the conditional INLA models. The results shows that the model with $\alpha_1 = 1$ gave the highest marginal log-likelihood value and with a weight close to 1, while the rest of the α_1 values have weights close to 0. The marginal log-likelihoods $\log \pi\Big(\mathbf{Y}|\alpha_1^{(k)}\Big)$ and the corresponding BMA weights w_k are shown in Table A.3 of Appendix A.

Table 4.3 shows the posterior estimates of the model fixed effects for the stations-only model and the proposed data fusion model. The estimates are quite similar,

although the `ClimateType` variable is not significant in the stations-only model. The results show that there is a significant non-linear relationship between temperature and relative humidity, and that the elevation variable is negatively associated with relative humidity. Moreover, the `ClimateType` variable is positively related with relative humidity, which means that areas in the eastern section of the country have higher relative humidity, on average, than the western part.

Table 4.3: Posterior estimates of fixed effects for the relative humidity model – stations-only model versus proposed data fusion model

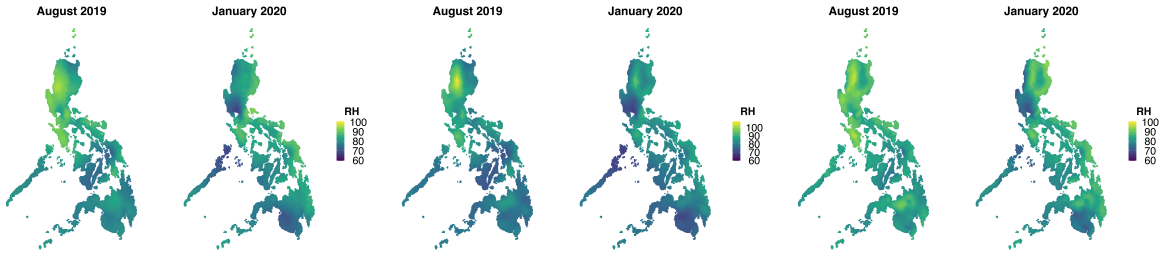
	Stations-only model				Proposed model			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
β_0	4.451	0.030	4.392	4.509	4.451	0.050	4.353	4.548
β_1, log(Temperature)	0.567	0.048	0.473	0.661	0.790	0.039	0.714	0.866
β_2, log(Temperature)^2	-0.173	0.014	-0.201	-0.145	-0.237	0.012	-0.260	-0.215
β_3, log(Elevation)	-0.008	0.003	-0.014	-0.002	-0.013	0.002	-0.017	-0.010
β_4, Climate Type	0.028	0.0180	-0.007	0.063	0.026	0.009	0.009	0.043

As in Section 4.6.1, the range ρ_1 of the spatial field in the latent process $\xi(\mathbf{s}, t)$ is estimated to be larger than the range ρ_2 of the error field $\alpha_0(\mathbf{s}, t)$ (see Table A.4 of Appendix A). Moreover, the estimated marginal variance of the spatial field is also larger than that of the error field. The estimates of the AR parameter are both close to 1, although the estimated value of the parameter for the spatial field is higher than that of the error field. Moreover, Table A.5 of Appendix A shows the posterior estimates of the regression calibration model for relative humidity. The 2.5$^{\text{th}}$ and 97.5$^{\text{th}}$ percentile of the multiplicative bias estimates of the regression calibration model are 0.96 and 1.02, respectively. The values are close to 1, which justifies the assumption of a constant α_1 in the proposed model, and agrees with the results from the proposed model.

Figure 4.24 shows the estimated relative humidity fields for two different months: August 2019 and January 2020. These two specific months were chosen since August is a rainy month while January is a dry month (PAGASA, 2023). The predicted fields show very similar structure, although it is apparent that there is more smoothing in the estimates from the stations-only model. The estimated fields show that in the eastern section of the country, the level of relative humidity is similar for the two months. On the other hand, in the western section, particularly in the northwestern section, relative humidity is very high in August, and very low in January. These

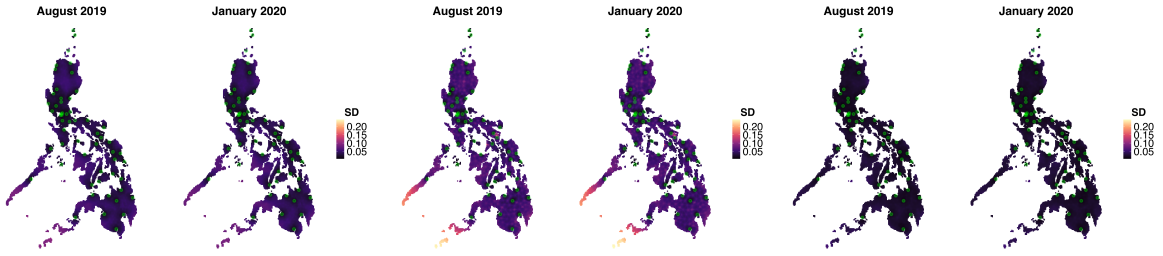
dynamics in relative humidity is consistent with the climate types in the Philippines (Coronas, 1920; Kintanar, 1984a; PAGASA, 2023). Figure 4.25 shows the corresponding uncertainty estimates of the estimated relative humidity fields. As expected, the proposed data fusion model has the lowest posterior uncertainty. The average of the posterior standard deviation in the predicted fields from the stations-only, regression calibration, and the proposed data fusion model is 0.040, 0.055, and 0.022, respectively.

The estimated spatial field of the latent process, $\hat{\xi}(\mathbf{s}, t)$, for the same two months are shown in Figure A.4 of Appendix A. The spatio-temporal dynamics observed in Figure 4.24 are also evident in the estimated spatial fields. The estimated error fields for the same two months are shown in Figure A.5 of Appendix A. Finally, Figure A.6 of Appendix A shows different scatterplots that indicate a close correspondence between the observed and predicted values, and a strong bias in the GSM values.



(a) Stations-only model **(b)** Regression calibration model **(c)** Proposed data fusion model

Figure 4.24: Comparison of estimated relative humidity fields for August 2019 and January 2020: (a) stations-only model, (b) regression calibration model, and (c) proposed data fusion model. There is more smoothing in the estimated fields using the stations-only model.



(a) Stations-only model **(b)** Regression calibration model **(c)** Proposed data fusion model

Figure 4.25: Posterior uncertainty of the estimated relative humidity fields in Figure 4.24. The posterior uncertainty in the estimated field from using the stations-only model is much higher.

4.6.3 Rainfall

The mean (standard deviation) of the cumulative monthly rainfall (in mm) is 220.27 (207.69) for the stations data and 166.12 (114.99) for the GSM data. Since the Philippines has high amounts rainfall and the values are aggregated monthly, there are very few zeros in the data (1.52% for the stations data and 0% for the GSM).

The predictor expression for the fixed effects is as follows:

$$\begin{aligned}\log\Big(\text{Rainfall}(\mathbf{s},t)+1\Big) = \beta_0 &+ \beta_1 \log \text{Temperature}(\mathbf{s},t) + \beta_2\Big(\log \text{Temperature}(\mathbf{s},t)\Big)^2 \\ &+ \beta_3 \text{Season}(\mathbf{s},t) + \beta_4 \text{ClimateType}(\mathbf{s},t) \\ &+ \beta_5 \text{ClimateType}(\mathbf{s},t) \times \text{Season}(\mathbf{s},t).\end{aligned} \tag{4.20}$$

The `Season` variable is binary and takes a value of '1' for June to November (characterized as a wet period), and a value of '0' for the rest of the year (characterized as a dry period). As with the relative humidity model, the log temperature and its squared term are included as predictors, as recommended by PAGASA (2023). I also used the predictions from the temperature model as input in Equation (4.20), but without accounting for the posterior uncertainty when fitting the model. An interaction effect between `ClimateType` and `Season` was included to capture the climate dynamics of the country, as recommended by PAGASA (2023).

As for the previous models, PC priors are used for the Matérn field parameters. For the range parameters, I used the same values as before, while for the marginal standard deviations, I set $\sigma_{1_0} = 1.35$, and $\sigma_{2_0} = 0.01$. The variance parameters of $e_1(\mathbf{s}_i, t)$ and $e_2(\mathbf{g}_j, t)$ are also given PC priors, with $\sigma_{e_{1_0}} = 0.5$ and $\sigma_{e_{2_0}} = 0.26$. The probability value in the PC priors is equal to 0.50. The rest of the model parameters are given the default non-informative priors.

Figure 4.26 shows the estimated marginal posterior distribution of α_1, $\pi(\alpha_1|\mathbf{Y})$. The posterior mean is 0.6733 while the 95% credible interval estimate is (0.5607, 0.8353). Unlike the two previous climate variables, the multiplicative bias parameter for the GSM outcomes for rainfall is significantly different from 1, implying a more severe bias for rainfall outcomes. This is expected since Figure 4.2c shows a large

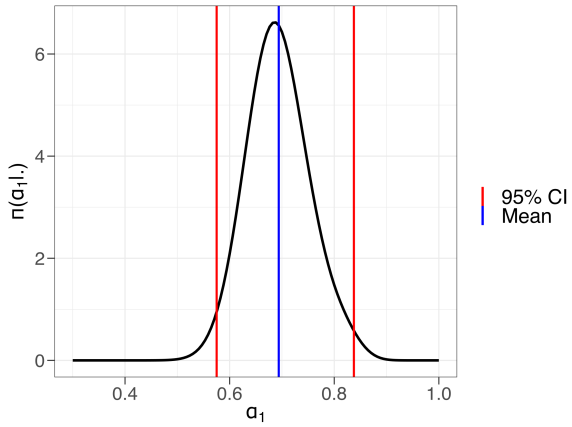


Figure 4.26: Estimated marginal posterior of α_1, $\pi(\alpha_1|\mathbf{Y})$, for the rainfall data fusion model. The posterior mean is 0.6733, while the 95% credible interval estimate is $(0.5607, 0.8353)$.

discrepancy between the interpolated GSM outcomes and the observed values at the weather stations for log-transformed rainfall. This also agrees with the insights from the LGOCV results discussed in Section 4.6.4.

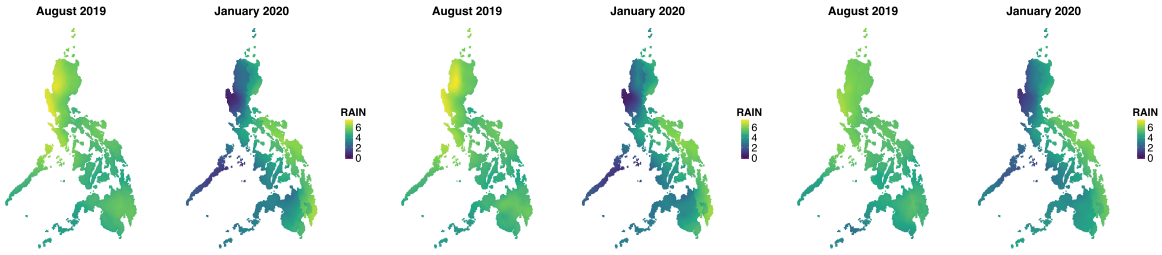
Table 4.4 shows the posterior estimates of the fixed effects for the stations-only model and proposed data fusion model. The results show that log temperature has a non-linear association with log rainfall amounts. Moreover, there is a significant interaction between `Season` and `ClimateType`: the western part of the country has a pronounced dry and wet season, while the eastern part of the country has a less pronounced dry and wet season and with more or less evenly distributed rainfall for the whole year. This can be confirmed in Figure 4.27 which shows the predicted log rainfall fields for two months - August 2019 (rainy month) and January 2020 (dry month). This climatic pattern is consistent with theory (Coronas, 1920; Kintanar, 1984a; PAGASA, 2023), and are the same seasonal dynamics observed for relative humidity in Section 4.6.2.

Table 4.4: Posterior estimates of fixed effects for the log rainfall model – stations-only model versus proposed data fusion model

	Stations-only model				**Proposed model**			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
β_0	4.427	0.360	3.722	5.132	4.759	0.377	3.931	5.484
β_1, log(Temperature)	2.186	0.454	1.296	3.076	1.672	0.430	0.911	2.541
β_2, log(Temperature)^2	-0.699	0.134	-0.961	-0.437	-0.570	0.122	-0.809	-0.354
β_3, Season	0.795	0.306	0.195	1.395	0.444	0.261	-0.020	0.973
β_4, Climate Type	1.183	0.146	0.898	1.469	0.657	0.112	0.458	0.873
β_5, Climate Type $\times$ Season	-0.844	0.162	-1.161	-0.527	-0.287	0.099	-0.461	-0.106

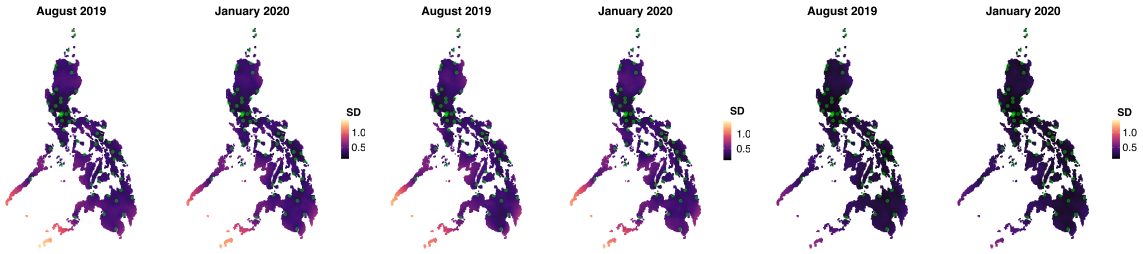
Table A.6 of Appendix A shows the posterior estimates of the hyperparameters

for the stations-only model and the proposed data fusion model. Similar to the previous meteorological variables, the estimated range $\hat{\rho}_1$ of the spatial field in the latent process is larger than the estimated range $\hat{\rho}_2$ of the error field. This is also true for the estimated marginal variances of the two fields. The estimated autocorrelation parameters of the two fields are very different, with the AR parameter for the error field being much larger. Table A.7 of Appendix A shows the posterior estimates of the regression calibration model for rainfall. The 2.5^{th} and 97.5^{th} percentiles of the multiplicative bias estimates of the regression calibration model are 0.35 and 1.35, respectively. The values vary significantly in space and time, raising doubt about the assumption of a constant α_1 in the proposed model. This agrees with the exploratory plot in Figure 4.2c, which is also noted as a limitation of the model for the rainfall variable. A model that specifies a spatially and temporally-varying multiplicative bias in the proposed model can be explored in a future work. This can be computationally difficult since this involves estimating the product of two random fields.



(a) Stations-only model **(b)** Regression calibration model **(c)** Proposed data fusion model

Figure 4.27: Comparison of estimated log rainfall fields for August 2019 (wet season) and January 2020 (dry season) between (a) stations-only model, (b) regression calibration model, and (c) proposed data fusion model. The figures show that the western section of the country has a pronounced dry and wet season.



(a) Stations-only model **(b)** Regression calibration model **(c)** Proposed data fusion model

Figure 4.28: Posterior uncertainty of the estimated log rainfall fields in Figure 4.27 for three approaches: (a) stations-only model, (b) regression calibration model, (c) proposed data fusion model. The posterior uncertainty in the estimated fields from the proposed data fusion model is the smallest.

The estimated log rainfall fields from the three approaches look similar (see Figure

4.27), but the uncertainty in the predictions from the proposed data fusion model is the smallest as expected, which are shown in Figure 4.28. The estimated spatial fields from the three modelling approaches for the same two months are shown in Figure A.7 of Appendix A, while the estimated error fields for the same two months from the proposed data fusion model are shown in Figure A.8 of Appendix A. Finally, Figure A.9 of Appendix A shows different scatterplots that indicate the correspondence between the observed and predicted values, although these are not as strong as the previous two meteorological variables. In particular, Figure A.9b of Appendix A shows the severe bias in the GSM values for rainfall.

4.6.4 Leave-group-out cross-validation

This section evaluates the predictive accuracy of the three modelling approaches using the leave-group-out cross-validation (LGOCV) approach (Adin et al., 2023; Liu and Rue, 2022). Contrary to the leave-one-out cross-validation method which estimates the predictive density for an observation at location $\mathbf{s}_i$ at time t by removing the same observation from the training set, the LGOCV approach computes the predictive densities by leaving out a set, say $I_{\mathbf{s}_i}$, of data points which includes the testing point and observations most related to it. The LGOCV is a better alternative than the leave-one-out cross-validation to evaluate prediction accuracy for structured models such as multi-level models, time series models, and spatial models (Adin et al., 2023; Liu and Rue, 2022) as it makes the unobserved data less dependent on the observed data, which is desirable when the goal of the prediction is extrapolation at unobserved locations. The LGOCV is efficiently implemented in the `INLA` library, with the details of the implementation found in Liu and Rue (2022).

Liu and Rue (2022) proposed two strategies to determine the leave-out sets $I_{\mathbf{s}_i}$: an automatic procedure based on the estimated correlation of the elements of the latent field, and a manual or user-defined approach. In this study, I use the latter strategy and implement the following: for each station $\mathbf{s}_i$, the leave-out set $I_{\mathbf{s}_i}$ for predicting $\mathrm{w}_1(\mathbf{s}_i, t)$ consists of the station at the spatial location $\mathbf{s}_i$ and all stations within its proximity (see Figure 4.29). In particular, four values are considered for the radius of the leave-out set: 60, 80, 125, and 150 km. Note that the testing point

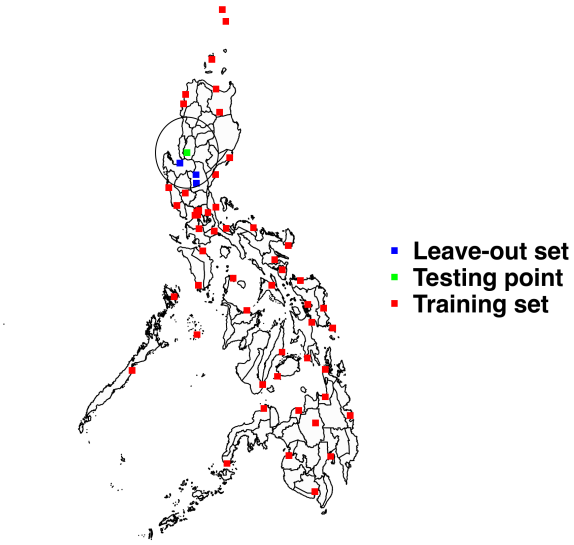


Figure 4.29: Illustration of the LGOCV approach. The model is fit on the training set (red) after excluding the leave-out set (blue and green), and then predictions are made on the testing point (green).

is also considered part of the leave-out set. Also, I remove the observed values for all time points for each station in $I_{\mathbf{s}_i}$.

In implementing the LGOCV, I use the approximation of the LGOCV predictive density for INLA models proposed in Liu and Rue (2022). The LGOCV predictive density for a station at location $\mathbf{s}_i$ and time t is given by $\pi\big(\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big)$. Note that the training set $\mathbf{Y}_{-I_{\mathbf{s}_i}}$ includes all data from the GSM and that only data from stations are left out. In particular, for this specific application, the approximation is done via the following nested integrals:

$$\pi\big(\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big) = \int_{\boldsymbol{\theta}} \pi\big(\mathrm{w}_1(\mathbf{s}_i,t)|\boldsymbol{\theta},\mathbf{Y}_{-I_{\mathbf{s}_i}}\big)\pi\big(\boldsymbol{\theta}|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big)d\boldsymbol{\theta}$$
$$\pi\big(\mathrm{w}_1(\mathbf{s}_i,t)|\boldsymbol{\theta},\mathbf{Y}_{-I_{\mathbf{s}_i}}\big) = \int \pi\big(\mathrm{w}_1(\mathbf{s}_i,t)|\eta(\mathbf{s}_i,t),\boldsymbol{\theta}\big)\pi\big(\eta(\mathbf{s}_i,t)|\boldsymbol{\theta},\mathbf{Y}_{-I_{\mathbf{s}_i}}\big)d\eta(\mathbf{s}_i,t),$$

where $\boldsymbol{\theta}$ denotes the model hyperparameters, while $\eta(\boldsymbol{s}_i,t) \equiv \mathbb{E}[\mathrm{w}_1(\mathbf{s}_i,t)] = x(\mathbf{s}_i,t)$. The details on how the nested integrals are approximated are detailed in Liu and Rue (2022).

I consider the mean of the predictive density $\pi\Big(\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\Big)$ as the predicted value at a testing point $\mathbf{s}_i$, where $\mathbf{Y}_{-I_{\mathbf{s}_i}}$ denotes the training set. Suppose N denotes the total number of data points from the stations. The following are the posterior prediction scores to compare the three modelling approaches:

1. LGOCV logarithmic utility (ULGOCV): $\dfrac{1}{N}\displaystyle\sum_{\forall i,t}\log\pi\Big(\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\Big)$

2. Root mean squared error (RMSE): $\sqrt{\frac{1}{N}\sum_{\forall i,t}\Big(\mathrm{w}_1(\mathbf{s}_i,t)-\mathbb{E}\big[\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big]\Big)^2}$

3. Mean absolute error (MAE): $\frac{1}{N}\sum_{\forall i,t}\big|\mathrm{w}_1(\mathbf{s}_i,t)-\mathbb{E}\big[\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big]\big|$

4. Mean absolute percentage error (MAPE): $\frac{1}{N}\sum_{\forall i,t}\left|\frac{\mathrm{w}_1(\mathbf{s}_i,t)-\mathbb{E}\big[\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big]}{\mathrm{w}_1(\mathbf{s}_i,t)}\right|$

5. Mean of the SD of predictive density of $x(\mathbf{s}_i,t)$ (MSD): $\frac{1}{N}\sum_{\forall i,t}\sqrt{\mathbb{V}\big[x(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big]}$

6. Mean Kullback-Leibler divergence (MKLD): $\frac{1}{N}\sum_{\forall i,t}D_{\mathrm{KL}}\Big(\pi\big(x(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\big)\big|\big|\pi\big(x(\mathbf{s}_i,t)|\mathbf{Y}\big)\Big)$

The LGOCV logarithmic utility (ULGOCV) is the mean of the log predictive densities $\pi\Big(\mathrm{w}_1(\mathbf{s}_i,t)|\mathbf{Y}_{-I_{\mathbf{s}_i}}\Big)$ which is related to the conditional predictive ordinate (Pettit, 1990). A higher value for the ULGOCV implies better model fit. Moreover, $D_{\mathrm{KL}}(\cdot||\cdot)$ denotes the Kullback-Leibler (KL) divergence metric, so that the MKLD is the mean of the KL divergence between the predictive density for $x(\mathbf{s}_i,t)$ given the complete data and the predictive density when excluding $I_{\mathbf{s}_i}$. A smaller value for the MKLD implies a better model fit.

Figure 4.30 shows a comparison of the posterior prediction scores for the temperature model. The proposed data fusion model generally has the highest ULGOCV especially when the leave-out set is large. The proposed model also has the smallest RMSE, MAE, MAPE, MKLD, and MSD. The prediction scores for the stations-only model and the regression calibration model deteriorate with the size of the leave-out set, while the scores for the proposed model are stable. The LGOCV results for the temperature model show that the proposed data fusion model outperforms the other two approaches, and that the stations-only model fares better than the regression calibration approach.

Similar results hold for relative humidity (Figure A.10 of Appendix A) and log rainfall (Figure A.11 of Appendix A). The benefits from doing data fusion is smaller for rainfall since there is no substantial difference in the scores, particularly for RMSE, MAPE, and MKLD, especially when the leave-out-sets are small. One potential reason for this is that the quality of the GSM outcomes for rainfall is lower compared to the other two meteorological variables, which is apparent from Figure 4.2 and from the model results, particularly with the estimated value of $\hat{\alpha}_1 < 1$ for the rainfall data fusion model. Nonetheless, the LGOCV results show that the proposed data fusion

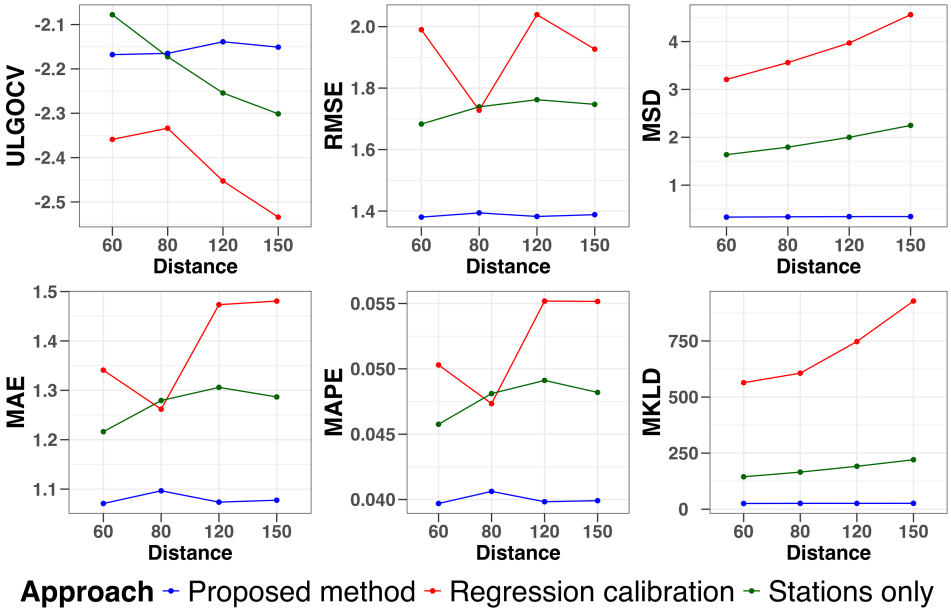


Figure 4.30: Comparison of LGOCV results for temperature from three models: stations-only model, regression calibration model, and the proposed data fusion model

model generally outperforms the other two approaches for the three meteorological variables considered.

4.7 Conclusions

Data fusion, which combines information from different data sources, has the potential benefit to improve model accuracy and prediction quality, while dealing with varying quality of the data sources (Bauer et al., 2015; Gettelman et al., 2022; Lawson et al., 2016). This chapter addresses a data fusion challenge motivated by meteorological data in the Philippines. The proposed model is based on the Bayesian melding model, which assumes a common latent process across data sources, and extends existing work in the literature (Forlani et al., 2020; Moraga et al., 2017; Villejo et al., 2023; Zhong and Moraga, 2023). In particular, I introduce a time-varying random field to model the additive bias in the numerical forecast model, termed *error field*, along with a constant multiplicative bias parameter. The goal of the proposed model is to perform spatial interpolation at a present time rather than make predictions for future time points.

The model offers several advantages: it defines a unified latent process for all data outcomes, accounts for measurement errors for all data sources, provides flexibility

in addressing biases, accommodates multiple spatially-misaligned data sources, and gauges the relative quality of the data sources. Although I assume that the multiplicative bias parameter α_1 is constant, the model can be extended to allow α_1 to vary over space or time, adding complexity but potentially improving accuracy in other applications. This extension is challenging because it involves estimating the product of two random fields. Furthermore, the proposed model treats the areal proxy data as point-referenced at the centroids of the grid cells. A (preliminary) simulation study was conducted to verify this assumption. The results show that this assumption is reasonable since the performance of the (full) Bayesian melding model and the simplified version are comparable in terms of the average squared error of the estimated field, and that the simplified version yields smaller posterior uncertainties.

A simulation study was performed to compare the proposed data fusion model to two benchmark models: a stations-only model and a regression calibration model. The main goal was to evaluate the model's performance under varying levels of data sparsity and prior specifications. The results showed that the proposed fusion model achieved lower squared errors and lower posterior uncertainty in the estimated fields, especially with sparse stations data. The proposed model also has lower Dawid-Sebastiani scores. In terms of the relative error and posterior uncertainty in the model parameter estimates, the proposed data fusion model also outperformed the two benchmark models.

In the data application, where three important meteorological variables are considered, the proposed model outperformed the stations-only model and the regression calibration model based on the leave-group-out cross-validation (LGOCV). The LGOCV results, calculated via the INLA method (Liu and Rue, 2022), showed that the proposed model provides better predictions with higher log predictive densities, smaller mean Kullback-Leibler divergence (KLD), and more accurate predictive scores (RMSE, MAE, MAPE, posterior uncertainty). Notably, the model's advantage grew with larger leave-out sets.

While this study considered only two meteorological data sources for the Philippines, the framework can be extended to include additional sources, such as satellite data or a second numerical forecast model. This would introduce new bias parame-

ters and error fields, broadening the model's utility. Future work could also involve incorporating longer time series data, which could enhance the model's capabilities.

The proposed data fusion framework, while developed for meteorological data, is applicable in other fields, such as air quality modelling (see the motivating example in Section 1.2.2 of Chapter 1). For example, in the UK, combining data from a network of monitoring stations called the Automatic Urban and Rural Network (AURN) (Lee et al., 2017), outcomes from a weather and chemical transport model called the Air Quality Unified Model (AQUM) (DEFRA, 2024; Forlani et al., 2020), and outcomes of dispersion models like the Pollution Climate Mapping (PCM) model which are run by Ricardo Energy & Environment (DEFRA, 2024; Forlani et al., 2020), could improve air quality predictions, which are crucial for public health. Unlike many existing models that consider one additional data source at a time, the proposed approach in this chapter allows for the joint use of multiple sources, while accounting for biases in each.

Another area where the proposed data fusion framework is applicable is in species distributions modelling. With technological advancements, data collection efforts to study the natural world have significantly increased, which are accompanied with a surge in data contributions from the general public, commonly referred to as citizen science data (August et al., 2015; Belmont et al., 2024). However, there are also concerns with regards to the quality of these data, particularly due to systemic biases (August et al., 2015; Koh and Opitz, 2023; Van Strien et al., 2013). These biases stem from uneven geographical coverage, differences in observer expertise, and variable effort in data collection. The appropriate use of citizen science data together with data from well-planned surveys falls within the realm of data fusion. The data fusion framework proposed in this chapter views these different data sources as realizations of the same latent process and, therefore, can be extended to the ecological context by assuming some probabilistic structure in the biases in the citizen science data while borrowing strength from the accuracy of the outcomes from well-planned surveys.

The INLA and the SPDE approach was used for model inference because they provide fast and reliable fitting of complex spatio-temporal models (Lindgren et al., 2011; Rue et al., 2009). A potential challenge in the computational aspect is that

the multiplicative bias parameter α_1 can be hard to identify and, therefore, can lead to numerical problems. To overcome these, a Bayesian model averaging approach is used, which allowed the fitting of the data fusion model conditional on fixed values of α_1. This is a viable approach, since there is an intuitive understanding of the plausible values of this bias parameter. Another approach for fitting the model is to include the parameter α_1 in the latent Gaussian field, and perform a linearization on the non-linear predictors using a first-order Taylor approximation, and to iteratively do this by looking for the optimal linearization point. This can be implemented using the `inlabru` library (Bachl et al., 2019; Lindgren et al., 2024; Serafini et al., 2023). Since the convergence of this approach and the properties of the approximation depend on the non-linear nature of the problem, it can also be computationally challenging for some cases. The use of a model averaging approach successfully removed the computational challenges, but the linearized INLA approach, previously described, is also a viable approach for the problem and will be further explored in a future work.

An immediate forthcoming work is to use the predicted fields from the data fusion model as input in an epidemiological model, in order to understand the link between climate data and health outcomes. This introduces another layer of spatial misalignment, since typically data for the health outcomes are areal while the predicted fields from the climate data fusion models are point-referenced. This is pursued in the next chapter, Chapter 5, which aims to link climate and dengue incidence in the Philippines. Moreover, in this two-stage modelling framework, accounting for the uncertainty in the data fusion model when fitting the health model should be carefully considered. The problem of uncertainty propagation has been studied in the context of health modelling (Blangiardo et al., 2016; Gryparis et al., 2009; Lee et al., 2017) and which generally falls under the area of measurement error models (Berry et al., 2002). This is formally discussed in Chapter 6.

Chapter 5

Linking climate and dengue in the Philippines

5.1 Introduction

Dengue fever is an infectious disease caused by the dengue arbovirus and commonly transmitted by two mosquito species: *Aedes aegypti* and *Aedes albopictus*. Section 1.2.1 explained the public importance of controlling dengue transmission. Relevant statistics on the global incidence and projections of the disease are also presented in Section 1.2.1.

This chapter focuses on the Philippines, a tropical country in Southeast Asia that has consistently been among the nations with the highest dengue incidence in the region (Undurraga et al., 2017, 2013). In the period 2008 – 2012, the country's Health Department reported an annual average of 117,065 dengue cases, and a fatality rate of 0.55% (Edillo et al., 2015). The last dengue epidemic in the country occurred in 2019, with 437,563 recorded cases, the highest number ever recorded worldwide (Ong et al., 2022).

Dengue virus spend part of their life cycle in the external environment; thus, disease transmission is particularly influenced by climatic factors. Section 5.3 presents a literature review on the association between dengue and climate variables, particularly temperature, rainfall, and relative humidity.

Studies relating dengue to climate factors in the Philippines have mainly employed relatively simple statistical methodology and limited the analysis to a subarea rather than the entire country, relying on rather sparse covariate data measured at a small number of weather stations. These analyses include correlation analysis or classical linear models such as MANOVA (Dulay et al., 2013; Duque-Lee et al., 2020; Edillo et al., 2022, 2024; Marigmen and Addawe, 2022a,b; Murphy et al., 2022; Su, 2008), generalized additive modelling (Carvajal et al., 2018; Cawiding et al., 2025; Cruz et al., 2024), deterministic climate-dengue risk functions (Xu et al., 2020), generalized linear models like a Poisson regression (Francisco et al., 2021; Iguchi et al., 2018), classical time series approaches such as ARIMA models (Pineda-Cortel et al., 2019), or spectral analysis methods, which also focus on the temporal and/or seasonal variations of the relationship between climate and dengue (Francisco et al., 2021; Subido and Aniversario, 2022; Sumi et al., 2017), or machine learning algorithms (Buczak et al., 2014; Carvajal et al., 2018). Seposo et al. (2023) employed a mixed modelling framework, but they only considered unstructured random intercepts in space and time.

This chapter aims to contribute to the literature on the association between climate variables and dengue disease. The specific goal is to understand the covariate effect, rather than to predict future outbreaks or to assess the impacts of climate change on dengue incidence. The novelty of this chapter lies in the use of a complex two-stage Bayesian spatio-temporal model that incorporates both structured and unstructured random effects across space and time, including their interactions. The models account for the complex spatial structure of the Philippines, which forms an archipelago consisting of more than 7000 islands, while employing a data-fusion approach making use of climate data from weather stations and a weather prediction model.

To assess the association between climate and dengue, I use a two-stage model in a Bayesian framework (Figure 1.6 in Chapter 1). The first stage fits climate models, and produces predicted surfaces of the climate variables. The second-stage model is the health model, where dengue incidence is the outcome variable, and the climate predictions from the first-stage model are the primary covariates of interest. As

discussed in Section 1.3, a two-stage modeling framework is typically used in spatial analysis, particularly in cases where response variable and predictors are spatially misaligned, and is common in spatial epidemiology (Blangiardo et al., 2016; Cameletti et al., 2019; Gryparis et al., 2009; Lee et al., 2017; Liu et al., 2017; Szpiro et al., 2011). In this work, dengue incidence (the response) is areal, while the climate variables are point-referenced, resulting in spatial misalignment.

Section 1.3 in Chapter 1 presents a justification for the use of a two-stage modelling framework, rather than a joint modelling approach. To emphasize an important point, a two-stage modelling framework is appropriate and reasonable in this specific context for the following reasons: (a) it is computationally efficient, (b) it offers an intuitive physical interpretation, (i.e. climate affects dengue but not the other way around) and (c) it avoids potential feedback effects. In this chapter, I use the posterior sampling approach (Blangiardo et al., 2016; Cameletti et al., 2019; Liu et al., 2017; Villejo et al., 2023; Zhu et al., 2003) to account for the uncertainty in the climate predictions.

In most studies examining the relationship between dengue and climate, including those focusing on the Philippines, data from meteorological stations are the primary source of climate data. The measurements taken at these stations are considered the gold standard due to their high accuracy. However, weather station networks are typically sparse, as is the case in the Philippines (see Figure 1.3a or Figure 4.1). Models trained on sparse data often produce predictions with high uncertainty and potential biases (Lawson et al., 2016). To mitigate these challenges, I use the data fusion models developed in Chapter 4, which were based on two climate data sources: weather synoptic stations and the GSM forecast model.

For inference, I use integrated nested Laplace approximation (INLA), as it provides fast and accurate posterior estimates (Rue et al., 2009). In addition, I use the stochastic partial differential equations (SPDE) approach (Lindgren et al., 2011) to represent spatial Gaussian Matérn fields in the climate models. The combination of INLA and the SPDE approach has proven to be a powerful tool for spatial or spatio-temporal analysis (Bakka et al., 2018; Blangiardo and Cameletti, 2015; Cameletti et al., 2013; Lindgren and Rue, 2015; Schrödle and Held, 2011).

This chapter is structured as follows: Section 5.2 presents the data sources and an initial data exploration. Section 5.3 presents a literature review on the link between climate and dengue. The proposed models are discussed in Section 5.4. Estimation strategies are presented in 5.5, which highlights the approach for uncertainty propagation in Section 5.5.2. Results and discussion are presented in Section 5.6, followed by conclusions and future work in Section 5.7.

5.2 Data

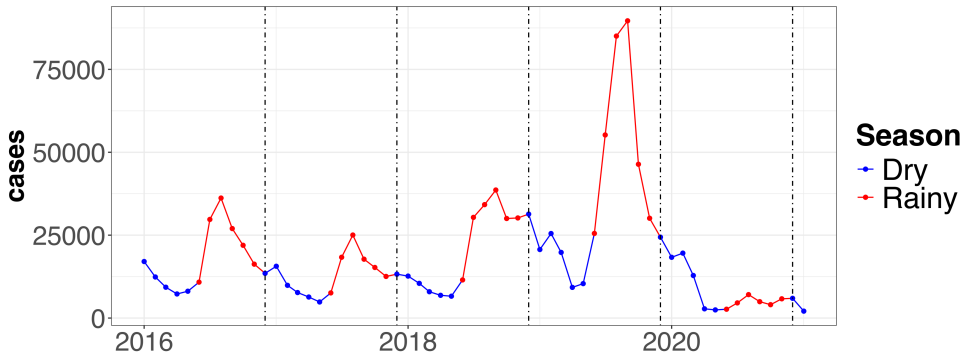


Figure 5.1: Time series plot of the number of dengue cases in the Philippines from January 2016 to January 2021

Data on dengue cases were retrieved from the United Nations Office for the Coordination of Humanitarian Affairs, whose primary mandate is to support humanitarian organizations worldwide through data collection, information dissemination, funding and resource mobilization, development of policies to meet the needs of crisis-affected people, and implementation of campaigns that advocate for humanitarian action. The data consist of weekly dengue case counts at the provincial level from 2016 to 2020. Data prior to 2016 were excluded due to differences in surveillance and reporting practices. Specifically, by 2015, the country's Health Department mandated nationwide reporting to include probable cases, whereas previously only clinically-confirmed cases were recorded in selected sentinel sites (Seposo, 2021). For the analysis, I aggregate the data to monthly level. Figure 5.1 shows a plot of monthly dengue cases in the Philippines from January 2016 to January 2021. As noted in Section 1.2.1 in Chapter 1, seasonality is evident, with cases generally higher during the rainy season (June to November). Notably, there is an unusually high number of cases from August to October 2019, which was a period of national dengue alert and epidemic (BBC,

2019). Moreover, during 2020, coinciding with the onset of the COVID-19 pandemic, the number of reported cases is very low. This global phenomenon (WHO, 2023a) was attributed to both reduced mobility – several studies have shown that limited household movement is linked to lower transmission (Stoddard et al., 2013) – and reporting hesitancy, as individuals were afraid of contracting COVID-19 when visiting health facilities (Seposo, 2021).

Figure 1.2 in Chapter 1 shows a plot of the annual (2016 to 2020) total number of dengue cases in the country. As already noted, 2019 had the highest number of cases, and it also shows specific areas with the highest number of recorded cases. The areas with high cases are the same areas identified by the Health Department of the country as requiring immediate emergency attention (BBC, 2019).

In epidemiological applications, a measure of risk, which accounts for the differences in the sizes and demographic structure of the provinces or areas, is typically mapped (Waller and Carlin, 2010; Waller and Gotway, 2004). A common measure of risk is the standardized incidence ratio (SIR), and is computed as the ratio of observed and expected cases. The expected cases are typically computed using indirect standardization, which involves applying age-specific rates (incidence proportions) of a standard population on the study population (Waller and Gotway, 2004). The age-specific national rates can be used, which are then applied on each age stratum for each province. However, a limitation of the data is that there are no information on age-specific rates. Hence, I used internal standardization to compute the expected cases. This approach uses an estimate for the baseline individual risk based on the national observed disease rate (Waller and Carlin, 2010). This is further detailed in Section 5.4.1. Figure 5.2 shows a plot of dengue SIRs for specific months: August 2019 to November 2019. The reason for choosing these specific months is that they correspond to the period of high incidence of dengue, based on Figure 5.1. The plot shows, specifically for August 2019, that the area with the highest number of reported cases (see Figure 1.2 in Chapter 1) is also the area with the highest SIR.

For climate variables, I used the same data from Chapter 4: a sparse network of weather stations, and outcomes of GSM. For the main results in this chapter, I used existing climate predictions from the data fusion models in Chapter 4, since

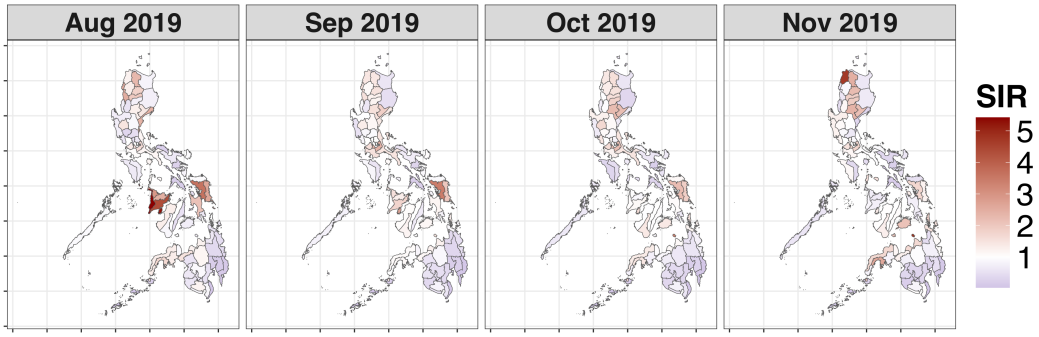


Figure 5.2: Plot of standardized incidence ratios (SIR) of dengue in the Philippines from August 2019 to November 2019

predicted fields based on sparse stations data either have high uncertainty or more biased (Lawson et al., 2016). However, the data fusion models cover only years 2019 to 2020, due to data availability constraints. Thus, I also explored the use of longer monthly time series, albeit for stations only, from 2016 to 2020, for the same climate variables: temperature (in °C), relative humidity (in %), and total rainfall (in mm). The results from these models are presented in Section B.2 of Appendix B.

5.3 Climate and dengue

Temperature affects the extrinsic incubation period of the pathogen, as well as the reproductive rate and biting rate of the mosquito (Ewing et al., 2016; Promprou et al., 2005). Higher temperatures accelerate mosquito breeding and shorten the virus incubation period, allowing mosquitoes to become infectious more quickly (Macdonald. G, 1957). Furthermore, increasing temperatures enhance the mosquito's biting behavior; raising the risk of virus transmission. However, there is also a temperature threshold beyond which mosquito survival decreases. The temperature range shown to be optimal for dengue transmission is 21.3–34.0°C for *Ae. aegypti* and 19.9–29.4°C for *Ae. albopictus* (Ryan et al., 2019). Rising global temperatures are expected to increase the risk of mosquito-borne diseases, particularly dengue and malaria. Under worst climate-change scenarios, a 1°C increase in global mean temperature could put 2.4 billion people at risk of both diseases by 2100 (Colón-González et al., 2021), with the higher risk concentrated in densely populated areas of Africa, Southeast Asia and the Americas. In the Philippines, temperature has been shown to have a positive

effect on dengue incidence (Cawiding et al., 2025; Francisco et al., 2021; Subido and Aniversario, 2022). In Seposo et al. (2024), they showed that between 2010 and 2019, 72.1% of reported dengue cases in the Philippines were attributable to temperature, which implies that it is a significant driver of dengue transmission. Edillo et al. (2022) looked at the effect of temperature at the vector level, particularly looking at three development-related phenotypes: percent pharate larvae, hatch rates, and reproductive outputs. Their results show that temperature, together with season and latitudinal differences of the islands, significantly influence the phenotypes of *Aedes aegypti*. The latitude of a spatial location brings variation in the amount of sunlight received, which affects the suitability of the breeding sites of mosquitoes. Lastly, Xu et al. (2020) showed a non-linear association between temperature and dengue.

Increasing rainfall creates more breeding sites for mosquitoes, leading to an increase in the mosquito population and a higher risk of virus transmission (Ewing et al., 2016; Promprou et al., 2005). However, excessive rainfall can wash away breeding sites, decreasing the risk of dengue. In consistently wet regions, a decrease in rainfall, such as during droughts, can cause water stagnation in rivers and lead to increased water storage, both of which create ideal breeding conditions for *Aedes* mosquitoes (McMichael, 2003). In the Philippines, rainfall has also been shown to have a positive relationship with dengue, with some lagged effect (Francisco et al., 2021). Cawiding et al. (2025) showed that the effect of rainfall could vary depending on the location. In western areas of the country, which experience pronounced dry and wet season, sporadic rainfall could create new breeding sites, thus increasing dengue incidence. However, for areas in the eastern part of the country, with a uniform amount (low variation) of rainfall, rainfall tends to flush out stagnant water, reducing mosquito breeding sites.

Since relative humidity is positively related to rainfall (MetOffice, 2024), it is also linked to dengue transmission. Virus transmission tends to be higher during months of high humidity (McMichael, 2003), as increased humidity favours mosquito survival (Gubler et al., 2001). In contrast, mosquitoes dessicate easily under dry conditions (Focks et al., 1995; Hales et al., 2002). Xu et al. (2020) showed that relative humidity is the only factor to be associated with a future seasonal peak of

dengue in the Philippines.

Other important factors are house structure, human behaviour and general socioeconomic conditions (Patz et al., 2000). Dengue transmission is further exacerbated by ineffective vector and disease surveillance, inadequate public health infrastructure, population growth, unplanned and uncontrolled urbanization, and increased travel (ECDC, 2024; Gubler et al., 2014; McMichael, 2003; Murphy and Nathanson, 1994; Rigau-Pérez et al., 1998).

5.4 Proposed model

This section mainly presents the proposed second-stage (dengue) models. The first-stage (climate) data fusion models are discussed in Chapter 4. Section 5.4.1 presents the health model. This section also discusses the formula for computing block-averages of climate field (Section 5.4.2), specification of spatial and temporal effects (Section 5.4.3), instrinsic conditional autoregressive specification for disconnected graphs (Section 5.4.4), models for interaction effects (Section 5.4.5), and prior specification for second-stage model parameters (Section 5.4.6). The results from a stations-only model input, which also covers a longer time series (2016 to 2020), are provided in Section B.2 of Appendix B.

5.4.1 Poisson model for dengue

Let $\mathrm{y}(B_i, t)$ be the number of observed cases in area B_i, $i = 1, \ldots, N$, and time t, $t = 1, \ldots, T$. The observed cases are assumed to be Poisson distributed with mean $\mu(B_i, t)$. The model is given by:

$$\begin{aligned} \mathrm{y}(B_i, t) &\sim \text{Poisson}\Big(\mu(B_i, t)\Big) \\ \mathbb{E}\Big[\mathrm{y}(B_i, t)\Big] &= \mu(B_i, t) = \lambda(B_i, t) \times \mathrm{E}(B_i, t) \\ \log\Big(\lambda(B_i, t)\Big) &= \gamma_0 + \gamma_1 \hat{x}(B_i, t) + \boldsymbol{\gamma}_2^{\intercal} \mathbf{z}_2 + \varphi(B_i, t), \end{aligned} \tag{5.1}$$

where $\{\gamma_0, \gamma_1, \boldsymbol{\gamma}_2\}$ are fixed effects, $\mathbf{z}_2$ is a set of covariates, $\hat{x}(B_i, t)$ is the block-level value for a climate variable at time t, and $\varphi(B_i, t)$ is a spatio-temporal random

effect. In Equation (5.1), the Poisson mean is expressed as a product of $\lambda(B_i, t)$ and $\mathrm{E}(B_i, t)$, where $\lambda(B, t)$ is the risk while $\mathrm{E}(B_i, t)$ is the expected cases. $\mathrm{E}(B_i, t)$ are known quantities which are computed using internal standardization (Waller and Carlin, 2010). In particular, suppose that $\mathrm{n}(B_i, t)$ is the population at risk at area B_i and time t. Moreover, suppose that r_t is the constant baseline risk per person at time t, which is estimated using aggregate population data, i.e., $\hat{r}_t = \dfrac{\sum_{\forall B_i} \mathrm{y}(B_i, t)}{\sum_{\forall B_i} \mathrm{n}(B_i, t)}$, and is interpreted as the global observed disease rate for time t. The expected number of cases in B_i at time t is then computed as $\mathrm{E}(B_i, t) = \mathrm{n}(B_i, t) \times \hat{r}_t$. Since we have $\log\Big(\lambda(B_i, t)\Big) = \log\left(\dfrac{\mu(B_i, t)}{\mathrm{E}(B_i, t)}\right)$, this implies that $\log\Big(\mu(B_i, t)\Big) = \log\Big(\lambda(B_i, t)\Big) + \log\Big(\mathrm{E}(B_i, t)\Big)$ i.e., $\log\Big(\mathrm{E}(B_i, t)\Big)$ operates as an offset parameter in the Poisson model.

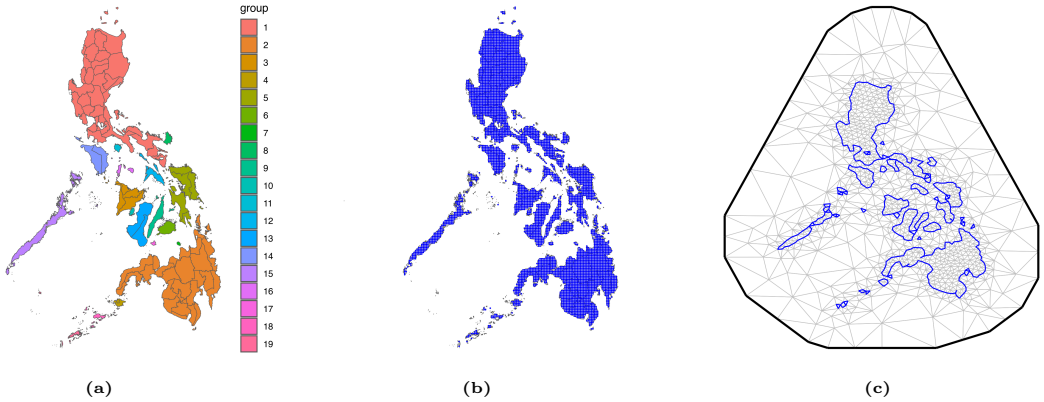


Figure 5.3: (a) Plot showing the 19 disconnected graphs for the iCAR model. Out of the 19 graphs, 12 of them are singletons (isolated islands), (b) prediction grid, (c) mesh used for the SPDE approximation

5.4.2 Block averages of climate variables

Since the climate model is point-referenced, while the health model is areal, the two models are spatially misaligned. To address this, I use the definition of the average level of $x(\mathbf{s}, t)$ for an area B_i at time t in Gelfand et al. (2010), denoted by $x(B_i, t)$, as follows:

$$x(B_i, t) = \int_{B_i} x(\mathbf{s}, t) p(\mathbf{s}) d\mathbf{s}, \tag{5.2}$$

for a weighting function $p(\mathbf{s})$ such that $\int_{B_i} p(\mathbf{s}) d\mathbf{s} = 1$ (see also Equation (2.1) of Chapter 2). Cameletti et al. (2019) mentioned two ways to approximate Equation (5.2) (see also Section 3.2.2). The first one is a linear combination based on neighbourhood intersections, while the second one is an unweighted (simple) mean of predicted

values of $x(\mathbf{s}, t)$ that lie inside B_i. In this work, I used the latter, so that

$$\hat{x}(B_i, t) = \sum_{\forall \mathbf{s}_i^* \in B_i} \hat{x}(\mathbf{s}_i^*, t) p(\mathbf{s}_i^*) = \frac{1}{\#B_i} \sum_{\forall \mathbf{s}_i^* \in B_i} \hat{x}(\mathbf{s}_i^*, t), \tag{5.3}$$

where $\#B_i$ denotes the number of prediction points in block B_i and $\hat{x}(\mathbf{s}_i^*, t)$ is the predicted value of the climate variable at spatial location $\mathbf{s}_i^*$ and time t. Equation (5.3) is computed using a fine prediction grid shown in Figure 5.3b.

5.4.3 Spatio-temporal effects $\varphi(B_i, t)$

I assume the following form for the spatio-temporal effects:

$$\begin{aligned}
\varphi(B_i, t) &= \psi(B_i) + \zeta(t) + \nu(t) + \upsilon(B_i, t) \\
\psi(B_i) &= \left[\frac{1}{\sqrt{\tau_\psi}} \sqrt{1-\phi}\mathrm{v}(B_i) + \sqrt{\phi}\mathrm{u}(B_i) \right] \\
\mathrm{v}(B_i) &\sim \mathcal{N}(0, 1) \\
\mathrm{u}(B_i) &\sim \text{scaled iCAR on a disconnected graph} \\
\zeta(t) &\overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\zeta^2) \\
\nu(t) &\text{ is a random walk in time of order 2} \\
\upsilon(B_i, t) &\text{ is in interaction term between space and time}
\end{aligned} \tag{5.4}$$

The model specification for $\psi(B_i)$ follows that of Riebler et al. (2016). It provides a compromise between pure overdispersion, denoted by $\mathrm{v}(B_i)$, and spatially-structured correlation, denoted by $\mathrm{u}(B_i)$. The total (marginal) variance of the spatial main effect $\psi(B_i)$ is $1/\tau_\psi$, and the proportion of the variance explained by the structured effect is ϕ. Thus, ϕ is a mixing parameter between the unstructured and structured spatial effect. The structured effect $\mathrm{u}(B_i)$ is defined as an intrinsic conditionally autoregressive (iCAR) process (Besag et al., 1991) (see also Section 2.3). In particular, since the Philippines is an archipelago, which means that there are provinces without neighbours (islands), the iCAR process is defined on a disconnected graph; the details of which are provided in Section 5.4.4. Similarly, Equation (5.4) also specifies an unstructured and structured effect in time, denoted by $\zeta(t)$ and $\nu(t)$, respectively. I assume a random walk of order 2 for the structured time effect. Finally, $\upsilon(B_i, t)$

specifies the interaction term between space and time (see Section 5.4.5 for details).

5.4.4 Specifying the iCAR process on a disconnected graph

Under the assumption of an iCAR model for $\mathrm{u}(B_i)$, the joint distribution of the effects $\mathbf{u} = \Big(\mathrm{u}(B_1), \ldots, \mathrm{u}(B_\mathrm{N})\Big)^\intercal$ is then given by:

$$\pi(\mathbf{u}|\sigma_\mathrm{u}^2) = \left(\frac{1}{2\pi\sigma_\mathrm{u}^2}\right)^{(\mathrm{N}-1)/2} |\mathbf{R}_\mathrm{u}|_*^{1/2} \exp\left(-\frac{1}{2\sigma_\mathrm{u}^2}\sum_{i\sim j}\Big(\mathrm{u}(B_i) - \mathrm{u}(B_j)\Big)^2\right), \tag{5.5}$$

where σ_u^2 is the marginal variance, $i \sim j$ means that $\mathrm{u}(B_i)$ and $\mathrm{u}(B_j)$ are neighbours, $|\cdot|_*$ is the generalized determinant, and $\mathbf{R}_\mathrm{u}$ is the structure matrix representing the neighborhood structure, whose ij^{th} element, denoted by R_{ij}, is given by:

$$R_{ij} = \begin{cases} \mathrm{n}_i & i = j \\ -1 & i \sim j \\ 0 & \text{otherwise} \end{cases}, \tag{5.6}$$

where n_i denotes the number of neighbours of area B_i. It should be noted that an iCAR process is defined with respect to a specific undirected graph, a set of vertices (which here refer to the areas $\{B_1, \ldots, B_\mathrm{N}\}$), and the edges (which refer to the set of neighbours).

The problem in the data application is that there are areas or provinces which do not have neighbours because they are islands. In particular, the entire archipelago can be viewed as a collection of disconnected graphs, where some of them are singletons. In the data, there are 19 disconnected graphs, where 12 of them are singletons. These are shown in Figure 5.3a. When working on disconnected graphs, the precision parameters for each connected graph are not comparable, and the presence of singletons can lead to improper posterior distribution (Freni-Sterrantino et al., 2018).

To circumvent the aforementioned issues, I used the proposed method in Freni-Sterrantino et al. (2018), which is based on the proposed scaling of intrinsic GMRFs in Sørbye and Rue (2014). The first step is to define an iCAR model for each connected graph (not including the singletons). In the data application, there are 7 connected

graphs. Suppose we denote by $\mathbf{R}_j$ the structure matrix for the j^{th} connected graph. Moreover, suppose $s_{ij}^{(\kappa)}$ are the marginal variances, i.e., the diagonal elements of the generalized inverse of $\kappa\mathbf{R}_j$. The geometric mean of the marginal variances are then computed, given by:

$$\mathcal{S}_j^{(\kappa)} = \exp\left(\frac{1}{\mathrm{n}_j}\sum_i \log\left(s_{ij}^{(\kappa)}\right)\right). \tag{5.7}$$

Equation 5.7 is computed conditional on κ, where n_j refers to the number of nodes in the j^{th} connected graph. The scaled precision matrix of the j^{th} connected graph is then given by

$$\tau\mathcal{S}_j^{(1)}\mathbf{R}_j.$$

The parameter τ is now interpreted as the common precision parameter for all the connected graphs. For singletons, a standard Gaussian with precision τ is assumed, i.e., singletons has a non-spatial random effect. This gives the following scaled precision matrix, denoted by $\mathbf{Q}_{\mathrm{u}}$, for the iCAR with disconnected graph in the data application:

$$\mathbf{Q}_{\mathrm{u}} = \tau\mathcal{S}_1^{(1)}\begin{pmatrix}\mathbf{R}_1 & & & & \\ & \ddots & & & \\ & & \mathbf{0} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix} + \cdots + \tau\mathcal{S}_7^{(1)}\begin{pmatrix}\mathbf{0} & & & & \\ & \ddots & & & \\ & & \mathbf{R}_7 & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 0\end{pmatrix} + \cdots + \tau\begin{pmatrix}\mathbf{0} & & & & \\ & \ddots & & & \\ & & \mathbf{0} & & \\ & & & 0 & \\ & & & & \ddots \\ & & & & & 1\end{pmatrix} \tag{5.8}$$

This specification means that if a block B_i has a neighbour, then $\mathrm{u}(B_i)$ shrinks to the local mean of the graph where it belongs. On the other hand, if a block B_i does not have a neighbour, then $\mathrm{u}(B_i)$ shrinks to the overall or global mean. Also, τ operates as the precision parameter for all connected components, i.e., it regulates the degree to which each $\mathrm{u}(B_i)$ shrinks to either the local mean or the global mean.

In addition to applying a sum to zero constraint, I also added a separate intercept for each connected component, as recommended by Freni-Sterrantino et al. (2018). This implies that random effects for each node within a connected component can deviate randomly from the local intercept, just as singletons deviate from the global

(overall) intercept. The size of the deviation only depends on τ and not on the graph structures of the connected components. Neither the sum-to-zero constraint nor a separate intercept is required for singletons.

5.4.5 Interaction term $\upsilon(B_i, t)$

I explored four possible specifications for the interaction term $\upsilon(B, t)$ following Knorr-Held (2000). These are summarized below:

1. Type I interaction assumes that both unstructured effects in space and time interact, i.e., $\upsilon(B_i, t) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma_\upsilon^2\big)$. The structure matrix for this interaction is simply the identity matrix, i.e., $\mathbf{R}_\upsilon = \mathbb{I}$, so that $\boldsymbol{\upsilon} \sim \mathcal{N}\big(\mathbf{0}, \sigma_\upsilon^2 \mathbb{I}\big)$.

2. Type II interaction assumes that a structured temporal effect interacts with an unstructured spatial effect. The structure matrix here is the Kronecker product of the structure matrices for the iid effect in space and the first-order autoregressive effect in time. This type of interaction implies that each block B_i has its own temporal structure, which is independent of the other blocks/areas.

3. Type III interaction assumes an interaction between an unstructured temporal effect and a structured spatial effect. The structure matrix is the Kronecker product between the iid effect in time and the iCAR effect in space. This type of interaction implies that for each time point t, there is a spatial structure, which is independent of the other time points.

4. Type IV interaction assumes that both structured effects in space and time interact. This type of interaction specifies a temporal (autoregressive) structure for each area which depends on the temporal patterns of the neighbouring areas.

5.4.6 Specification of priors

I assigned non-informative priors for the fixed effects, particularly, $\gamma_0 \sim \mathcal{N}(0, \infty)$, $\gamma_1 \sim \mathcal{N}(0, 1000)$, and $\boldsymbol{\gamma}_2 \sim \mathcal{N}(\mathbf{0}, 1000\mathbb{I})$. The precision parameters of the time effects are given the following vague priors: $\log\left(1/\sigma_\nu^2\right) \sim \log \text{Gamma}(1, 0.00005)$ and $\log\left(1/\sigma_\zeta^2\right) \sim \log \text{Gamma}(1, 0.00005)$. The variance parameter for the spatial effect

$\psi(B_i)$ is given a PC prior, particularly, $\mathbb{P}\Big(1/\sqrt{\tau_\psi} > 1\Big) = 0.01$. This prior choice assumes a probability of 0.99 of having residual relative risks smaller than 1. The mixing parameter ϕ of the spatial effect $\psi(B_i)$ is also given a PC prior, particularly, $\mathbb{P}\Big(\phi < 0.5\Big) = 2/3$. This gives more mass to values $\phi < 0.5$. This is more conservative since it assumes that the unstructured term accounts for more of the variation in $\psi(B_i)$ (Riebler et al., 2016). The parameters for the spatio-temporal interaction effect $\upsilon(B_i, t)$ are given non-informative priors. For the Type I interaction, I specified $\log\Big(1/\sigma_\upsilon^2\Big) \sim \log \text{Gamma}(1, 0.00005)$. For the Type II interaction, the variance parameter has the same prior as Type I, but in addition, I specified a log normal prior for the transformed AR parameter, similar to the prior choice for ϕ_1 of the first-stage model. For the Type III interaction, both the precision parameters of the unstructured time effect and structured time effects are also given the $\log \text{Gamma}(1, 0.00005)$ distribution. Finally, for the Type IV interaction, the precision parameters and the AR parameter are also given similar priors as the other interaction types. The prior distributions for the first-stage model parameters are discussed in Chapter 4.

5.5 Model Estimation

5.5.1 INLA and SPDE approach

I used the integrated nested Laplace approximation (INLA) approach (Rue et al., 2009) for performing inference. Moreover, I also used the stochastic partial differential equations (SPDE) approach, which is an efficient method for estimating Gaussian fields of the Matérn class (Lindgren et al., 2011). Both are discussed in Sections 2.5.2 and 2.6 of Chapter 2, respectively.

5.5.2 Uncertainty propagation

A crude two-stage modelling approach is to first evaluate Equation (5.3) using the posterior mean of $x(\mathbf{s}, t)$ at the prediction grid spatial locations, and then plug-in the block-averages into the second stage model; the formal steps for this are given in Algorithm 5.1. However, this will potentially underestimate the posterior uncertainty

in the second-stage model parameters, since the uncertainty in $\hat{x}(\mathbf{s}, t)$ is ignored. In order to account for the uncertainty, we use a resampling approach, which essentially generates several samples from the posterior distribution of first-stage parameters, and then uses each resample as plug-in values to the second-stage model; the formal steps are outlined in Algorithm 5.2.

Algorithm 5.1 Implementation of the crude plug-in method

Step 1: Suppose $\hat{\beta}_0$, $\hat{\boldsymbol{\beta}}_1$ and $\{\hat{\boldsymbol{\xi}}_t\}_{t=1,\ldots,T}$ are posterior means from the first-stage model (see Section 4.3.2 of Chapter 4). Also, suppose that $\mathbf{B}_{\text{grid}}$ is the projection matrix from the SPDE mesh nodes (Figure 5.3c) to the point locations of the prediction grid (Figure 5.3b). Compute the following:

$$\hat{x}(\mathbf{s}^*, t) = \hat{\beta}_0 + \hat{\boldsymbol{\beta}}_1^\intercal \mathbf{z}(\mathbf{s}^*, t) + \mathbf{b}_{\mathbf{s}^*}^\intercal \hat{\boldsymbol{\xi}}_t, \tag{5.9}$$

where $\mathbf{b}_{\mathbf{s}^*}$ is the row in $\mathbf{B}_{\text{grid}}$ which corresponds to the spatial location $\mathbf{s}^*$, and $\mathbf{z}(\mathbf{s}^*, t)$ is the set of covariates for the first-stage model (see Equation (4.5) of Chapter 4).

Step 2: Compute the spatial average given in Equation (5.3) for all B_i and t.

Step 3: Plug in the values of $\hat{x}(B_i, t), \forall B_i, t$, in the second-stage model.

When the first-stage model uses data solely from stations, the crude plug-in method and resampling method are straightforwardly implemented using Algorithms 5.1 and 5.2, respectively. However, there is a slight modification in the steps when the data fusion model is used as the first-stage model. The reason is that the data fusion model is estimated conditional on the multiplicative bias parameter α_1 (see Equation (4.7) in Chapter 4), which in practice is defined on a grid of values centered on 1.

Suppose that $\hat{\beta}_0^{(\ell)}$, $\hat{\boldsymbol{\beta}}_1^{(\ell)}$ and $\{\hat{\boldsymbol{\xi}}_t^{(\ell)}\}_{t=1,\ldots,T}$ denote the posterior means for the (first-stage) model which was fitted conditional on $\alpha_1 = \alpha_1^{(\ell)}$. Here, the first-stage model is indexed by $\ell = 1, \ldots, L$, where L is the number of α_1 values considered. In addition, suppose that w_ℓ denotes the weight for the ℓ^{th} model, i.e., the model fitted conditional on $\alpha_1 = \alpha_1^{(\ell)}$. The spatial averages are then computed as:

$$\hat{x}(B_i, t) = \frac{1}{L}\frac{1}{\#B_i}\sum_{\ell=1}^{L}\sum_{\forall \mathbf{s}^* \in B_i} \hat{x}^{(\ell)}(\mathbf{s}^*, t) w_\ell = \frac{1}{L}\frac{1}{\#B_i}\sum_{\ell=1}^{L}\sum_{\forall \mathbf{s}^* \in B_i} \left(\hat{\beta}_0^{(\ell)} + \hat{\boldsymbol{\beta}}_1^{(\ell)\intercal}\mathbf{z}(\mathbf{s}^*, t) + \mathbf{b}_{\mathbf{s}^*}^\intercal \hat{\boldsymbol{\xi}}_t^{(\ell)}\right) w_\ell. \tag{5.12}$$

Equation (5.12) implies that the spatial average for block B_i and time t is a weighted

Algorithm 5.2 Implementation of the resampling method

Repeat steps 1–5 for $j = 1, 2, \ldots, J$:

Step 1: Generate posterior samples of the latent parameters from the first-stage model (see Section 4.3.2 of Chapter 4).

Step 2: Suppose $\tilde{\beta}_0^{(j)}$, $\tilde{\boldsymbol{\beta}}_1^{(j)}$ and $\{\tilde{\boldsymbol{\xi}}_t^{(j)}\}_{t=1,\ldots,T}$ are the j^{th} posterior samples from the first-stage model. Compute the following:

$$\tilde{x}^{(j)}(\mathbf{s}^*, t) = \tilde{\beta}_0^{(j)} + \tilde{\boldsymbol{\beta}}_1^{(j)\intercal}\mathbf{z}(\mathbf{s}^*, t) + \mathbf{b}_{\mathbf{s}^*}^{\intercal}\tilde{\boldsymbol{\xi}}_t^{(j)}, \tag{5.10}$$

where $\mathbf{b}_{\mathbf{s}^*}$ and $\mathbf{z}(\mathbf{s}^*, t)$ are the same as in Equation (5.9).

Step 3: Compute the spatial average given in Equation (5.3) for all B_i and t. Here, I denote the spatial averages by $\tilde{x}^{(j)}(B_i, t)$, since these are computed using posterior samples of the latent parameters, not the posterior means.

Step 4: Plug in the values of $\tilde{x}^{(j)}(B_i, t), \forall B_i, t$, in the second-stage model.

Step 5: Store all the posterior estimates, such as $\pi^{(j)}(\gamma_0|\mathbf{y})$ and $\pi^{(j)}(\gamma_1|\mathbf{y})$, with the superscript (j) denoting that these posteriors are computed using the j^{th} posterior samples.

Step 6: Combine all results via model averaging, e.g.,

$$\pi(\gamma_0|\mathbf{y}) = \frac{1}{J}\sum_j^J \pi^{(j)}(\gamma_0|\mathbf{y}) \quad \text{and} \quad \pi(\gamma_1|\mathbf{y}) = \frac{1}{J}\sum_j^J \pi^{(j)}(\gamma_1|\mathbf{y}). \tag{5.11}$$

average of the point predictions from each first-stage model estimated conditional on $\alpha_1 = \alpha_1^{(\ell)}, \ell = 1, \ldots, L$.

A similar idea is implemented with the resampling method, but the spatial averages are computed by further averaging all results across the J posterior samples, i.e.,

$$\hat{x}(B_i, t) = \frac{1}{J}\frac{1}{L}\frac{1}{\#B_i}\sum_{j=1}^{J}\sum_{\ell=1}^{L}\sum_{\forall \mathbf{s}^* \in B_i}\left(\tilde{\beta}_0^{(j\ell)} + \tilde{\boldsymbol{\beta}}_1^{(j\ell)\intercal}\mathbf{z}(\mathbf{s}^*, t) + \mathbf{b}_{\mathbf{s}^*}^{\intercal}\tilde{\boldsymbol{\xi}}_t^{(j\ell)}\right)w_\ell,$$

where $\tilde{\beta}_0^{(j\ell)}$, $\tilde{\beta}_1^{(j\ell)}$, and $\{\tilde{\boldsymbol{\xi}}_t^{(j\ell)}\}_{t=1,\ldots,T}$ are the j^{th} posterior samples from the first-stage model fitted conditional on $\alpha_1 = \alpha_1^{(\ell)}, \ell = 1, \ldots, L$.

5.6 Results

This section presents the results for the dengue models, where the input (first-stage) models are the data fusion models in Chapter 4. The results for a stations-only model

input, and from a longer time series, can be found in Section B.2 of Appendix B.

Figure B.1 in Appendix B shows a pairwise scatter plot of the three climate variables. It shows that relative humidity and log rainfall are positively correlated, while temperature and relative humidity are negatively correlated. On the other hand, temperature and log rainfall are not correlated. Hence, I considered two health models: the first model considers temperature and log rainfall as climate variables, while the second model only has relative humidity as the climate variable. The results for each are presented in Sections 5.6.1 and 5.6.2, respectively. I considered the following additional covariates in the model at the province level: population density (`PopDensity`), and a binary variable which indicates if a time point is during the COVID-19 pandemic (`covid`).

5.6.1 Temperature and log rainfall

The linear predictor of the health model with temperature and log rainfall as climate covariates is given as follows:

$$\begin{aligned}\eta\Big(B_i, t\Big) = \gamma_0 + \gamma_1 \widehat{\text{Temperature}}(B_i, t) + \gamma_2 \widehat{\text{Temperature}}^2(B_i, t) + \gamma_3 \log \widehat{\text{Rain}}(B_i, t) + \\ \gamma_4 \text{ClimateType}(B_i, t) + \gamma_5 \log \widehat{\text{Rain}}(B_i, t) \times \text{ClimateType}(B_i, t) \\ + \gamma_6 \text{covid} + \gamma_7 \log \text{PopDensity} + \varphi(B_i, t)\end{aligned} \tag{5.13}$$

In Equation (5.13), I considered a non-linear effect of temperature, following Xu et al. (2020). The `ClimateType` variable is a binary variable which takes a value of '1' for the eastern section of the country, and takes '0' for the western section, as defined in Chapter 4. The country's western section has a pronounced dry and wet season, while the eastern part has relatively high rainfall all year round (Coronas, 1920; Kintanar, 1984a). Moreover, I considered an interaction effect between log rainfall and climate type. This is based on the results in Cawiding et al. (2025), which showed that the effect of rainfall varies for different regions of the country (see Section 5.3). The values $\widehat{\text{Temperature}}(B_i, t)$ and $\widehat{\text{Rain}}(B_i, t)$ are computed using Equation (5.3); see also Algorithms 5.1 and 5.2.

Table 5.1 shows the marginal log likelihood (MLik), Watanabe-Akaike Informa-

tion Criterion (WAIC), and the conditional predictive ordinate (CPO) values for the different models considered. These values are based on the results from the crude plug-in method (Algorithm 5.1). Results show that Type II interaction model has the highest MLik, the smallest WAIC, and the smallest CPO value as well. Hence, the Type II interaction model was considered for further investigation.

Model	MLik	WAIC	CPO
Type I	-7368.66	10658.92	14182.67
Type II	-6814.81	10526.76	7694.59
Type III	-7337.89	10705.02	12526.74
Type IV	-11488.96	22881.93	14366.48

Table 5.1: Marginal log likelihood (MLik), WAIC, and $-\sum \log \text{CPO}_i$ for different dengue models with temperature and log rainfall as climate covariates

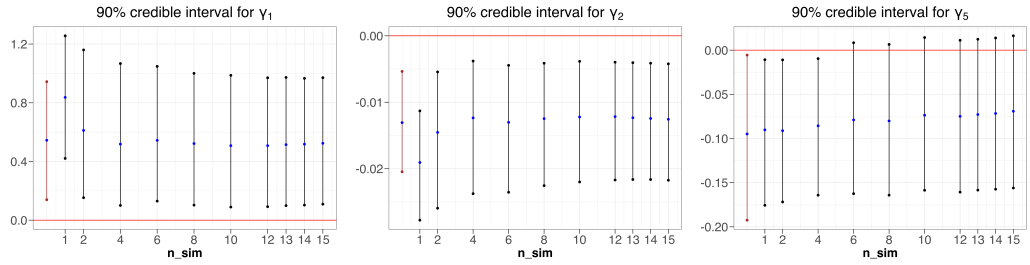
Table 5.2 shows the fixed effects estimates. The results show that temperature has a non-linear relationship with dengue. In particular, the higher the temperature, the higher the risk; however, too high temperature leads to a decline in the risk. Moreover, although the main effect of log rainfall is not significant in the model, the interaction between log rainfall and climate type is significant and is negative, with the plug-in approach. The results imply that for a 10% increase in the amount of rainfall, there is an expected decline in the risks by around 0.43% for areas in the eastern section of the country, i.e., for areas with uniform amounts of rainfall and relatively wet all year round, log rainfall and dengue are negatively related. This agrees with the results from Cawiding et al. (2025), which explained that the constant amount of rainfall tends to flush out breeding sites for mosquitoes; thus decreasing the risk of dengue. Note that the resampling approach increased the posterior standard deviation of the coefficient (γ_5) of this interaction effect, causing it to be no longer significant. Lastly, population density and dengue are positively associated, while the `covid` binary variable is not significant. For the case with a stations-only climate model as input (see Section B.2 of Appendix B), the coefficient of log rainfall is significant. The results suggest that for areas in the western section of the country, a 10% increase in rainfall is associated with an increase in the risks by around 0.09% or 0.07% based on the plug-in method and resampling method, respectively.

The results show that the posterior standard deviations from the resampling method are generally larger compared to those of the plug-in method, except for

Parameter	Plug-in method				Resampling method			
	Mean	SD	P5%	P95%	Mean	SD	P5%	P95%
γ_0, Intercept	-6.2806	3.0463	-11.4991	-1.9470	-6.3940	3.3271	-12.0272	-1.0479
γ_1, Temperature	0.5332	0.2368	0.1545	0.9158	0.5326	0.2665	0.1043	0.9781
γ_2, Temperature^2	-0.0132	0.0050	-0.0206	-0.0053	-0.0127	0.0053	-0.0218	-0.0042
γ_3, log Rain	-0.0176	0.0275	-0.0645	0.0242	-0.0207	0.0239	-0.0607	0.0182
γ_4, ClimateType	0.3493	0.3479	-0.2184	0.9296	0.2016	0.3578	-0.3920	0.7910
γ_5, log Rain × ClimateType	-0.0886	0.0543	-0.1786	-0.0075	-0.0689	0.0528	-0.1561	0.0185
γ_6, covid	-0.1396	0.0922	-0.2879	0.0228	-0.1375	0.0915	-0.2892	0.0127
γ_7, log PopDensity	0.2124	0.0934	0.0498	0.3590	0.1998	0.0967	0.0404	0.3585

Table 5.2: Comparison of estimates of fixed effects between the plug-in method and the resampling method for the dengue model with temperature and log rainfall as climate covariates

coefficients whose credible intervals (CI) contain the null value. Figure 5.4 shows a clear comparison of the posterior uncertainties for γ_1, γ_2, and γ_5. An attenuation of the estimated posterior means to the null risk can be observed using the resampling method. This is also observed in Lee et al. (2017) and Liu et al. (2017), where they argue that it is due to the posterior predictive distribution from the first-stage model outweighing the spatio-temporal variation in the data. The same plots for the other covariates are shown in Figure B.2 of Appendix B.



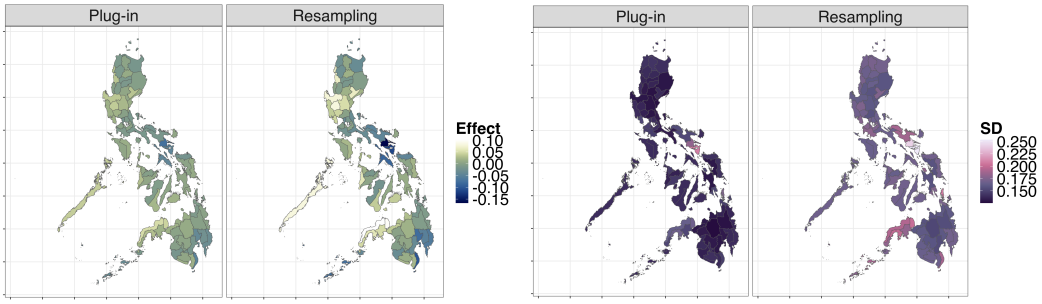
(a) γ_1, Temperature **(b)** γ_2, Temperature^2 **(c)** γ_3, log Rain × ClimateType

Figure 5.4: Plot showing the posterior means and 90% credible intervals for the following parameters: (a) γ_1 (b) γ_2 (c) γ_5; for the model with temperature and log rainfall as climate covariates. The first vertical line shows the estimates for the plug-in method, while the rest of the lines show the estimates for the resampling method for different number of resamples, from 1 to 15.

Table B.1 in Appendix B shows the estimates of the hyperparameters. The results show that the posterior uncertainty is generally higher for the resampling method compared to the plug-in method. Moreover, the structured effect in time accounts for more variability in the data compared to the unstructured effect. The mixing parameter ϕ of the spatial effect is less than 0.5 for the plug-in method, but greater than 0.5 for the resampling method. This suggests that the variability explained by the structured spatial effect is smaller compared to the unstructured effect for the plug-in method, but it is the opposite for the resampling method. The results also

suggest that each province has its own temporal structure, which is independent of the other provinces. The time dependence in the interaction effect is strong, since the estimated autoregressive parameter is close to 1.

Figure 5.5b shows a comparison of the posterior standard deviation of the space effects $\psi(B_i)$ between the plug-in method and resampling method. The figure shows that the resampling method give higher uncertainty in the spatial effects. The posterior means of $\psi(B_i)$ are quite similar between the two methods as shown in Figure 5.5a.

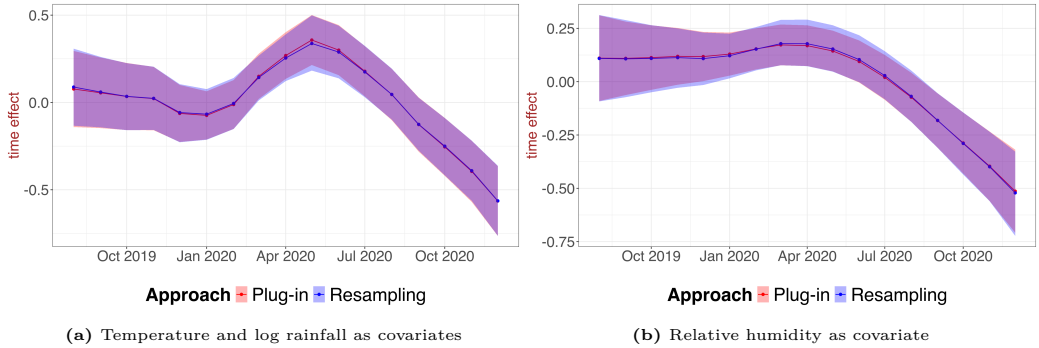


(a) Posterior mean **(b)** Posterior standard deviation

Figure 5.5: Comparison of (a) posterior mean and (b) posterior standard deviation, of the space effects $\psi(B_i)$ between the plug-in method and resampling method, for the dengue model with temperature and log rainfall as climate covariates

Figure 5.6a shows a plot of the estimated structured temporal effects with the corresponding 95% credible intervals. The plot shows that the posterior uncertainty is very similar between the two methods. Moreover, the plot also shows that the estimated temporal effect is decreasing during 2020, which is the COVID-19 episode. This potentially explains why the `covid` binary variable in the model is not significant, since the drop in the reported dengue cases during this year is already accounted for by the structured temporal effect.

Figure 5.7 shows the estimated space-time interaction effect $\nu(B_i, t)$ using the Type II interaction model (see Section 5.4.5) for five provinces. Four of the provinces, which are contiguous and constitute a major island, have decreasing space-time interaction effect pre-pandemic. This also agrees with the trend in the SIRs for the same period (shown in the top portion of Figure 5.7). For the fifth province considered (Ilocos Norte), which is located in the north, the estimated space time effect shows an increasing trend pre-pandemic, which also agrees with the SIRs. During



(a) Temperature and log rainfall as covariates

(b) Relative humidity as covariate

Figure 5.6: Plot of the estimated structured time effects $\nu(t)$ with the 95% credible intervals between the plug-in method and resampling method: (a) temperature and log rain as climate covariates (b) relative humidity as covariate

the COVID-19 pandemic, the estimated effects vary for the 5 provinces, and which shows patterns distinct to each province. This figure confirms the Type II interaction in space and time, which means that in addition to the overall temporal effect, each province exhibits its own temporal structure which is independent of the other provinces.

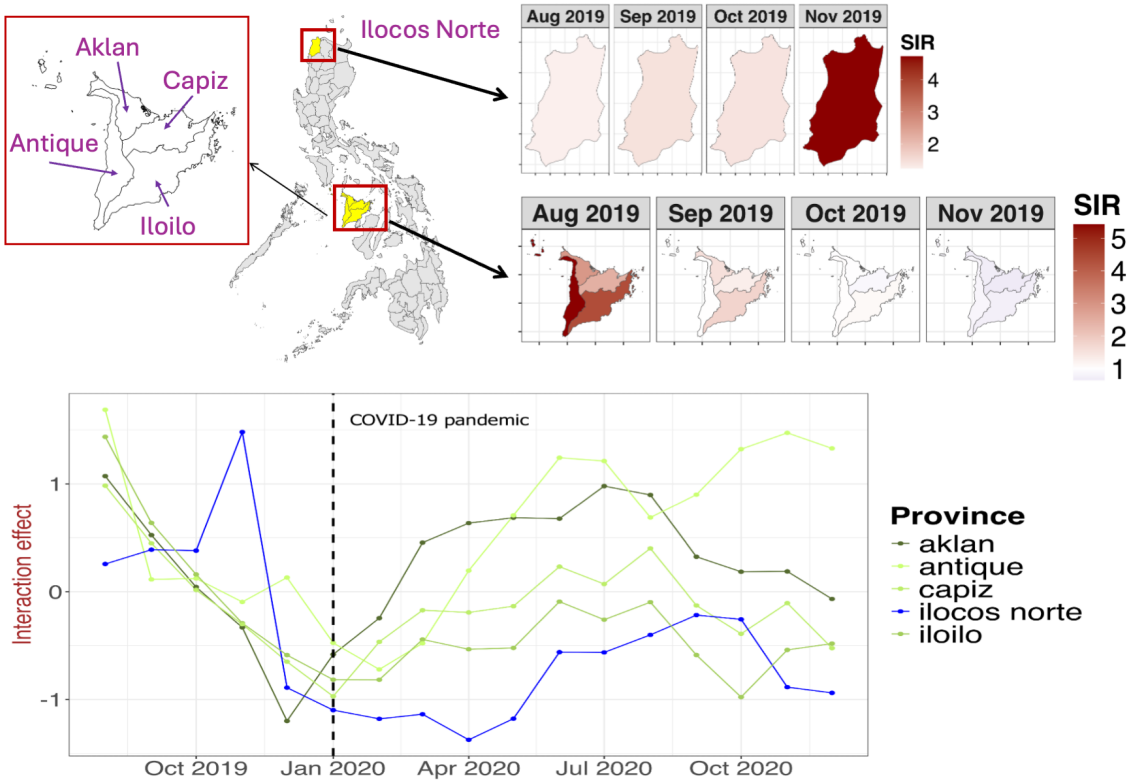


Figure 5.7: Estimated space-time interaction effect $\upsilon(B_i, t)$ for five provinces. Four of them are contiguous provinces which exhibit the same temporal structure pre-pandemic, and which also agrees with the trend in the SIRs. The fifth province (located in the north) has a decreasing trend in the SIRs for the same time period, and is also accounted for by the space-time effect. The temporal structure during the pandemic varies for the five provinces.

5.6.2 Relative humidity

The linear predictor of the health model with relative humidity as climate covariate has the following form:

$$\eta\Big(B_i,t\Big) = \gamma_0 + \gamma_1\widehat{\text{RH}}(B_i,t) + \gamma_2\text{ClimateType}(B_i,t) + \gamma_3\widehat{\text{RH}}(B_i,t) \times \text{ClimateType}(B_i,t) + \gamma_4\text{covid} + \gamma_5 \log \text{PopDensity} + \varphi(B_i,t) \tag{5.14}$$

Table 5.3 shows a summary of the metrics to compare the models with different interaction types. The results show that the Type II interaction has the highest marginal log likelihood, the smallest WAIC, and the smallest CPO value. Thus, similar to the model with temperature and log rainfall as climate covariates, the model with Type II interaction was considered for further investigation. The values $\widehat{\text{RH}}(B_i,t)$ are computed using Equation (5.3); see also Algorithms 5.1 and 5.2.

Model	MLik	WAIC	CPO
Type I	-7363.05	10652.39	14247.44
Type II	-6808.63	10522.15	7756.46
Type III	-7326.95	10701.90	12542.34
Type IV	-12091.47	23698.59	14558.67

Table 5.3: Marginal log likelihood (MLik), WAIC, and $-\sum \log \text{CPO}_i$ for different dengue models with relative humidity as climate covariate

Table 5.4 shows the estimates of the fixed effects. The main effect of relative humidity is significant and positive. The interaction between relative humidity and climate type is also significant and negative, both for the plug-in approach and the resampling approach. This is the same relationship that log rainfall has with dengue, which is expected since relative humidity and log rainfall are positively correlated. For areas in the eastern section of the country, a one-unit increase in RH is associated with a 1.84% or 1.40% decline in the risks based on the plug-in method and the resampling method, respectively. On the other hand, for areas in the western section of the country, a one-unit increase in RH is associated with a 1.59% or 1.56% increase in risks based on the plug-in method and the resampling method, respectively. Moreover, both population density and the `covid` variable are not significant. The non-significance of the `covid` variable is potentially due the temporal random effect

accounting for the decline in the dengue risks (see Figure 5.6b). The posterior standard deviations in the coefficients are generally higher for the resampling approach compared to the plug-in method. A comparison of the 90% credible intervals, similar to Figure 5.4, is shown in Figure B.3 of Appendix B.

	Plug-in method				Resampling method			
Parameter	Mean	SD	P5%	P95%	Mean	SD	P5%	P95%
γ_0, Intercept	-2.2360	0.8622	-3.6905	-0.8305	-2.1124	0.8681	-3.5365	-0.6733
γ_1, RH	0.0170	0.0071	0.0039	0.0294	0.0155	0.0073	0.0035	0.0276
γ_2, ClimateType	2.5001	1.4705	0.5529	5.0257	2.2384	1.4407	-0.1160	4.5985
γ_3, RH × ClimateType	-0.0356	0.0163	-0.0616	-0.0072	-0.0296	0.0167	-0.0577	-0.0021
γ_4, covid	-0.0668	0.0632	-0.1589	0.0497	-0.0645	0.0680	-0.1777	0.0479
γ_5, log PopDensity	0.0953	0.0896	-0.0452	0.2347	0.1035	0.0928	-0.0487	0.2582

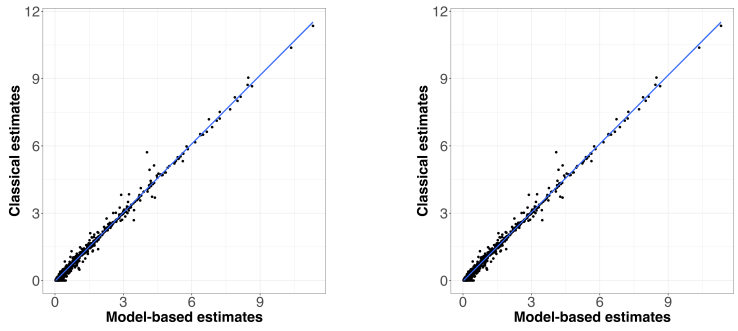
Table 5.4: Comparison of estimates of fixed effects between the plug-in method and the resampling method for the dengue model with relative humidity as climate covariate.

Table B.2 in Appendix B shows the estimated hyperparameters. The results show that the posterior standard deviations for the hyperparameters are significantly higher for the resampling method. Figure B.4b in Appendix B shows a comparison of the posterior SD of the spatial effects $\psi(B_i)$. Here, it is also apparent that the posterior standard deviation from the resampling method is higher compared to the plug-in method. The posterior means of $\psi(B_i)$ are provided in Figure B.4a in Appendix B, which shows that the estimated posterior means between the plug-in method and resampling method are similar. Finally, Figure 5.6b shows a comparison of the posterior means and 95% credible intervals for the structured time effect $\nu(t)$.

5.6.3 Estimated risks

Figure 5.8 shows a comparison of the observed SIRs, which are viewed as classical estimates of risks, versus the model-based estimates $\hat{\lambda}(B_i, t)$ using Equation (5.13), i.e., the model with temperature and log rainfall as climate covariates. The figure shows an agreement between the classical estimates and model-based estimates, from both the plug-in method and resampling method. Figure B.5 in Appendix B shows the same scatter plots, but from the dengue model with relative humidity as the climate covariate. The results also show a general agreement between the classical estimates and the model-based estimates of SIR.

Figure 5.9 shows the estimated risks $\hat{\lambda}(B_i, t)$ for months August to November



(a) Plug-in method **(b)** Resampling method

Figure 5.8: Comparison of classical SIR estimates and model-based SIR estimates from the health model with temperature and log rainfall as climate covariates: (a) plug-in method (b) resampling method

2019, and from using the plug-in method and resampling method, on the health model with temperature and log rainfall as climate covariates. Firstly, the plot shows that both the plug-in method and resampling method have equivalent estimates, which is expected based on Figure 5.8. Secondly, the plot also agrees with Figure 5.2, which shows the classical SIR estimates. These maps show specific areas with elevated SIRs. As noted in Section 5.2 and Section 1.2.1 in Chapter 1, the Philippines declared dengue epidemic during August 2019 due to a surge in dengue cases (BBC, 2019; Santos, 2019; Yeung and Faidell, 2019). The same spatial structure is also evident when producing the same maps of SIRs based on the health model with relative humidity as the climate covariate, since Figure B.5 in Appendix B also shows a general agreement between the model estimates and the classical estimates of SIR.

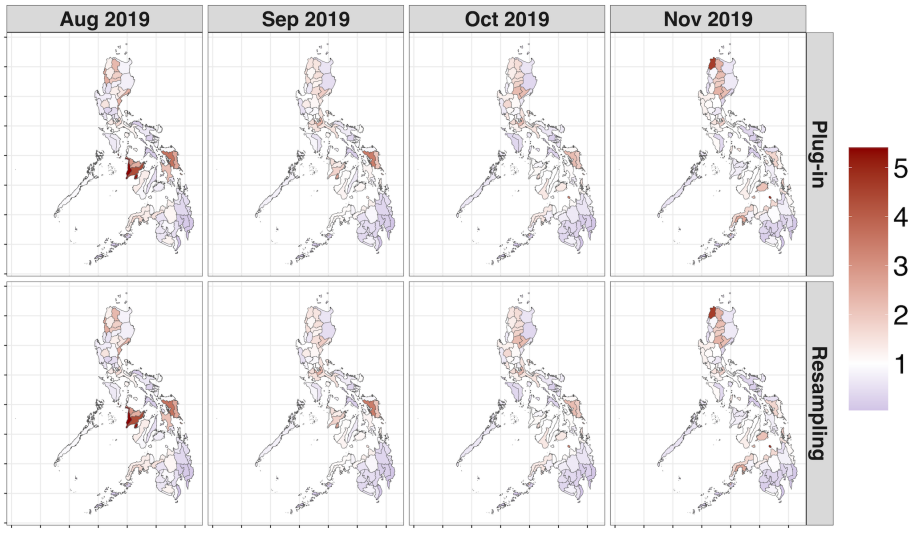


Figure 5.9: Model-based estimates of dengue risks from August 2019 to November 2019, for both plug-in method and resampling method on the dengue model with temperature and log rainfall as climate covariates

Figure B.6 in Appendix B shows the corresponding posterior uncertainty of the estimated SIRs in Figure 5.9. The results show that there is no difference in the posterior uncertainty in the predicted SIRs between the plug-in method and the resampling method. To investigate this further, I looked at the variance-covariance structure of the different components of the linear predictor (Equation 5.13) across the different resamples. In particular, for each posterior sample, I first computed the posterior variance-covariance matrix of the model components. I then averaged the values for all resamples. Matrix (B.1.1) in Appendix B shows the variance-covariance structure for the fixed effects (across resamples) of the linear predictor in Matrix (5.13). Note that most of the covariances are negative. Moreover, Matrix (B.1.2) in Appendix B shows the variance-covariance matrix for the random effects in the linear predictor. The results show that most of the covariances are close to zero. Finally, Matrix (B.1.3) in Appendix B shows the cross-covariance between the fixed and random effects in the linear predictor, which shows an equal mix of positive and negative linear association between the components. Since most of the pairs of components in the linear predictor in Equation (5.13) are negatively correlated across the resamples, then this potentially explains why the posterior uncertainty in the dengue risks in the resampling method is similar to the plug-in method. Although the resampling method generally gives higher uncertainty for individual components of the linear predictor, the uncertainty in a linear combination of these components can be washed away because of the latent correlation structure.

Figure 5.10 shows a comparison of the posterior standard deviations in the estimated risks $\lambda(\hat{B}_i, t)$ between three approaches: a classical approach based on the asymptotic (Gaussian) distribution of the SIR, model-based estimates from the plug-in approach, and model-based estimates from the resampling approach. Note that there are relatively higher uncertainty values for the classical approach, while the estimates from the model-based plug-in and resampling approaches are almost indistinguishable. On average, the classical approach has higher uncertainty estimates than the model-based approaches (see the broken lines in Figure 5.10).

Finally, Figure 5.11 shows the probability that the dengue risks $\lambda(B_i, t)$ exceed 1, i.e., $\mathbb{P}\Big(\lambda(B_i, t) > 1\Big)$ for August 2019 to November 2019, for both the plug-in method

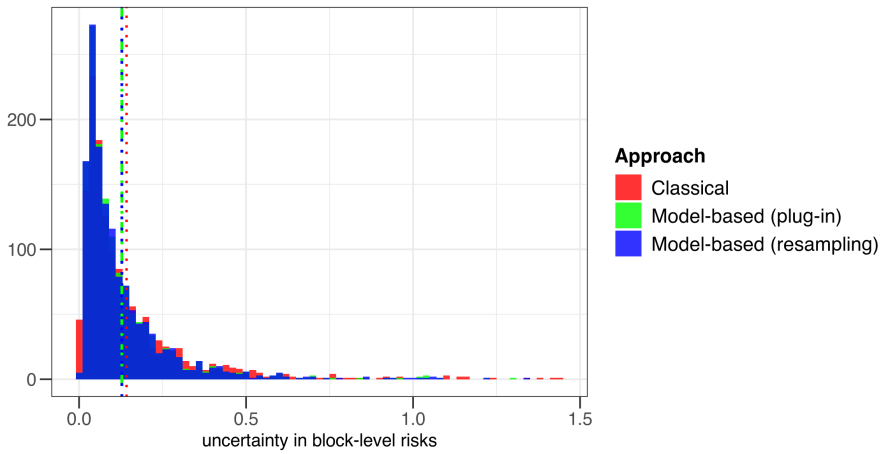


Figure 5.10: Comparison of the posterior standard deviations in the estimated risks $\lambda(\hat{B_i}, t)$ between three approaches: classical approach based on the asymptotic (Gaussian) distribution of the SIR, model-based estimates from the plug-in approach, model-based estimates from the resampling approach. The model here has temperature and log rainfall as climate covariates. The broken lines are the means of the values for each approach.

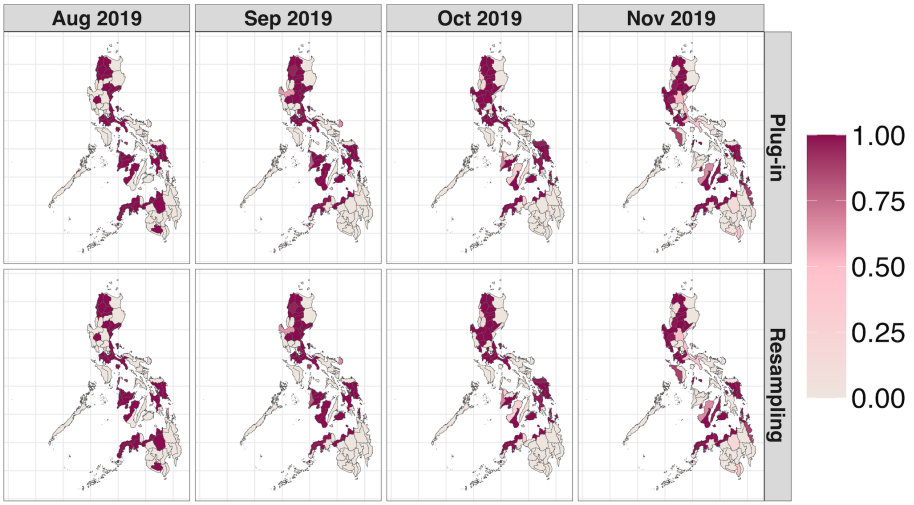


Figure 5.11: Probability of exceedence, i.e., $\mathbb{P}(\lambda(B_i, t) > 1)$ from August 2019 to November 2019, for both plug-in method and resampling method on the dengue model with temperature and log rainfall as climate covariates

and resampling method. Note that most of the areas with an estimated probability of exceedence equal to 1 are the same areas badly hit by dengue during the epidemic (Yeung and Faidell, 2019).

5.7 Conclusions

The main aim of this chapter is to provide additional evidence on the relationship between climate and dengue in the Philippines. I proposed a Bayesian spatio-temporal model for dengue, where the effect of climate covariates are considered as fixed ef-

fects, and which incorporates both structured and unstructured random effects in space and time, including their interaction, in order to account for extra variability in the data unexplained by the climate variables. I used the integrated nested Laplace approximation (INLA) approach (Rue and Martino, 2007; Van Niekerk et al., 2023) to perform model inference. The Bayesian modelling framework that this work employed has not been used based on currently published work which links climate and dengue in the Philippines (see Section 5.1). Moreover, most of the existing work only looks at certain regions of the country; while this work looks at the data for the entire country.

The association between climate and dengue has long been established in the literature (Hales et al., 2002; McMichael, 2003; Naish et al., 2014). The results in this chapter agree with existing studies on climate and dengue. Results show that temperature has a non-linear relationship. In particular, for very high temperature values, the association becomes negative (Liu et al., 2023), which is explained by the fact that excessively high temperature can shorten the lifespan of mosquitoes and reduce their population size (Myer et al., 2020). Moreover, results show that rainfall has varying effects on dengue, depending on the spatial location of an area, which is defined based on the climate type of the region. In particular, I segmented the country into the eastern and western section. In the western section, rainfall and dengue are positively related, while in the eastern section, the relationship is negative. The eastern section of the country has a low variation in the amount of rainfall and is relatively wet all year round. This phenomenon tends to wash away the breeding sites of mosquitoes, thus explaining its negative relationship with dengue. On the other hand, for the western section, episodes of wet and dry conditions are more pronounced. Phenomena of sporadic rainfall during dry conditions create more breeding sites for mosquitoes, which enhances dengue transmission. This agrees with the results in Cawiding et al. (2025), who also looked into this spatially varying effect of rainfall on dengue in the Philippines. Relative humidity, which is highly correlated with rainfall, also exhibits the same association with dengue.

The spatial and temporal effects in the model capture extra variation in the data unexplained by the climate variables and other covariates. When it comes to the

spatial effect, most of the variability in space is explained by the conditional autoregressive component, i.e., there are remaining spatial correlations in the data after accounting for the covariate effects. Similarly, for the temporal effect, the structured (random walk) time effect accounts for more of the variability in the data compared to the unstructured effect. In fact, there is a clear decline in the estimated temporal effects during 2020, which is the start of the COVID-19 pandemic, and also a year of very low number of reported dengue cases. The `covid` binary variable in the model was not significant, most likely since the information is already captured by the random walk effect in time. Moreover, an interaction effect between space and time is shown to be important in the model. In particular, the interaction specifies that each province has its own temporal structure that is independent of the other provinces. A more advanced specification of the random effects is to use the variance partitioning approach proposed in Franco-Villoria et al. (2022). Essentially, it specifies a global (space-time) precision parameter, which is partitioned into the main effect (combined space and time) and the interaction effect via a mixing parameter. The variance explained by the main effect is further partitioned into the space and time effects via another mixing parameter. This kind of specification enhances model interpretability, and also allows an intuitive prior specification. This is further discussed in Section 7.7.1.2 in Chapter 7.

This chapter used a two-stage modelling approach to link climate and dengue. The first stage fits the climate models, and then the second stage fits the health model for dengue using the climate predictions from the first stage as input. This is a practical framework for doing analysis since the climate models are complex, especially the data fusion models. Also, this chapter used climate predictions from existing data fusion models in Chapter 4. Thus, a joint modelling framework, i.e., fitting both climate model and health model, is unnecessary. More important reasons for not pursuing a joint modelling framework are discussed in Section 1.3.

The use of a two-stage approach requires proper propagation of the uncertainty from the first-stage model to the second-stage model. In this work, I used the posterior sampling approach. This approach considers different realizations from the estimated posteriors in the first stage, specifically the climate predictions which we

input to the second-stage model. This implies a set of estimated health models, one for each posterior sample. The final posterior estimates for the second-stage model parameters are then combined using model averaging. The posterior uncertainties were stable after around 12 resamples. When it comes to the differences in the posterior uncertainties, the resampling method generally gave higher posterior standard deviations of model parameters compared to the plug-in method. A methodological innovation in this area is to consider an uncertainty propagation approach which does not require resampling from the posteriors and which does not require fitting the second-stage model several times. This is pursued in Chapter 6.

In this chapter, I computed the block-level estimates of the first-stage latent process $x(\mathbf{s}, t)$, which are then used as a covariate in the second-stage model. A future work related to the current problem of linking a point-referenced covariate and an areal response variable is to use a new model specification which defines a latent intensity field in the second stage. The predictor expression in the second stage would take the following form: $\log\Big(\mu(B,t)\Big) = \log\Big(\int_B \mu(\mathbf{s},t)d\mathbf{s}\Big) = \log\Big(\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t)\Big\}d\mathbf{s}\Big)$, where $\mu(B,t)$ is the mean of the Poisson count for block B at time t. This assumes that the mean count at block B at time t is an aggregation of an intensity field $\mu(\mathbf{s},t)$ over B, and where $\mu(\mathbf{s},t)$ is non-linearly related to the first-stage latent process $x(\mathbf{s},t)$. Whereas the model used in this work specifies the predictor expression in the second-stage as linear with respect to $x(\mathbf{s},t)$, the new specification is a highly non-linear model. This type of model is straightforward to implement using the `inlabru` library, which extends the class of models that can be fitted using INLA, specifically models which are non-linear in the latent parameters (Bachl et al., 2019; Lindgren et al., 2024). This new specification of the Poisson model is further discussed in Chapter 6 and in Section 7.7.1.3 of Chapter 7.

There are several important indicators of dengue which I did not incorporate in this chapter. One important index is the Southern Oscillation Index (SOI), which is an indicator of *El Niño* and *La Niña* episodes. The former is an episode of above-average temperature levels, while the latter implies colder and wetter conditions. A positive SOI is associated with much warmer and wetter conditions than the average, which is ideal for breeding of mosquitoes (McMichael, 2003). SOI is shown to be an

important indicator of dengue transmission, but the magnitude of its effects could vary for different countries (Hales et al., 1999, 1996; McMichael, 2003). Another extension of the model is to incorporate lagged effects of the climate variables (Carvajal et al., 2018; Cruz et al., 2024), which are also shown to be significant indicators of dengue transmission. This can be pursued in a future work, but I think that this should be used on data with higher time resolution, such as considering weekly cases and daily records of the climate indicators. Future models should also consider vector abundance and biological characteristics of pathogens (Murphy et al., 2022). Finally, social factors and economic factors are also important indicators in the model. Examples of social factors are human behaviour, such as water storage practices, and land use, such as irrigation/forest clearance/livestock and agricultual practices. Moreover, some economic factors are poverty, population displacement/travel, housing, urbanization, and public health infrastructure (McMichael, 2003). There is a complex interaction among the aforementioned factors and the transmission of infectious diseases (Foster, 2001), which is something that should be carefully considered for future work.

Chapter 6

Validating uncertainty propagation approaches for two-stage Bayesan models using simulation-based calibration

Chapter 1 discussed the motivation for employing a two-stage modelling framework. Essentially, a two-stage model is either more practical or more appropriate in many situations due to three important reasons. Firstly, it is more computationally efficient when the first-stage model is already complex in itself. Secondly, it has an intuitive physical interpretation since there is a clear one-directional relationship between the two physical processes, e.g., climate and concentration of pollutants affect disease risks but not the other way around. Thirdly, it avoid the 'feedback' problem which happens in a joint modelling approach, as explained in Section 1.3.1.

The first main contribution of this chapter is that I evaluate the correctness of two existing approaches for doing two-stage modelling, particularly the crude plug-in method and the posterior sampling approach (or resampling approach). These were used in Chapter 5, and discussed in Algorithms 5.1 and 5.2, respectively. The correctness of the two aforementioned methods are evaluated using the simulation-based calibration (SBC) method (Talts et al., 2018), which tests for the self-consistency

property of Bayesian algorithms. The SBC is a method used to validate a Bayesian algorithm. It has a frequentist interpretation since it is interpreted as the expected behavior averaged over all potential data outcomes. Talts et al. (2018) argued that SBC is an integral part of a robust *Bayesian workflow*, which includes the following key three steps: model building, inference, and model checking/improvement (Gelman et al., 2020).

The second main contribution is that this chapter proposes a new approach for uncertainty propagation in two-stage Bayesian models, which is called the **Q** uncertainty method and is implemented using INLA. In this work, I illustrate and validate the correctness of the **Q** uncertainty method in spatial applications. The proposed method introduces a new model component, called an *error component*, in the second-stage model. The error component is given a Gaussian prior with zero mean and $\mathbf{Q}^{-1}$ covariance matrix which encodes the full uncertainty from the first-stage model. The proposed method has a similar flavor as the *prior exposure method* proposed in Cameletti et al. (2019), but here I consider the full covariance structure of the latent parameters of the first-stage model. This is also similar to the ones in Chang et al. (2011) and Peng and Bell (2010), which used the posterior results of the first-stage model as the prior model in the second-stage model, but here I instead introduce a new model component which accounts for the uncertainty in the first-stage model parameters. Moreover, I explore a low-rank approximation of the error component, which can be useful for spatial models with high dimensions, such as large spatio-temporal models. Thus, there are two versions of the proposed method: the *full* **Q** *uncertainty approach* and the *low rank* **Q** *uncertainty approach*. I also validate the correctness of the proposed methods using the SBC method.

Thirdly, this chapter proposes a variation in the original SBC method, which is motivated by scenarios wherein some model parameters in the first-stage model violate the self-consistency property, but the primary parameters of interest are the second-stage model parameters. I implemented both the original SBC and the proposed variation in the simulation experiments.

Section 6.1 formally discusses the two-stage modelling framework and the uncertainty propagation problem. Section 6.1.1 discusses two current approaches for

two-stage modeling: the crude plug-in approach and the resampling approach. In Section 6.1.2, I discuss the proposed methods, which are applied in the INLA framework and are illustrated in the context of spatial applications. Section 6.2 revisits the self-consistency property of Bayesian models and the SBC method, and presents the proposed variant of the SBC method. Section 6.3 discusses results from simulation experiments that validates four uncertainty propagation approaches: the crude plug-in approach, the resampling approach, the full $\mathbf{Q}$ approach, and the low rank $\mathbf{Q}$ approach. I start with a two-stage spatial model with a Gaussian likelihood in Section 6.3.1, and then consider the case of a Poisson likelihood in Section 6.3.2. I highlight the SBC results for the second-stage model parameters, since these are the parameters whose posterior uncertainty is potentially underestimated in a two-stage modelling framework. I then demonstrate the proposed method in a real data application in Section 6.4. The data application aims to link relative humidity and the case counts of dengue fever in the Philippines for August 2018. Finally, I end with conclusions and future work in Section 6.5.

6.1 Uncertainty propagation problem

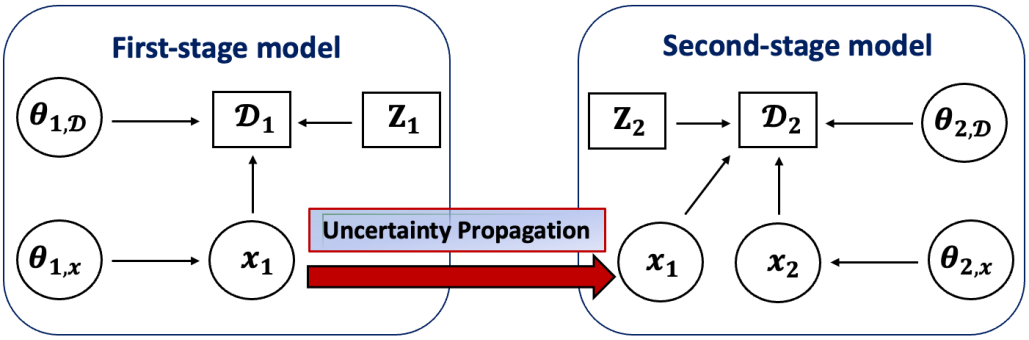


Figure 6.1: Two-stage modelling framework for uncertainty propagation

Here, it is assumed that model inference for a physical process of interest is performed in two stages, as shown in Figure 6.1. The observed data is $\mathcal{D} = \{\mathcal{D}_1, \mathcal{D}_2\}$, partitioned into the first-stage data and second-stage data, respectively. First-stage inference is performed using $\mathcal{D}_1$ only, without looking at $\mathcal{D}_2$, since for instance, the health counts are not intended to inform the estimation of exposures or the health biomarkers causes the probability of survival and not the other way around.

I model such process using a Bayesian hierarchical model. Suppose that $\boldsymbol{x}_1$ and $\boldsymbol{\theta}_1$ are the latent parameters and hyperparameters linked to $\mathfrak{D}_1$, respectively. I partition $\boldsymbol{\theta}_1$ into $\{\boldsymbol{\theta}_{1,\boldsymbol{x}}, \boldsymbol{\theta}_{1,\mathfrak{D}}\}$, for which $\boldsymbol{\theta}_{1,\boldsymbol{x}}$ are the hyperparameters linked to $\boldsymbol{x}_1$, while $\boldsymbol{\theta}_{1,\mathfrak{D}}$ are variance/scaling parameters of the assumed distribution (likelihood) for the data $\mathfrak{D}_1$. Also, let $\mathbf{Z}_1$ be a set of fixed inputs/covariates. I assume that $\mathfrak{D}_1 \sim \mathcal{F}_1(\mathfrak{D}_1|\boldsymbol{x}_1, \boldsymbol{\theta}_1, \mathbf{Z}_1)$. Similarly, let $\boldsymbol{x}_2$ and $\boldsymbol{\theta}_2 = \{\boldsymbol{\theta}_{2,\boldsymbol{x}}, \boldsymbol{\theta}_{2,\mathfrak{D}}\}$ be the latent parameters and hyperparameters linked to $\mathfrak{D}_2$, respectively, and $\mathbf{Z}_2$ be a set of fixed inputs. The modeling framework in Figure 6.1 assumes that $\boldsymbol{x}_1$ from the first-stage model is also linked to $\mathfrak{D}_2$, so that $\mathfrak{D}_2 \sim \mathcal{F}_2(\mathfrak{D}_2|\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{\theta}_2, \mathbf{Z}_2)$. This gives us the full data model: $\{\mathfrak{D}_1, \mathfrak{D}_2\} \sim \mathcal{F}_2(\mathfrak{D}_2|\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{\theta}_2, \mathbf{Z}_2)\mathcal{F}_1(\mathfrak{D}_1|\boldsymbol{x}_1, \boldsymbol{\theta}_1, \mathbf{Z}_1)$.

Figure 6.1 considers $\boldsymbol{x}_1$, or some function of it, as an input when fitting the second-stage model. However, in practice, $\boldsymbol{x}_1$ is unknown and needs to be estimated in the first stage. Hence, its uncertainty due to estimation error or model misspecification error needs to be correctly propagated into the second stage; otherwise, the standard errors of the second-stage model parameters may be underestimated. The end goal, therefore, is to correctly estimate the following posterior distributions:

- The posterior distribution of the first-stage parameters, given by $\pi(\boldsymbol{x}_1, \boldsymbol{\theta}_1|\mathfrak{D}_1)$.
- The posterior distribution of the second-stage parameters. For the plug-in method, this is given by $\pi(\boldsymbol{x}_2, \boldsymbol{\theta}_2|\mathfrak{D}_2, \boldsymbol{x}_1^*)$, were $\boldsymbol{x}_1^*$ denotes the posterior mean of $\boldsymbol{x}_1$ from the first-stage model results. For the resampling method, this is given by
$$\int \pi(\boldsymbol{x}_2, \boldsymbol{\theta}_2|\mathfrak{D}_2, \boldsymbol{x}_1)\pi(\boldsymbol{x}_1|\mathfrak{D}_1)d\boldsymbol{x}_1.$$

When estimating the posterior distribution of the second-stage model, the uncertainty in $\boldsymbol{x}_1^*$ needs to be accounted for, which is fundamentally the *uncertainty propagation problem*.

6.1.1 Current approaches

This section discusses two existing approaches to fit the two-stage model in Figure 6.1. Although not exhaustive, these approaches serve as benchmarks for comparison

with our proposed approaches. Related methods, which use first-stage posteriors as priors for the second stage, are discussed in Section 6.1.2.

1. **Plug-in Method** – Let $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1} = \mathbb{E}[\boldsymbol{x}_1|\mathcal{D}_1]$ be the posterior mean of $\boldsymbol{x}_1$ estimated from the first-stage model. The crude plug-in method simply uses this as an input to the second stage. The linear predictor of the second-stage model is then:

$$g\Big(\mathbb{E}[\mathcal{D}_2|\cdots]\Big) = \gamma_0\mathbf{1} + \gamma_1\mathbf{h}(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}, \cdot) + \mathbf{Z}_2\boldsymbol{\gamma}_2, \tag{6.1}$$

where $g(\cdot)$ is the link function, $\{\gamma_0, \gamma_1, \boldsymbol{\gamma}_2\}$ are model fixed effects, and $\mathbf{h}(\cdot)$ is a vector-valued function $\mathbf{h}$: $\boldsymbol{x}_1$-space $\rightarrow \mathbb{R}^{\dim(\mathcal{D}_2)}$. Note that Equation (6.1) can also include random effects. The estimated uncertainty in the second-stage posterior distribution, $\pi(\gamma_0, \gamma_1, \boldsymbol{\gamma}_2, \boldsymbol{\theta}_2|\mathcal{D}_2, \hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1})$, is possibly underestimated since it fails to account for the uncertainty in $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}$. This is the approach implemented in Algorithm 5.1 in Chapter 5.

2. **Resampling method** – The resampling method, described in Algorithm 6.1, accounts for the uncertainty in the first-stage model in a natural way, but can be computationally expensive since it requires fitting the second-stage model several times. This approach was adopted in Blangiardo et al. (2016); Liu et al. (2017), and Zhu et al. (2003). A related approach was also implemented in Lee et al. (2017), where a new value $\tilde{\boldsymbol{\mu}}_{\boldsymbol{x}_1}^{(j)}$ is sampled at each iteration of the MCMC algorithm, and then the second-stage model is fitted for each sample. The resampling method was also implemented in Chapter 5 (see Algorithm 6.1).

Algorithm 6.1 Implementation of the resampling method

Repeat steps 1–2 for $j = 1, 2, \ldots, J$:

Step 1: Sample $\tilde{\boldsymbol{\mu}}_{\boldsymbol{x}_1}^{(j)} \sim \hat{\pi}(\boldsymbol{x}_1|\mathcal{D}_1)$.

Step 2: Plug-in the sampled values in the second stage model, i.e., plug-in $\tilde{\boldsymbol{\mu}}_{\boldsymbol{x}_1}^{(j)}$ instead of $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}$ in Equation (6.1). Store all posterior marginals, such as $\hat{\pi}\big(\boldsymbol{\gamma}_1^{(j)}|\mathcal{D}_2, \tilde{\boldsymbol{\mu}}_{\boldsymbol{x}_1}^{(j)}\big)$.

Step 3: All J results are then combined using model averaging, e.g., $\hat{\pi}(\gamma_1|\mathcal{D}_2) = \frac{1}{J}\sum_{j=1}^{J}\hat{\pi}\big(\boldsymbol{\gamma}_1^{(j)}|\mathcal{D}_2, \tilde{\boldsymbol{\mu}}_{\boldsymbol{x}_1}^{(j)}\big)$.

6.1.2 Proposed method – Q uncertainty

In this section, I present the proposed **Q** uncertainty method for uncertainty propagation. This approach avoids multiple runs of the Bayesian algorithm for the second-stage model, offering potential computational efficiency over the resampling method. The method shares similarities with MCMC algorithms used by Chang et al. (2011), Peng and Bell (2010), and Gryparis et al. (2009), where the second-stage model is fitted using first-stage results as an informative prior. Two implementation approaches are noted in the literature: one that allows feedback by updating the prior distribution with second-stage data (Chang et al., 2011; Gryparis et al., 2009), and another that cuts feedback by fixing the prior at each iteration (Peng and Bell, 2010). Our method aligns with the latter, as we also cut feedback. It is also related to the prior-exposure method in Cameletti et al. (2019), but the proposed **Q** method accounts for the full uncertainty in first-stage latent parameters. I propose two versions: the *full* **Q** *uncertainty method* and the *low rank* **Q** *uncertainty method*, the latter being an approximation useful for large **Q** matrices, such as in large spatio-temporal applications. Both methods are implemented within the INLA framework (Rue et al., 2009; Van Niekerk et al., 2023) and demonstrated in spatial applications.

6.1.2.1 Full Q uncertainty method

The first-stage Bayesian hierarchical model, based on Figure 6.1, is as follows:

$$\begin{aligned} \boldsymbol{\mathcal{D}}_1|\boldsymbol{x}_1, \boldsymbol{\theta}_{1,\mathcal{D}} &\sim \prod_{i=1}^{n_1} \pi(\boldsymbol{\mathcal{D}}_{1i}|\boldsymbol{x}_1, \boldsymbol{\theta}_{1,\mathcal{D}}) \\ \boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}} &\sim \mathcal{N}\Big(\mathbf{0}, \mathbf{Q}^{-1}_{\text{prior}}(\boldsymbol{\theta}_{1,\boldsymbol{x}})\Big) \\ \boldsymbol{\theta}_1 = \{\boldsymbol{\theta}_{1,\mathcal{D}}, \boldsymbol{\theta}_{1,\boldsymbol{x}}\} &\sim \pi(\boldsymbol{\theta}_1) \end{aligned} \tag{6.2}$$

Equation (6.2) is a latent Gaussian model since a Gaussian prior is assumed for the latent parameters $\boldsymbol{x}_1$. For inference, the needed quantities are the posteriors $\pi(\boldsymbol{\theta}_1|\boldsymbol{\mathcal{D}}_1)$ and $\pi(\boldsymbol{x}_1|\boldsymbol{\mathcal{D}}_1)$.

The INLA methodology provides a Gaussian approximation to $\pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1, \boldsymbol{\mathcal{D}}_1)$ given by $\pi_G(\boldsymbol{x}_1|\boldsymbol{\theta}_1, \boldsymbol{\mathcal{D}}_1)$, which is computed from a second-order expansion of the log-

posterior density around its mode. In particular, $\pi_G(\boldsymbol{x}_1|\boldsymbol{\theta}_1, \mathcal{D}_1)$ is given by

$$\boldsymbol{x}_1|\boldsymbol{\theta}_1, \mathcal{D}_1 \approx \mathcal{N}\Big(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1), \mathbf{Q}^{-1}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)\Big), \tag{6.3}$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$ is the mean of the Gaussian approximation for a given $\boldsymbol{\theta}_1$, and $\mathbf{Q}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$ is a sparse precision matrix which primarily depends on two components: the graph obtained from the prior of $\boldsymbol{x}_1$ and the graph based on the mapping from $\boldsymbol{x}_1$ to the linear predictors (Van Niekerk et al., 2023), and is also computed given $\boldsymbol{\theta}_1$. The details are found in Section 2.5.2.

The variance-covariance matrix $\mathbf{Q}^{-1}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$ in Equation (6.3) encodes the uncertainty in the latent parameters $\boldsymbol{x}_1$. Its inverse is what we refer to as the $\mathbf{Q}$ matrix, i.e., $\mathbf{Q} \equiv \mathbf{Q}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$. This information is then used when fitting the second-stage model. In particular, in the second-stage hierarchical model, I introduce a new model component $\boldsymbol{\epsilon}$, which I call an *error component*. Its prior model is derived from Equation (6.3), i.e., $\boldsymbol{\epsilon} \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{Q}^{-1}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)\Big)$. In practice, I propose to evaluate $\mathbf{Q}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$ at the mode of $\hat{\pi}(\boldsymbol{\theta}_1|\mathcal{D}_1)$. The predictor in the second stage is then given by

$$g\Big(\mathbb{E}[\mathcal{D}_2|\cdots]\Big) = \gamma_0\mathbf{1} + \gamma_1\mathbf{h}\Big(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1) + \boldsymbol{\epsilon}, \cdot\Big) + \mathbf{Z}_2\boldsymbol{\gamma}_2, \tag{6.4}$$

I call this approach the *full* $\mathbf{Q}$ *uncertainty method*. As defined in Equation (6.1), the $\mathbf{h}(\cdot)$ function accounts for the unequal dimensions between $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}$ and $\mathcal{D}_2$. The domain of the function is the latent $\boldsymbol{x}_1$-space. It then evaluates the first-stage predictor expression, which yields a vector matching the dimension of the second-stage data $\mathcal{D}_2$. The specific functional form of $\mathbf{h}(\cdot)$ depends on the context and data application, and is consequently defined by the user.

In fitting Equation (6.4), the quantity $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$ is assumed to be fixed and known. Since $\mathbf{h}(\cdot)$ is generally a linear function of $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)+\boldsymbol{\epsilon}$, it follows that $\mathbf{h}\big(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)+\boldsymbol{\epsilon}, \cdot\big)$ is also Gaussian. However, γ_1 is also unknown and assigned a Gaussian prior; thus, Equation (6.4) is not a latent Gaussian model since the predictor involves a product of two components, each with a Gaussian prior.

One way to fit the model is to linearize the predictor in Equation (6.4) using a first-order Taylor approximation as implemented in the `R` library `inlabru` (Lindgren

et al., 2024). Another approach is to specify a grid of values for γ_1, and then fit Equation (6.4) conditional on each γ_1. All estimates are then combined using model averaging (Gómez-Rubio, 2020). A third approach is to fit the model using a hybrid INLA with MCMC or importance sampling approach. Here γ_1 is estimated using sampling, while the rest of the parameters are estimated using INLA (Berild et al., 2022; Gómez-Rubio and Rue, 2018).

In the INLA framework, the model component $\boldsymbol{\epsilon}$ in Equation (6.4) involves a scaling or precision parameter, say $\tau_{\boldsymbol{\epsilon}}$, for $\mathbf{Q}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$. I propose fixing the value of this scaling parameter at $\tau_{\boldsymbol{\epsilon}} = 1$. Note that fixing this value to higher (lower) values implies a reduction (increase) in the uncertainty carried over from the first-stage model to the second-stage model. An approach for determining the optimal value of $\tau_{\boldsymbol{\epsilon}}$ could be pursued in future work.

An application to spatial models

In spatial applications, the linear predictor in the first-stage model is a combination of fixed effects and a random field. Under the scenario of a Gaussian likelihood model, a common specification for the function $\mathbf{h}(\cdot)$ in Equation (6.4) is as follows:

$$\mathbf{h}(\boldsymbol{x}_1) = \mathbf{Z}_1\boldsymbol{\beta} + \boldsymbol{\xi}, \tag{6.5}$$

where $\boldsymbol{\beta}$ are fixed effects and $\boldsymbol{\xi}$ is a random field. An efficient method for estimating the random field is the stochastic partial differential equations (SPDE) approach (Lindgren et al., 2011). This approach provides a finite-dimensional but continuously-indexed approximation of Gaussian fields with Matérn covariance function. The discretization is defined on a mesh and expresses the approximation as

$$\xi(\mathbf{s}) \approx \sum_{k=1}^{K} \psi_k \omega_k, \tag{6.6}$$

where K is the number of mesh nodes or vertices, $\{\psi_k\}$ are basis functions chosen to be piecewise linear in each triangle, i.e., $\psi_k = 1$ at vertex k and $\psi_k = 0$ otherwise, and $\{\omega_k\}$ are Gaussian-distributed weights. Using the SPDE approach, we can write Equation (6.5) as

$$\mathbf{h}(\boldsymbol{x}_1) = \mathbf{Z}_1\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\omega},$$

where the latent parameter vector is $\boldsymbol{x}_1 = \begin{pmatrix} \boldsymbol{\beta} & \omega_1 & \dots & \omega_K \end{pmatrix}^\intercal$, and $\mathbf{A}$ is the mapping matrix from the mesh nodes to the observed data points. With the full $\mathbf{Q}$ uncertainty method, the posterior mean $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1} = \begin{pmatrix} \hat{\boldsymbol{\beta}} & \hat{\boldsymbol{\omega}} \end{pmatrix}^\intercal$ and the precision matrix $\mathbf{Q}^{-1}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)$ from the Gaussian approximation in Equation (6.3) are used. The linear predictor in the second-stage model is then specified as:

$$g\Big(\mathbb{E}[\mathcal{D}_2|\cdots]\Big) = \gamma_0\mathbf{1} + \gamma_1\left\{ \begin{bmatrix} \mathbf{Z}_1 & \mathbf{A} \end{bmatrix} \left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\omega}} \end{bmatrix} + \boldsymbol{\epsilon} \right) \right\} + \mathbf{Z}_2\boldsymbol{\gamma}_2, \tag{6.7}$$

where γ_0, γ_1, and $\boldsymbol{\gamma}_2$ are the model parameters; $\mathbf{Z}_1$,$\mathbf{Z}_2$ and $\mathbf{A}$ are known matrices; and $\boldsymbol{\epsilon}$ is the error component with prior given by $\boldsymbol{\epsilon} \sim \mathcal{N}\Big(\mathbf{0}, \mathbf{Q}^{-1}_{\boldsymbol{x}_1}(\boldsymbol{\theta}_1)\Big)$.

6.1.2.2 Low rank Q uncertainty method

A potential problem with the specification in Equation (6.7) is that when extending this to large spatio-temporal data, the dimension of $\hat{\boldsymbol{\omega}}$ scales linearly as the number of time points, which in effect also applies to the dimension of the error component $\boldsymbol{\epsilon}$. Thus, I propose a low rank approximation of $\mathbf{Q}$, which then expresses Equation (6.7) as follows:

$$g\Big(\mathbb{E}[\mathcal{D}_2|\cdots]\Big) = \gamma_0\mathbf{1} + \gamma_1\left\{ \begin{bmatrix} \mathbf{Z}_1 & \mathbf{A} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} + \boldsymbol{\epsilon}_\beta \\ \hat{\boldsymbol{\omega}} + \mathbf{B}\boldsymbol{\epsilon}^*_\omega \end{bmatrix} \right\} + \mathbf{Z}_2\boldsymbol{\gamma}_2 \tag{6.8}$$

$$= \gamma_0\mathbf{1} + \gamma_1\left\{ \begin{bmatrix} \mathbf{Z}_1 & \mathbf{A} \end{bmatrix} \left(\begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\omega}} \end{bmatrix} + \begin{bmatrix} \mathbb{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \boldsymbol{\epsilon}_\beta \\ \boldsymbol{\epsilon}^*_\omega \end{bmatrix} \right) \right\} + \mathbf{Z}_2\boldsymbol{\gamma}_2, \tag{6.9}$$

where $\boldsymbol{\epsilon}$ is implicitly partitioned into $\begin{pmatrix} \boldsymbol{\epsilon}_\beta & \boldsymbol{\epsilon}_\omega \end{pmatrix}^\intercal$, with $\boldsymbol{\epsilon}_\beta$ being the fixed effects error component which is time-invariant, and $\boldsymbol{\epsilon}_\omega$ is the spatial error component which varies in time. In Equation (6.8), $\mathbf{B}\boldsymbol{\epsilon}^*_\omega$ is used to approximate $\boldsymbol{\epsilon}_\omega$, where $\boldsymbol{\epsilon}^*_\omega$ is defined on a coarser mesh compared to the mesh used for $\hat{\boldsymbol{\omega}}$, and $\mathbf{B}$ is the appropriate projection matrix from the coarse mesh to the fine mesh.

The probability model for $\boldsymbol{\epsilon}^*_\omega$ depends on the distribution of the weights at the coarser mesh, say $\boldsymbol{\phi} \in \mathbb{R}^M$, $M \ll K$, K being the dimension of $\hat{\boldsymbol{\omega}}$. The probability

model for $\boldsymbol{\epsilon}_\omega^*$ is given by $\boldsymbol{\epsilon}_\omega^* \sim \mathcal{N}\left(\mathbf{0}, \left(\mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \mathbf{B}\right)^{-1}\right)$ as stated in Theorem 6.1.

Theorem 6.1. *Suppose that $\boldsymbol{\omega}$ is defined on a discretization of dimension $\mathbb{R}^K$, i.e. $\boldsymbol{\omega} \in \mathbb{R}^K$, with probability model $\boldsymbol{\omega} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{Q}_{\boldsymbol{\omega}}^{-1}\right)$. Moreover, suppose that we define a coarser discretization specified via the weights $\boldsymbol{\phi} \in \mathbb{R}^M, M \ll K$, such that $\boldsymbol{\omega} = \mathbf{B}\boldsymbol{\phi}$. Then, the precision matrix of $\boldsymbol{\phi}$ is given by $\mathbf{Q}_{\boldsymbol{\phi}} = \mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \mathbf{B}$ and the probability model of $\boldsymbol{\phi}$ is given by*

$$\boldsymbol{\phi} \sim \mathcal{N}\left(\left(\mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \mathbf{B}\right)^{-1} \mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \boldsymbol{\omega}, \mathbf{Q}_{\boldsymbol{\phi}}^{-1}\right).$$

Proof. This is simply a generalized least squares problem, i.e., $\hat{\boldsymbol{\phi}} := \text{argmin}_{\boldsymbol{\phi}} \|\boldsymbol{\omega} - \mathbf{B}\boldsymbol{\phi}\|^2_{\mathbf{Q}_{\boldsymbol{\omega}}^{-1}}$. This yields $\mathbb{E}[\hat{\boldsymbol{\phi}}] = \left(\mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \mathbf{B}\right)^{-1} \mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \boldsymbol{\omega}$ and $\mathbb{V}[\hat{\boldsymbol{\phi}}] = \mathbf{Q}_{\boldsymbol{\phi}} = \left(\mathbf{B}^\intercal \mathbf{Q}_{\boldsymbol{\omega}} \mathbf{B}\right)^{-1}$. □

6.2 Simulation-based calibration for model validation

I validate the uncertainty propagation methods discussed in Sections 6.1.1 and 6.1.2 using the simulation-based calibration (SBC) approach, originally proposed by Talts et al. (2018) and based on ideas from Cook et al. (2006). The SBC method tests for the self-consistency property of Bayesian models, which states that the posterior distribution, averaged over all possible outcomes from the full generative model, is equal to the prior distribution. Formally, suppose $\pi(\boldsymbol{\theta})$ is the prior model, $\pi(\mathbf{y}|\boldsymbol{\theta})$ is the observation density or probability mass function, and $\pi(\boldsymbol{\theta}|\mathbf{y})$ is the posterior distribution. The self-consistency property is stated as:

$$\pi(\boldsymbol{\theta}') = \int \pi(\boldsymbol{\theta}'|\mathbf{y})\pi(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) d\mathbf{y} d\boldsymbol{\theta}.$$

Any discrepancy between the prior model and the data-averaged posterior indicates an error in the Bayesian algorithm. The SBC method tests this property using rank statistics. It involves sampling from the data's generative model and applying the Bayesian algorithm to each data replicate. Specifically, consider the following

sequence of samples drawn from the Bayesian model:

$$\begin{aligned}\tilde{\boldsymbol{\theta}} &\sim \pi(\boldsymbol{\theta}) \\ \tilde{\mathbf{y}} &\sim \pi(\mathbf{y}|\tilde{\boldsymbol{\theta}}) \\ \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_L\} &\overset{iid}{\sim} \pi(\boldsymbol{\theta}|\tilde{\mathbf{y}}).\end{aligned}$$

If the Bayesian algorithm is correct, then for any one-dimensional function of the parameters, $f : \Theta \to \mathbb{R}$, where Θ is the $\boldsymbol{\theta}$-space, the *rank statistic* of the prior sample relative to the posterior sample, given by

$$r\Big(\{f(\boldsymbol{\theta}_1), \ldots, f(\boldsymbol{\theta}_L)\}, f(\tilde{\boldsymbol{\theta}})\Big) = \sum_{\ell=1}^{L} \mathbb{I}[f(\boldsymbol{\theta}_\ell) < f(\tilde{\boldsymbol{\theta}})], \quad \mathbb{I}[f(\boldsymbol{\theta}_\ell) < f(\tilde{\boldsymbol{\theta}})] = \begin{cases} 1 & \text{if } f(\boldsymbol{\theta}_\ell) < f(\tilde{\boldsymbol{\theta}}) \\ 0 & \text{if } f(\boldsymbol{\theta}_\ell) \geq f(\tilde{\boldsymbol{\theta}}) \end{cases} \tag{6.10}$$

should be uniformly distributed across the integers $\{0, 1, \ldots, L\}$.

Deviations from uniformity in the rank distribution provide insights into errors in the posteriors. A $\cap$-shaped rank distribution suggests that the data-averaged posterior is overdispersed compared to the prior, meaning uncertainty is overestimated, violating Bayesian self-consistency. Conversely, a $\cup$-shaped rank distribution indicates underdispersion, where the estimated posterior underestimates the true uncertainty. Asymmetry in the rank distribution reveals bias in the data-averaged posterior, deviating in the opposite direction relative to the prior distribution.

6.2.1 Implementation of SBC for the two-stage model

This section discusses how to implement the SBC in a two-stage modelling framework following Figure 6.1. I assume that $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}_2, \boldsymbol{x}_1 \in \boldsymbol{\chi}_1$, and $\boldsymbol{x}_2 \in \boldsymbol{\chi}_2$, and that $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2, \boldsymbol{\chi}_1$, and $\boldsymbol{\chi}_2$ are continuous spaces. The assumption of continuous spaces for the model parameters are crucial for the SBC method in Talts et al. (2018) to work. For cases when some of the spaces are discrete, an SBC variant is proposed in Modrák et al. (2023).

Following Equations (6.1), (6.7), and (6.8), the main interest is in the posterior marginals of γ_0 and γ_1, since these are the parameters whose posterior uncertainty are potentially underestimated when the uncertainty in the first stage model is not

properly propagated to the second stage. I use the individual parameters as test quantities to check that their uncertainty is correctly calibrated. Testing individual parameters allows the diagnosis of a large number of problems with posterior approximation (Modrák et al., 2023). However, Modrák et al. (2023) also recommended the use of test functions which are data-dependent, such as the joint likelihood of the data, since there are large classes of problems which cannot be detected when the test quantities are functions only of the parameters. I have not considered such test functions in this work, which can be done in a future work.

Algorithm 6.2 shows the steps to implement the SBC for the two-stage Bayesian model in Figure 6.1, where the test quantities are γ_0 and γ_1. To test for the uniformity of the rank statistic in Equation (6.10), I primarily use the graphical approach in Säilynoja et al. (2022), which generates simultaneous confidence bands for the difference between the empirical cumulative distribution function (ECDF) and the uniform CDF. The method is not sensitive to binning, does not require smoothing, and provides intuitive visual interpretation.

6.2.2 Variation in the SBC

In this section, I propose a variation of the SBC approach in a two-stage modeling framework. The motivation here is that the parameters of primary interest are the fixed effects of the second-stage model, namely γ_0 and γ_1, and I want to avoid the influence of certain parameters of the first-stage model which may violate the self-consistency property for specific priors or model specification. As an example, some parameters in $\boldsymbol{\theta}_{1,\boldsymbol{x}}$ (Figure 6.1) may violate the self-consistency property. Hence, we propose Theorem 6.2, which derives the distribution of the rank statistic for an arbitrary unidimensional test function conditional on $\boldsymbol{\theta}_{1,\boldsymbol{x}}$. This is then used as the theoretical basis for doing the SBC conditional on $\boldsymbol{\theta}_{1,\boldsymbol{x}}$.

Theorem 6.2. *Let $\pi(\boldsymbol{\theta}_1) = \pi(\boldsymbol{\theta}_{1,\boldsymbol{x}})\pi(\boldsymbol{\theta}_{1,\mathcal{D}})$ be the prior model, $\pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$ be the latent model, and $\pi(\mathcal{D}_1|\boldsymbol{x}_1, \boldsymbol{\theta}_1)$ be the observation density or probability mass function. Let $\boldsymbol{\chi}_1$ be the latent space, $\boldsymbol{\Theta}_{1,\boldsymbol{x}}$ be the $\boldsymbol{\theta}_{1,\boldsymbol{x}}$-space, and $\boldsymbol{\Theta}_{1,\mathcal{D}}$ be the $\boldsymbol{\theta}_{1,\mathcal{D}}$-space, all continuous. Suppose $\boldsymbol{\theta}_{1,\boldsymbol{x}} \in \boldsymbol{\Theta}_{1,\boldsymbol{x}}$ is fixed. Let $\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}}$ be a sample from the prior, i.e., $\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}} \sim \pi(\boldsymbol{\theta}_{1,\mathcal{D}})$, $\tilde{\boldsymbol{x}}_1$ be a sample from the latent model, i.e., $\tilde{\boldsymbol{x}}_1 \sim \pi(\boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}, \tilde{\boldsymbol{\theta}}_{1,\mathcal{D}})$, and $\tilde{\mathcal{D}}_1$*

Algorithm 6.2 Implementing SBC for Figure 6.1 with γ_0 and γ_1 as test quantities

Do for $k = 1, 2, \ldots, K$:

Step 1: Sample hyperparameter values: $\tilde{\boldsymbol{\theta}}_1^{(k)} \sim \pi(\boldsymbol{\theta}_1), \tilde{\boldsymbol{\theta}}_2^{(k)} \sim \pi(\boldsymbol{\theta}_2)$.

Step 2: Sample latent parameter values: $\tilde{\boldsymbol{x}}_1^{(k)} \sim \pi(\boldsymbol{x}_1|\tilde{\boldsymbol{\theta}}_1^{(k)}), \tilde{\boldsymbol{x}}_2^{(k)} \sim \pi(\boldsymbol{x}_2|\tilde{\boldsymbol{\theta}}_2^{(k)})$.

Step 3: Sample observed data values:

$$\tilde{\boldsymbol{\mathcal{D}}}_1^{(k)} \sim \pi_1(\boldsymbol{\mathcal{D}}_1|\tilde{\boldsymbol{x}}_1^{(k)}, \tilde{\boldsymbol{\theta}}_1^{(k)}), \quad \tilde{\boldsymbol{\mathcal{D}}}_2^{(k)} \sim \pi_2(\boldsymbol{\mathcal{D}}_2|\tilde{\boldsymbol{x}}_1^{(k)}, \tilde{\boldsymbol{x}}_2^{(k)}, \tilde{\boldsymbol{\theta}}_2^{(k)}).$$

Step 4: Perform inference in order to obtain estimated posteriors: $\hat{\pi}^{(k)}(\boldsymbol{\theta}_1, \boldsymbol{x}_1|\boldsymbol{\mathcal{D}}_1)$ and $\hat{\pi}^{(k)}(\boldsymbol{\theta}_2, \boldsymbol{x}_2|\boldsymbol{\mathcal{D}}_2)$.

Step 5: Generate L samples from the estimated posterior distributions of γ_0 and γ_1:

$$\gamma_{0,1}^{(k)}, \gamma_{0,2}^{(k)}, \ldots, \gamma_{0,L}^{(k)} \sim \hat{\pi}(\gamma_0|\boldsymbol{\mathcal{D}}_2)$$
$$\gamma_{1,1}^{(k)}, \gamma_{1,2}^{(k)}, \ldots, \gamma_{1,L}^{(k)} \sim \hat{\pi}(\gamma_1|\boldsymbol{\mathcal{D}}_2)$$

Step 6: Compute the ranks:

$$r\Big(\big\{\gamma_{0,1}^{(k)}, \gamma_{0,2}^{(k)}, \ldots, \gamma_{0,L}^{(k)}\big\}, \tilde{\gamma}_0^{(k)}\Big) = \sum_{\ell=1}^{L} \mathbb{I}\big[\gamma_{0,\ell}^{(k)} < \tilde{\gamma}_0^{(k)}\big], \quad \mathbb{I}\big[\gamma_{0,\ell}^{(k)} < \tilde{\gamma}_0^{(k)}\big] = \begin{cases} 1 & \gamma_{0,\ell}^{(k)} < \tilde{\gamma}_0^{(k)} \\ 0 & \gamma_{0,\ell}^{(k)} \geq \tilde{\gamma}_0^{(k)} \end{cases}$$
$$r\Big(\big\{\gamma_{1,1}^{(k)}, \gamma_{1,2}^{(k)}, \ldots, \gamma_{1,L}^{(k)}\big\}, \tilde{\gamma}_1^{(k)}\Big) = \sum_{\ell=1}^{L} \mathbb{I}\big[\gamma_{1,\ell}^{(k)} < \tilde{\gamma}_1^{(k)}\big], \quad \mathbb{I}\big[\gamma_{1,\ell}^{(k)} < \tilde{\gamma}_1^{(k)}\big] = \begin{cases} 1 & \gamma_{1,\ell}^{(k)} < \tilde{\gamma}_1^{(k)} \\ 0 & \gamma_{1,\ell}^{(k)} \geq \tilde{\gamma}_1^{(k)} \end{cases},$$

where $\tilde{\gamma}_0^{(k)}$ and $\tilde{\gamma}_1^{(k)}$ are prior samples. The ranks are normalized by computing

$$p_k = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}\big[\gamma_{0,\ell}^{(k)} < \tilde{\gamma}_0^{(k)}\big] \quad \text{and} \quad p_k = \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{I}\big[\gamma_{1,\ell}^{(k)} < \tilde{\gamma}_1^{(k)}\big] \tag{6.11}$$

---

a sample from the observation model, i.e., $\tilde{\boldsymbol{\mathcal{D}}}_1 \sim \pi(\boldsymbol{\mathcal{D}}_1|\tilde{\boldsymbol{x}}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}, \tilde{\boldsymbol{\theta}}_{1,\mathcal{D}})$. *Suppose that the approximate posteriors from applying the Bayesian algorithm are* $\hat{\pi}(\boldsymbol{x}_1|\tilde{\boldsymbol{\mathcal{D}}}_1)$ *and* $\hat{\pi}(\boldsymbol{\theta}_1|\tilde{\boldsymbol{\mathcal{D}}}_1)$. *Let* $\{\boldsymbol{x}_{1,\ell}\}$ *and* $\{\boldsymbol{\theta}_{1,\mathcal{D},\ell}\}, \ell = 1, \ldots, L$ *be independent samples from the posteriors, i.e.,* $\begin{pmatrix} \boldsymbol{x}_{1,1} & \boldsymbol{x}_{1,2} & \ldots \boldsymbol{x}_{1,L} \end{pmatrix} \overset{iid}{\sim} \hat{\pi}(\boldsymbol{x}_1|\tilde{\boldsymbol{\mathcal{D}}}_1)$, *and* $\begin{pmatrix} \boldsymbol{\theta}_{1,\mathcal{D},1} & \boldsymbol{\theta}_{1,\mathcal{D},2} & \ldots \boldsymbol{\theta}_{1,\mathcal{D},L} \end{pmatrix} \overset{iid}{\sim} \hat{\pi}(\boldsymbol{\theta}_{1,\mathcal{D}}|\tilde{\boldsymbol{\mathcal{D}}}_1)$. *Then, we have the following results:*

(1) For any uni-dimensional function $f : \boldsymbol{\chi}_1 \to \mathbb{R}$, *the distribution of the rank statistic*

$$r = \sum_{\ell=1}^{L} \mathbb{I}\big[f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1)\big], \quad \mathbb{I}\big[f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1)\big] = \begin{cases} 1 & \text{if } f(\mathbf{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1) \\ 0 & \text{if } f(\boldsymbol{x}_{1,\ell}) \geq f(\tilde{\boldsymbol{x}}_1) \end{cases}$$

is $\mathcal{U}(0, 1, \ldots, L)$.

(2) For any uni-dimensional function $f : \boldsymbol{\Theta}_{1,\mathcal{D}} \to \mathbb{R}$*, the distribution of the rank statistic*

$$r = \sum_{\ell=1}^{L} \mathbb{I}\big[f(\boldsymbol{\theta}_{1,\mathcal{D},\ell}) < f(\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}})\big], \quad \mathbb{I}\big[f(\boldsymbol{\theta}_{1,\mathcal{D},\ell}) < f(\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}})\big] = \begin{cases} 1 & \text{if } f(\boldsymbol{\theta}_{1,\mathcal{D},\ell}) < f(\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}}) \\ 0 & \text{if } f(\boldsymbol{\theta}_{1,\mathcal{D},\ell}) \geq f(\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}}) \end{cases}$$

is $\mathcal{U}(0, 1, \ldots, L)$.

Proof of Theorem 6.2 (1):

Proof. Let $f_\ell \equiv f(\boldsymbol{x}_{1,\ell})$ and $f \equiv f(\tilde{\boldsymbol{x}}_1)$. Also, let $\pi(f)$ and $\pi(f|\mathcal{D}_1)$ be the pushforward probability density function of $\pi(\boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}})$ and $\pi(\boldsymbol{x}_1|\mathcal{D}_1)$, respectively. Suppose $p_\ell = \mathbb{P}(f_\ell < f), \ell = 1, \ldots, L$. Also, we assume the ordering $f_1 \leq f_2 \leq \cdots \leq f_L$. We then have:

$$\begin{aligned}
\pi(r) &= \int d\boldsymbol{\theta}_{1,\mathcal{D}} df d\mathcal{D}_1 \pi(\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \binom{L}{r} \prod_{\ell=1}^{r} p_\ell \prod_{\ell=r+1}^{L} (1 - p_\ell) \\
&= \binom{L}{r} \int d\boldsymbol{\theta}_{1,\mathcal{D}} df d\mathcal{D}_1 \pi(\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \prod_{\ell=1}^{r} \left[\int_{-\infty}^{f} \pi(f_\ell|\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df_\ell \right] \times \\
&\qquad \prod_{\ell=r+1}^{L} \left[\int_{f}^{\infty} \pi(f_\ell|\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df_\ell \right]
\end{aligned}$$

The probability measure for generating f_ℓ depends only on $\mathcal{D}_1$ and is independent of the conditioning model configuration. Hence, we can write $\pi(f_\ell|\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}, \boldsymbol{\theta}_{1,\boldsymbol{x}}) = \pi(f_\ell|\mathcal{D}_1) = \pi(f_\ell|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})$. Further, since the model used to simulate data and construct posterior distributions is the same, then we have $\pi(f_\ell|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) = \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}), \ell = 1, \ldots, L$. This implies that

$$\pi(r) = \binom{L}{r} \int d\boldsymbol{\theta}_{1,\mathcal{D}} df d\mathcal{D}_1 \pi(\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \left[\int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^r \left[1 - \int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^{L-r}$$

We use the fact that

$$\begin{aligned}
\pi(\mathcal{D}_1, f, \boldsymbol{\theta}_{1,\mathcal{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) &= \pi(f, \boldsymbol{\theta}_{1,\mathcal{D}}|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \pi(\mathcal{D}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) && (6.12) \\
&= \pi(f|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \pi(\boldsymbol{\theta}_{1,\mathcal{D}}|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \pi(\mathcal{D}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) && (6.13) \\
&= \pi(f|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \pi(\mathcal{D}_1, \boldsymbol{\theta}_{1,\mathcal{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}). && (6.14)
\end{aligned}$$

Equation (6.13) is true since f and $\boldsymbol{\theta}_{1,\mathcal{D}}$ are d-separated given $\mathcal{D}_1$. This yields

$$\pi(r) = \binom{L}{r} \int d\boldsymbol{\theta}_{1,\mathcal{D}} df d\mathcal{D}_1 \pi(f|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \pi(\mathcal{D}_1, \boldsymbol{\theta}_{1,\mathcal{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \left[\int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^r \times$$

$$\left[1 - \int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df'\right]^{L-r}$$

$$= \binom{L}{r} \int d\boldsymbol{\theta}_{1,\mathfrak{D}} d\mathfrak{D}_1 \pi(\mathfrak{D}_1, \boldsymbol{\theta}_{1,\mathfrak{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \int df \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \left[\int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df'\right]^{r} \times$$

$$\left[1 - \int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df'\right]^{L-r}$$

Let $u = \int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df'$, so that $du = \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df$. Thus,

$$\begin{aligned}
\pi(r) &= \binom{L}{r} \int d\boldsymbol{\theta}_{1,\mathfrak{D}} d\mathfrak{D}_1 \pi(\mathfrak{D}_1, \boldsymbol{\theta}_{1,\mathfrak{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \int du (u)^r (1-u)^{L-r} \\
&= \binom{L}{r} B(r+1, L-r+1) \\
&= \frac{1}{L+1}
\end{aligned}$$

□ □

Proof of Theorem 6.2 (2):

Proof. Let $f_\ell \equiv f(\boldsymbol{\theta}_{1,\mathfrak{D},\ell})$ and $f \equiv f(\tilde{\boldsymbol{\theta}}_{1,\mathfrak{D}})$. Also, let $\pi(f)$ and $\pi(f|\mathfrak{D}_1)$ be the pushforward probability density function of $\pi(\boldsymbol{\theta}_{1,\mathfrak{D}}|\boldsymbol{\theta}_{1,\boldsymbol{x}}) = \pi(\boldsymbol{\theta}_{1,\mathfrak{D}})$ and $\pi(\boldsymbol{\theta}_{1,\mathfrak{D}}|\mathfrak{D}_1)$, respectively. Suppose $p_\ell = \mathbb{P}(f_\ell < f), \ell = 1, \ldots, L$. Also, we assume the ordering $f_1 \leq f_2 \leq \cdots \leq f_L$.

We can write the density of the rank statistic as

$$\pi(r) = \int df d\boldsymbol{x}_1 d\mathfrak{D}_1 \pi(\mathfrak{D}_1, \boldsymbol{x}_1, f|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \binom{L}{r} \prod_{\ell=1}^{r} p_\ell \prod_{\ell=r+1}^{L} (1-p_\ell).$$

We use the fact that

$$\begin{aligned}
\pi(\mathfrak{D}_1, \boldsymbol{x}_1, f|\boldsymbol{\theta}_{1,\boldsymbol{x}}) &= \pi(f, \boldsymbol{x}_1|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})\pi(\mathfrak{D}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) && (6.15) \\
&= \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})\pi(\boldsymbol{x}_1|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})\pi(\mathfrak{D}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) && (6.16) \\
&= \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})\pi(\mathfrak{D}_1, \boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) && (6.17)
\end{aligned}$$

Equation (6.16) is true since f and $\boldsymbol{x}_1$ are d-separated given $\mathfrak{D}_1$.

This yields

$$\pi(r) = \int \binom{L}{r} df d\boldsymbol{x}_1 d\mathfrak{D}_1 \pi(f|\mathfrak{D}_1)\pi(\mathfrak{D}_1, \boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \prod_{\ell=1}^{r} \left[\int_{-\infty}^{f} \pi(f_\ell|\mathfrak{D}_1, f, \boldsymbol{x}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df_\ell\right] \times$$
$$\prod_{\ell=r+1}^{L} \left[\int_{f}^{\infty} \pi(f_\ell|\mathfrak{D}_1, f, \boldsymbol{x}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}})df_\ell\right]$$

Similar to the argument in Result 1, we have $\pi(f_\ell|\mathfrak{D}_1, f, \boldsymbol{x}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) = \pi(f_\ell|\mathfrak{D}_1) = \pi(f_\ell|\mathfrak{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}), \ell =$

$1, \ldots, L$. Also, we have $\pi(f_\ell|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) = \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}), \ell = 1, \ldots, L$. This yields

$$\pi(r) = \binom{L}{r} \int df d\boldsymbol{x}_1 d\mathcal{D}_1 \pi(f|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \pi(\mathcal{D}_1, \boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \prod_{\ell=1}^{r} \left[\int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^r \times \left[1 - \int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^{L-r}$$

$$= \binom{L}{r} \int d\boldsymbol{x}_1 d\mathcal{D}_1 \pi(\mathcal{D}_1, \boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \int df \pi(f|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) \left[\int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^r \times \left[1 - \int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df' \right]^{L-r}$$

Let $u = \int_{-\infty}^{f} \pi(f'|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df'$, so that $du = \pi(f|\mathcal{D}_1, \boldsymbol{\theta}_{1,\boldsymbol{x}}) df$. Thus,

$$\pi(r) = \binom{L}{r} \int d\boldsymbol{x}_1 d\mathcal{D}_1 \pi(\mathcal{D}_1, \boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}) \int du (u)^r (1-u)^{L-r} = \binom{L}{r} B(r+1, L-r+1) = \frac{1}{L+1}$$

□

□

Theorem 6.2 implies a variation in the implementation of the original SBC. Instead of sampling from the full data generative model, we fix the value of $\boldsymbol{\theta}_{1,\boldsymbol{x}}$. In particular, the changes in Algorithm 6.2 only apply to Steps 1 – 3, which are the steps for generating data from the model. The changes are formalized in Algorithm 6.3. The remaining steps for doing the SBC conditional on $\boldsymbol{\theta}_{1,\boldsymbol{x}}$ are the same as the original SBC, i.e., the model inference is done without knowledge of $\boldsymbol{\theta}_{1,\boldsymbol{x}}$ and the test quantities for the SBC are γ_0 and γ_1. Moreover, I extend Theorem 6.2 to the case where we condition on the entire hyperaparameter vector $\boldsymbol{\theta}_1$. This is formalized in Theorem C.1 in Appendix C.

Algorithm 6.3 Data generation mechanism for the SBC conditional on $\boldsymbol{\theta}_{1,\boldsymbol{x}}$

Fix $\boldsymbol{\theta}_{1,\boldsymbol{x}} \in \Theta_{1,\boldsymbol{x}}$. Do for $k = 1, 2, \ldots, K$:

Step 1: Sample hyperparameter values: $\tilde{\boldsymbol{\theta}}_{1,\mathcal{D}}^{(k)} \sim \pi(\boldsymbol{\theta}_{1,\mathcal{D}}), \tilde{\boldsymbol{\theta}}_2^{(k)} \sim \pi(\boldsymbol{\theta}_2)$.

Step 2: Sample latent parameter values: $\tilde{\boldsymbol{x}}_1^{(k)} \sim \pi\big(\boldsymbol{x}_1|\boldsymbol{\theta}_{1,\boldsymbol{x}}\big), \tilde{\boldsymbol{x}}_2^{(k)} \sim \pi\big(\boldsymbol{x}_2|\tilde{\boldsymbol{\theta}}_2^{(k)}\big)$.

Step 3: Sample observed data values:

$$\tilde{\mathcal{D}}_1^{(k)} \sim \pi_1(\mathcal{D}_1|\tilde{\boldsymbol{x}}_1^{(k)}, \tilde{\boldsymbol{\theta}}_{1,\mathcal{D}}^{(k)}, \boldsymbol{\theta}_{1,\boldsymbol{x}}), \tilde{\mathcal{D}}_2^{(k)} \sim \pi_2(\mathcal{D}_2|\tilde{\boldsymbol{x}}_1^{(k)}, \tilde{\boldsymbol{x}}_2^{(k)}, \tilde{\boldsymbol{\theta}}_2^{(k)}).$$

6.3 Simulation experiments

This section presents the simulation experiments which compare the different uncertainty propagation approaches on two two-stage models: one with Gaussian observations (Section 6.3.1) and one with Poisson observations (Section 6.3.2). For each model, we highlight the SBC results for γ_0 and γ_1 using both Algorithms 6.2 and 6.3.

6.3.1 A two-stage spatial model with Gaussian likelihood

In the first experiment, the first-stage latent process $\mu(\mathbf{s})$ is given by $\mu(\mathbf{s}) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}) + \xi(\mathbf{s})$, where β_0 and β_1 are fixed effects, $\mathrm{z}(\mathbf{s})$ is a known covariate, and $\xi(\mathbf{s})$ is a Gaussian field with Matérn covariance function. The error-prone observed outcomes are $\mathcal{D}_1 \equiv \{\mathrm{w}(\mathbf{s}_i), i = 1, \ldots, n_{\mathrm{w}}\}$, which follows the classical error model, i.e.,

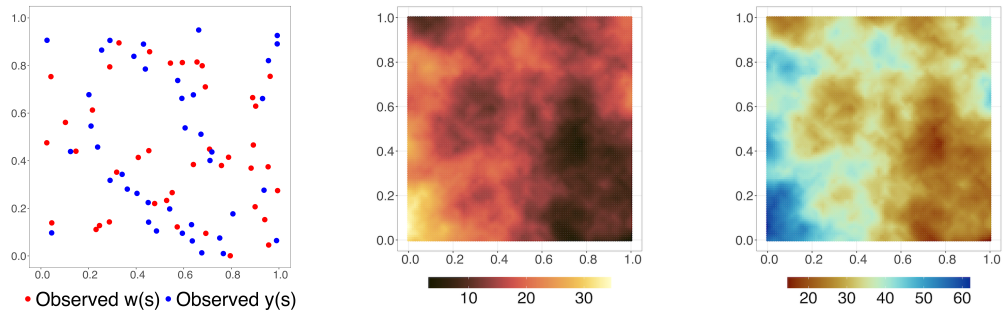
$$\mathrm{w}(\mathbf{s}_i) = \mu(\mathbf{s}_i) + e_1(\mathbf{s}_i), \quad e_1(\mathbf{s}_i) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_1}\big), \quad i = 1, \ldots, n_{\mathrm{w}} \tag{6.18}$$

The latent process $\mu(\mathbf{s})$ is an input in the second-stage model, i.e.,

$$\mathrm{y}(\mathbf{s}_j) = \gamma_0 + \gamma_1 \mu(\mathbf{s}_j) + e_2(\mathbf{s}_j), \quad e_2(\mathbf{s}_j) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_{e_2}\big), \quad j = 1, \ldots, n_{\mathrm{y}}, \tag{6.19}$$

Here, $\mathcal{D}_2 \equiv \{\mathrm{y}(\mathbf{s}_j), j = 1, \ldots, n_{\mathrm{y}}\}$ represents observations at locations different from where $\mathrm{w}(\mathbf{s}_i)$ is measured. This model belongs to the class of measurement error models because $\mu(\mathbf{s}_j)$ in Equation (6.19) is unobserved and needs to be estimated using Equation (6.18) (Banerjee and Gelfand, 2002; Berry et al., 2002; Madsen et al., 2008). Figure 6.2 shows a simulated observed locations for $\mathrm{w}(\mathbf{s}_i)$ and $\mathrm{y}(\mathbf{s}_j)$ where $n_{\mathrm{w}} = n_{\mathrm{y}} = 80$. Note that $\mathbb{E}\big[\mathrm{y}(\mathbf{s}) | \mu(\mathbf{s})\big] = \gamma_0 + \gamma_1 \mu(\mathbf{s})$ is also another latent field of interest.

I used INLA with SPDE representations of the spatial fields to fit this model. The first- and second-stage latent parameters are $\boldsymbol{x}_1 = \{\beta_0, \beta_1, \omega_1, \omega_2, \ldots, \omega_K\}$ and $\boldsymbol{x}_2 = \{\gamma_0, \gamma_1\}$, respectively. The $\{\omega_1, \omega_2, \ldots, \omega_K\}$ are the Gaussian weights of the SPDE approximation (Lindgren et al., 2011), i.e., $\xi(\mathbf{s}) \approx \sum_{i=1}^{K} \psi_i \omega_i$ where $\psi_i, i = 1, \ldots, K$, are basis functions as discussed in Section 6.1.2. Moreover, the first-stage hyperpa-



(a) Spatial locations of data (b) $\mu(\mathbf{s})$ (c) $\mathbb{E}(\mathrm{y}(\mathbf{s})|\mu(\mathbf{s}))$

Figure 6.2: Simulated data for the two-stage model in Section 6.3.1: (a) spatial locations of the data (b) simulated $\mu(\mathbf{s})$ (c) simulated second-stage field $\mathbb{E}[\mathrm{y}(\mathbf{s})|\mu(\mathbf{s})]$

rameters are $\boldsymbol{\theta}_1 = \{\sigma_{e_1}, \rho_\xi, \sigma_\xi\}$, where σ_ξ^2 and ρ_ξ are the marginal variance and range parameter of the random field $\xi(\mathbf{s})$, respectively. The second-stage hyperparameter is $\boldsymbol{\theta}_2 = \{\sigma_{e_2}\}$.

I used Gaussian priors for the fixed effects: $\beta_0 \sim \mathcal{N}(0, 10^2), \beta_1 \sim \mathcal{N}(0, 5^2), \gamma_0 \sim \mathcal{N}(0, 10^2), \gamma_1 \sim \mathcal{N}(0, 3^2)$; and penalized-complexity (PC) prior for σ_{e_1} and σ_{e_2} (Fuglstad et al., 2019; Simpson et al., 2017)). A PC prior for σ_{e_1} and σ_{e_2} penalizes deviation from the base model of zero variance. The formulation requires the user to specify a constant σ_0 and a probability value α such that $\mathbb{P}(\sigma_{e_i} > \sigma_0) = \alpha, i = 1, 2$. This is equivalent to the prior $\sigma_{e_i} \sim \text{Exp}\Big(\lambda = -(\ln \alpha)/\sigma_0\Big)$, where the rate parameter λ determines the magnitude of the penalty, with higher values corresponding to higher penalty. In particular, the PC priors for σ_{e_1} and σ_{e_2} are specified as $\mathbb{P}(\sigma_{e_1} > 1) = 0.5$ and $\mathbb{P}(\sigma_{e_2} > 1) = 0.5$. For the Matérn parameters, I used a joint normal prior for $\log(\tau)$ and $\log(\kappa)$ (Lindgren and Rue, 2015), where

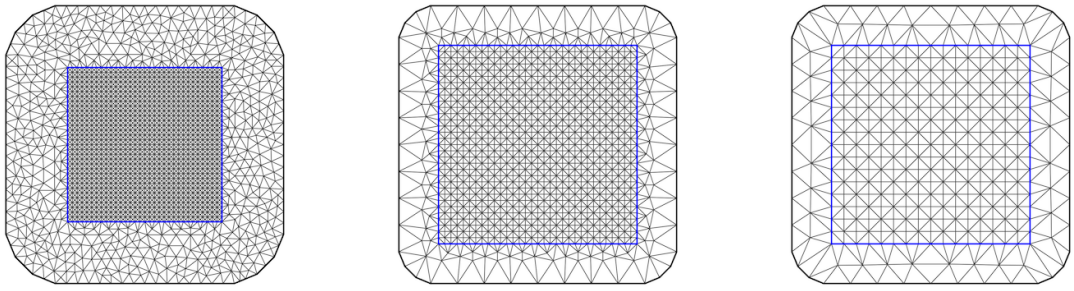
$$\log(\tau) = \frac{1}{2}\log\left(\frac{1}{4\pi}\right) - \log(\sigma_\xi) - \log(\kappa) \tag{6.20}$$

$$\log(\kappa) = \frac{\log(8)}{2} - \log(\rho_\xi). \tag{6.21}$$

In particular, this is given by $\begin{bmatrix}\log(\tau)\\ \log(\kappa)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}-0.7547 - \log(\kappa)\\ 1.0397\end{bmatrix}, \begin{bmatrix}20.67 & 0\\ 0 & 8.67\end{bmatrix}\right)$. This prior specification implies that the plausible values for the Matérn parameters are $0.3 \leq \sigma_\xi \leq 0.8$ and $0.6 \leq \rho_\xi \leq 1.5$. A joint Gaussian prior is used for the Matérn parameters because it simplifies the implementation of the SBC. The same prior is used across all Bayesian algorithms considered, as the primary objective of

this section is to validate the different algorithms using SBC.

In doing the SBC, I fixed the spatial locations of $w(\mathbf{s}_i)$ and $y(\mathbf{s}_j)$ for all the data replicates, which is shown in Figure 6.2a. The covariate field $z(\mathbf{s})$ was simulated from a Matérn process with range of 0.6, standard deviation of 2, and mean-squared differentiability parameter of 1. This is fixed for all data replicates since $z(\mathbf{s})$ is a known quantity in the model. The random field $\xi(\mathbf{s})$ was also simulated from a Matérn process, with range $\rho_\xi = 4$ and marginal standard deviation $\sigma_\xi = 0.6$. It varies for the different data replicates since this is an unknown quantity which needs to be estimated.



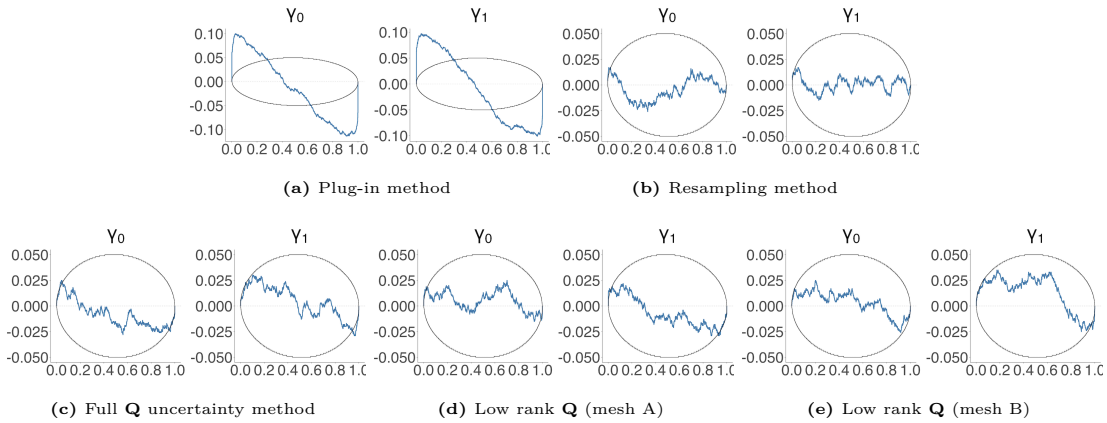
(a) Full **Q** mesh **(b)** Low rank **Q** mesh A **(c)** Low rank **Q** mesh B

Figure 6.3: Meshes used for the simulation experiments: (a) mesh for the full **Q** (b) slightly coarser mesh for the low rank **Q** method (c) very coarse mesh for the low rank **Q** method

Four uncertainty propagation approaches are compared: plug-in, resampling (with $J = 30$ resamples), full **Q**, and low rank **Q** uncertainty. The mesh used to fit the full **Q** uncertainty method is shown in Figure 6.3a. Here, the maximum triangle edge lengths for the inner domain and the outer extension are 0.04 and 0.1 units, respectively. I used the same mesh to simulate and estimate $\xi(\mathbf{s})$. For the low rank **Q** approach, I explored two meshes (A and B) with different level of coarseness: for mesh A (Figure 6.3b), the maximum edge lengths are 0.05 and 0.2 units, while for the coarser mesh B (Figure 6.3c) (low rank **Q** mesh B), they are 0.1 and 0.25, respectively. The number of nodes is 2672 for the full **Q** mesh, 1228 for mesh A, and 365 for mesh B.

Figure 6.4 shows the ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 from 1000 data replicates using Algorithm 6.2 (corresponding histograms are in Figure C.7 of Appendix C). Results show that for the plug-in method, the

hypotheses of uniform distribution of the ranks p_k is rejected for both γ_0 and γ_1 (Figure 6.4a). In addition, the ∪-shaped histogram (Figure C.7 of Appendix C) reveals an underestimation of the posterior uncertainty. The resampling method and the two proposed methods do not show deviations from uniformity, not even with the low rank $\mathbf{Q}$ approach using the coarser mesh B. This suggests that both the resampling approach and the proposed methods correctly capture the posterior uncertainty of γ_0 and γ_1.



(a) Plug-in method **(b)** Resampling method

(c) Full $\mathbf{Q}$ uncertainty method **(d)** Low rank $\mathbf{Q}$ (mesh A) **(e)** Low rank $\mathbf{Q}$ (mesh B)

Figure 6.4: ECDF difference plot of p_k for γ_0 and γ_1 using Algorithm 6.2 out of 1000 data replicates for the two-stage Gaussian spatial model (Section 6.3.1) using different approaches: (a) plug-in method (b) resampling method (c) full $\mathbf{Q}$ method (d) low rank $\mathbf{Q}$ (mesh A) (e) low rank $\mathbf{Q}$ (mesh B)

Although the primary focus is on the second-stage model parameters, I also examined the histogram and ECDF difference plot of the normalized ranks p_k for all first-stage model parameters. Section C.1.1 in Appendix C presents the SBC results for the first-stage model using Algorithm 6.2. The results show a uniform distribution of p_k for all first-stage model hyperparameters, except for σ_ξ which is slightly ∩-shaped. This motivates the use of Algorithm 6.3, based on Theorem 6.2, where SBC is applied conditional on $\boldsymbol{\theta}_{1,\boldsymbol{x}} = \{\sigma_\xi, \rho_\xi\}$. Moreover, three mesh nodes fail the Kolmogorov-Smirnov test for uniformity at 10% significance level (see Figure C.1 in Section C.1.1 of Appendix C).

The results from Algorithm 6.3 are shown in Sections C.1.2 and C.1.4 in Appendix C. The conclusions are consistent with those from Algorithm 6.2, i.e., the plug-in method underestimates the posterior uncertainty of γ_0 and γ_1, while the resampling method and the proposed methods are also correct, although there may be slight underestimation with the low rank $\mathbf{Q}$ approach. Finally, I also attempted to perform

the SBC on a non-spatial two-stage model, i.e., similar to Equations (6.18) and (6.19) but without the spatial field $\xi(\mathbf{s})$. I used both the INLA and no U-turn sampler (NUTS) (Hoffman et al., 2014). The results, which are given in Section C.3.1 of Appendix C, also show that the plug-in method underestimates the posterior uncertainty of both γ_0 and γ_1, while the resampling and the $\mathbf{Q}$ methods are correct. Note that the $\mathbf{Q}$ method is implemented only using the INLA method.

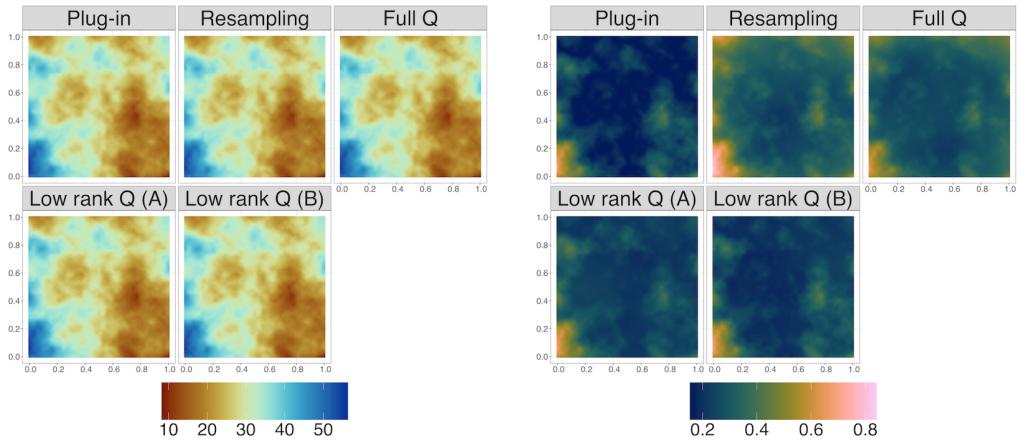
6.3.1.1 Illustration with simulated data

To gain additional insights, I analyze in detail one simulated data from the previous section with true values of the parameters as follows: $\beta_0 = 10, \beta_1 = 3, \gamma_0 = 10, \gamma_1 = 1.5, \sigma^2_{e_1} = 1, \sigma^2_{e_2} = 1, \sigma_\xi = 4$, and $\rho_\xi = 0.6$. Figure 6.2 shows: (a) the spatial locations for the data $\mathrm{w}(\mathbf{s}_i)$ and $\mathrm{y}(\mathbf{s}_j)$, (b) the simulated field $\mu(\mathbf{s})$, and (c) the simulated field $\mathbb{E}\big[\mathrm{y}(\mathbf{s})|\mu(\mathbf{s})\big]$, which we estimate using the different uncertainty propagation approaches.

The posterior means of $\mathbb{E}\big[\mathrm{y}(\mathbf{s})|\mu(\mathbf{s})\big]$ (Figure 6.5a) are all very similar and close to the truth (Figure 6.2c). Posterior standard deviations (Figures 6.5b) are smallest, as expected, for the plug-in method. The resampling method resulted in the highest overall uncertainty, while the full $\mathbf{Q}$ uncertainty method produced posterior uncertainties nearly identical to those of the resampling method. The posterior uncertainty from the low-rank $\mathbf{Q}$ method is lower than that of the full $\mathbf{Q}$ method, but higher than the uncertainty from the plug-in method.

Figure 6.6 shows the marginal posterior CDFs of γ_0 and γ_1 from the same simulated data. The plug-in method has the smallest posterior uncertainty for both parameters, but the difference is more apparent for γ_0. The resampling method has the highest posterior uncertainty, while the proposed methods provide a middle ground between the plug-in and resampling method. The posterior estimates from the full $\mathbf{Q}$ method and the low rank $\mathbf{Q}$ method are very similar.

Figures 6.7a and 6.7b show a comparison of the estimated posterior CDFs of γ_0 and γ_1, respectively, for different fixed values of the scaling parameter $\log(\tau_\epsilon)$ of the error component using the full $\mathbf{Q}$ method, as discussed in Section 6.1.2. The results show that as the log precision becomes larger, the estimated CDFs for γ_0 and



(a) Posterior mean of $\mathbb{E}(y(\mathbf{s})|\mu(\mathbf{s}))$

(b) Posterior SD of $\mathbb{E}(y(\mathbf{s})|\mu(\mathbf{s}))$

Figure 6.5: Comparison of the posterior mean and posterior SD of $\mathbb{E}[y(\mathbf{s})|\mu(\mathbf{s})] = \gamma_0 + \gamma_1\mu(\mathbf{s})$ for the two-stage Gaussian model in Section 6.3.1 from different approaches: plug-in method, resampling method, full **Q** method, low rank **Q** (mesh A), low rank **Q** (mesh B)

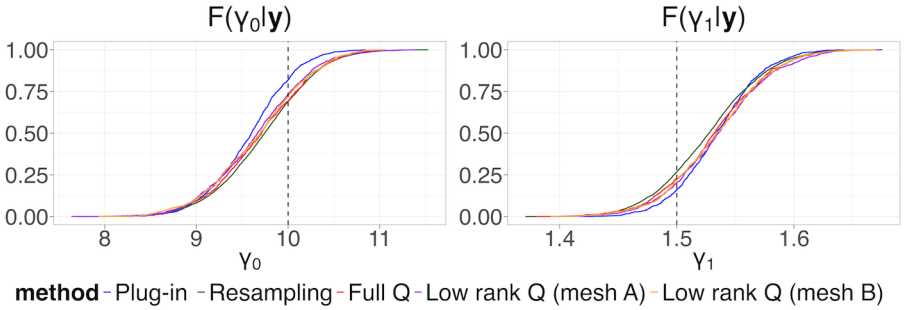
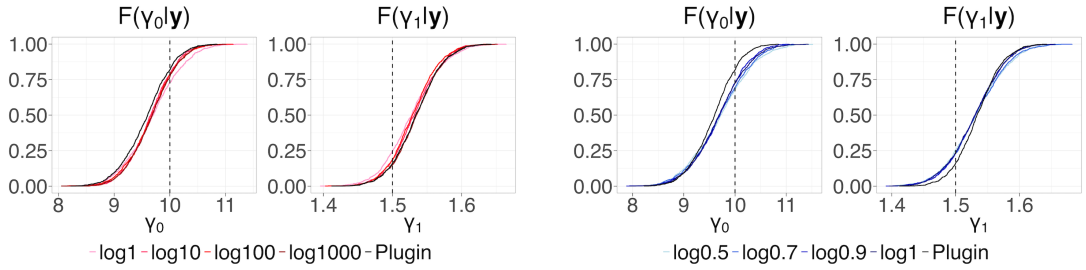


Figure 6.6: Estimated marginal posterior CDFs of γ_0 and γ_1 for a simulated dataset from the two-stage Gaussian model in Section 6.3.1 using four methods of uncertainty propagation: plug-in, resampling, full **Q** method, low rank **Q** (mesh A), and low rank **Q** (mesh B)



(a) Comparison with high log precision values

(b) Comparison with low log precision values

Figure 6.7: Comparison of the estimated marginal posterior CDFs of γ_0 and γ_1 for different fixed values of the log precision of the error component with the full **Q** uncertainty method using the simulated data example in Section 6.3.1

γ_1 approaches the estimated CDFs of the plug-in method. Also, as the log precision value becomes smaller, the estimated CDFs deviate more from the estimated CDFs of the plug-in method, i.e., the posterior uncertainty becomes larger. The same insights are true from the results of the low rank **Q** method (see Section C.1.5 of Appendix C).

In terms of computational time, the crude plug-in method took 2.98 seconds to fit the second-stage model, while the full $\mathbf{Q}$ method took 9.82 seconds. The low rank $\mathbf{Q}$ approach took 13.97 seconds with mesh A, and 5.80 seconds with mesh B. This suggests that using a coarse mesh with the low rank $\mathbf{Q}$ does not always lead to reduced computational time. A plausible reason for this is that the linear predictor of the low rank $\mathbf{Q}$ method, as shown in Equations (6.8) and (6.9), is more complex and involves additional operations compared to the linear predictor of the full $\mathbf{Q}$ method in Equation (6.7). The computational advantage of the low rank $\mathbf{Q}$ approach becomes evident from using a coarse enough mesh, similar to mesh B in Figure 6.3c. Lastly, the resampling method with $J = 30$, took 36.12 seconds using parallel computing with the `mclapply()` function in `R`.

The results from the SBC provide evidence that the plug-in method underestimates the posterior uncertainty, while the resampling method is correct. The full $\mathbf{Q}$ and the low rank $\mathbf{Q}$ methods are also expected to give correct posteriors for γ_0 and γ_1. The computational benefits from the low rank $\mathbf{Q}$ method potentially depends on the coarseness of the mesh for the error component.

6.3.2 A two-stage spatial model with Poisson likelihood

In this section, I perform the SBC in a two-stage spatial model with Poisson observations. I consider two model specifications: the first one, called the *classical specification*, is similar to Equations (1.4) and (1.5) in Section 1.3.1. Here, the second-stage model specifies the log mean of the Poisson counts in each block as linear with respect to the block averages of the first-stage field $\mu(\mathbf{s})$. This approach is often used in such spatial misalignment problems since it is straightforward to implement (Blangiardo et al., 2016; Cameletti et al., 2019; Lee et al., 2017, 2021; Liu et al., 2017; Villejo et al., 2023; Zhu et al., 2003). The second, named *new specification*, introduces a spatially continuous latent intensity field $\lambda(\mathbf{s})$ for the Poisson counts, which is then linked to the log mean in a nonlinear way (Lindgren et al., 2024). This approach better represents the physical process by assuming that the observed Poisson counts are function of the averages of a latent intensity field over the areas. Although this results in a highly nonlinear model, it can be efficiently fitted using an approximate iterative

method with INLA. This method extends the applicability of INLA beyond the linear predictor framework to accommodate more complex functional relationships and can be implemented with the `inlabru` library in `R` (Lindgren et al., 2024).

6.3.2.1 Classical specification

The first-stage latent model is similar to the one in Section 6.3.1, so that $\mu(\mathbf{s}) = \beta_0 + \beta_1 \mathrm{z}(\mathbf{s}) + \xi(\mathbf{s})$ and the observed data $\mathfrak{D}_1 = \{\mathrm{w}(\mathbf{s}_i), i = 1, \ldots, n_\mathrm{w}\}$ also follows the classical error model. The latent process $\mu(\mathbf{s})$ is an input in the second-stage model as follows:

$$\begin{aligned} \mathrm{y}(B) &\sim \text{Poisson}\Big(\mu_\mathrm{y}(B)\Big) \\ \mu_\mathrm{y}(B) &= \mathbb{E}[\mathrm{y}(B)] = \mathrm{E}(B) \times \lambda(B) \\ \log\Big(\lambda(B)\Big) &= \gamma_0 + \gamma_1 \frac{1}{|B|} \int_B \mu(\mathbf{s}) d\mathbf{s} \end{aligned} \tag{6.22}$$

The above model is closely related to the joint model in Equations (1.1) – (1.5) in Section 1.3.1, but here we have additional quantities $\mathrm{E}(B)$ which are introduced as an *offset* in order to account for the different sizes of the blocks B. For example, in spatial epidemiology where $\mathrm{y}(B)$ is the observed disease count, $\mathrm{E}(B)$ is the expected number of cases and are computed using the size and demographic structure of the population in block B (Lee, 2011). In this specification, $\lambda(B)$ is interpreted as the disease rate or risk (Blangiardo et al., 2016; Lee et al., 2017). The second-stage data is $\mathfrak{D}_2 = \{\mathrm{y}(B),\ \forall B\}$. Similar to Section 6.3.1, the first-stage and second-stage model latent parameters are $\boldsymbol{x}_1 = \{\beta_0, \beta_1, \omega_1, \omega_2, \ldots, \omega_K\}$ and $\boldsymbol{x}_2 = \{\gamma_0, \gamma_1\}$, respectively. The first-stage model hyperparameters are $\boldsymbol{\theta}_1 = \{\sigma_{e_1}, \sigma_\xi, \rho_\xi\}$. There are no second-stage model hyperparameters.

I simulate $\xi(\mathbf{s})$ and $\mathrm{z}(\mathbf{s})$ as in Section 6.3.1. The spatial locations of the first-stage observations and the meshes for the full $\mathbf{Q}$ and low rank $\mathbf{Q}$ are also as in Section 6.3.1. The configuration of the Poisson blocks is shown in Figure 6.10d.

I used the following priors for the fixed effects: $\beta_0 \sim \mathcal{N}(0, 10^2), \beta_1 \sim \mathcal{N}(0, 5^2), \gamma_0 \sim \mathcal{N}(-2, 1.5^2), \gamma_1 \sim \mathcal{N}(0, 0.1^2)$. I used the PC prior for σ_{e_1}, particularly $\mathbb{P}(\sigma_{e_1} > 1) = 0.5$, and the same joint log Gaussian prior for the Matérn parameters from Section

6.3.1. Again, four methods of uncertainty propagation are compared: plug-in method, resampling method, full $\mathbf{Q}$ method, and the low rank $\mathbf{Q}$ method. In fitting the plug-in method, I define the following as the $\mathbf{h}(\cdot)$ function:

$$\mathbf{h}(\hat{\boldsymbol{x}}_1) = \mathbf{C}\left[\mathbf{Z}_1\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\xi}}\right], \tag{6.23}$$

where $\mathbf{C}$ is an appropriate aggregation matrix that evaluates the integral in Equation (6.22). In Equation (6.23), $\mathbf{Z}_1\hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\xi}}$ is evaluated over a fine prediction grid, i.e., the $\mathbf{Z}_1$ covariate matrix contains the covariate information for all points in the prediction grid and $\hat{\boldsymbol{\xi}}$ is the corresponding estimated spatial field. The $\mathbf{C}$ matrix has dimension $\dim(\boldsymbol{\mathcal{D}}_2) \times \mathrm{n}_{\mathrm{grid}}$, where $\dim(\boldsymbol{\mathcal{D}}_2)$ is the number of blocks B in the second-stage data and $\mathrm{n}_{\mathrm{grid}}$ is the dimension of the vectorized prediction grid. The $\mathbf{C}$ matrix is a user-defined (sparse) matrix, created by first defining a binary matrix, whose $(i, j)^{\mathrm{th}}$ value is equal to 1 if the j^{th} element of the vectorized prediction grid is inside the i^{th} block, and then normalizing this matrix so that the row sums are equal to 1. The same $\mathbf{C}$ matrix is incorporated in the linear predictor of the full $\mathbf{Q}$ and low rank $\mathbf{Q}$ methods.

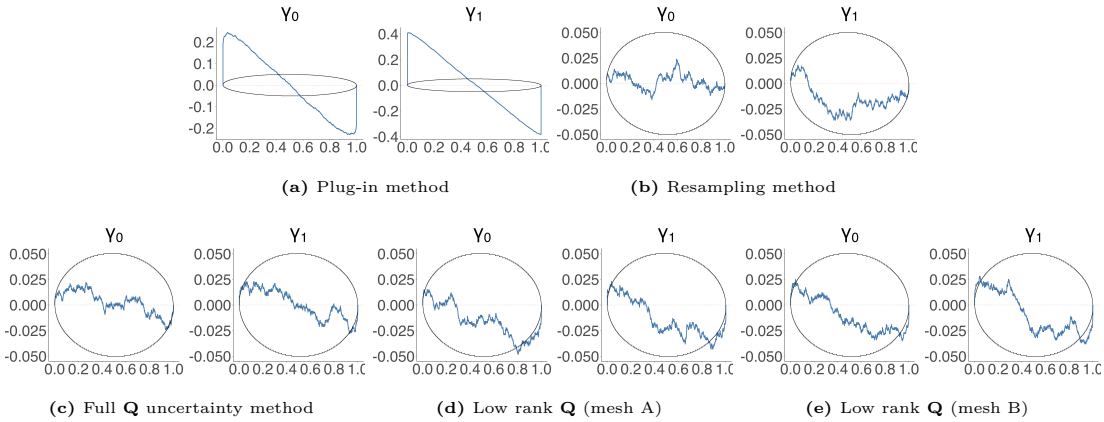


Figure 6.8: ECDF difference plot of p_k for γ_0 and γ_1 using Algorithm 6.2 out of 1000 data replicates for the classical specification of the two-stage Poisson spatial model (Section 6.3.2.1) and using different approaches: (a) plug-in method (b) resampling method (c) full $\mathbf{Q}$ method (d) low rank $\mathbf{Q}$ (mesh A) method (e) low rank $\mathbf{Q}$ (mesh B) method

Figure 6.8 shows the plot of the ECDF differences of p_k for γ_0 and γ_1 using Algorithm 6.2 with 1000 data replicates (corresponding histograms are in Section C.2.3 of Appendix C). Again, the plug-in method appears to underestimate the true posterior uncertainty for both parameters, while the resampling and the full $\mathbf{Q}$ uncertainty methods do not show deviations from uniformity. The two versions of the low rank

Q approach (mesh A and mesh B) show slight deviation from uniformity, but not as bad as the plug-in method. Results for the first-stage model parameters using Algorithm 6.2 are shown in Section C.2.1 of Appendix C. As in Section 6.3.1, the histogram of the normalized ranks p_k for σ_ξ is $\cap$-shaped. Moreover, there are some mesh nodes which also fail the uniformity test using the KS test at 10% significance level. Results using Algorithm 6.3 are shown in Section C.2.2 and Section C.2.4 in Appendix C for the first-stage and second-stage model parameters, respectively. The results are coherent with those from Algorithm 6.2.

Similar to Section 6.3.1, initial validation was done for a non-spatial two-stage model, i.e., without the spatial field $\xi(\mathbf{s})$ in the first stage. The results for both INLA and NUTS are in Section C.3.2 of Appendix C. The results are consistent with previous results; that the plug-in method underestimates the posterior uncertainty of both γ_0 and γ_1, while the resampling and the proposed method are correct.

6.3.2.2 New specification

For the new model specification, the second-stage model is as follows:

$$\begin{aligned}
&\mathrm{y}(B) \sim \text{Poisson}\Big(\mu_{\mathrm{y}}(B)\Big) \\
&\mu_{\mathrm{y}}(B) = \mathbb{E}\big[\mathrm{y}(B)\big] = \mathrm{E}(B) \times \lambda(B) \\
&\lambda(B) = \frac{1}{|B|}\int_B \lambda(\mathbf{s})d\mathbf{s} = \frac{1}{|B|}\int_B \exp\big\{\gamma_0 + \gamma_1\mu(\mathbf{s})\big\}d\mathbf{s}
\end{aligned} \tag{6.24}$$

The first-stage model is the same as the one used for the classical specification; but here we assume that $\mu(\mathbf{s})$ is linked to another latent intensity field which we denote by $\lambda(\mathbf{s})$. I used the same covariate $\mathrm{z}(\mathbf{s})$, the same spatial locations for the first-stage data, and the same configuration of the block/areas for the Poison outcomes as the classical specification. I also used the same priors for all model parameters, and compare the same uncertainty propagation methods.

The predictor expression in Equation (6.24) does not follow the general expression in Equation (6.1). The new model specification is non-linear with respect to both the first-stage latent parameters, and the second-stage fixed effects γ_0 and γ_1. To fit this model, I use the iterative linearized INLA approach, discussed in Section 2.5.4

in Chapter 2.

Figure 6.9 shows the plot of the ECDF differences of the normalized ranks p_k for γ_0 and γ_1 using Algorithm 6.2 and from 1000 data replicates. Histograms are provided in Section C.2.5 of Appendix C. The results show that the plug-in method underestimates the true posterior uncertainty for both parameters and introduces bias in the posterior distribution of γ_0. The resampling method correctly captures the posterior for γ_1, but shows some bias for γ_0, though less severe than the plug-in method. The full $\mathbf{Q}$ and low rank $\mathbf{Q}$ methods also show potential bias for γ_0 on the same direction as the plug-in and resampling method. Moreover, the ECDF difference plot reveals a slight deviation from uniformity for γ_1, though less pronounced than that of the plug-in method. This suggests that the $\mathbf{Q}$-based methods strike a balance between the plug-in and the resampling method. Results from Algorithm 6.3, shown in Section C.2.6 of Appendix C, align with the insights from Algorithm 6.2.

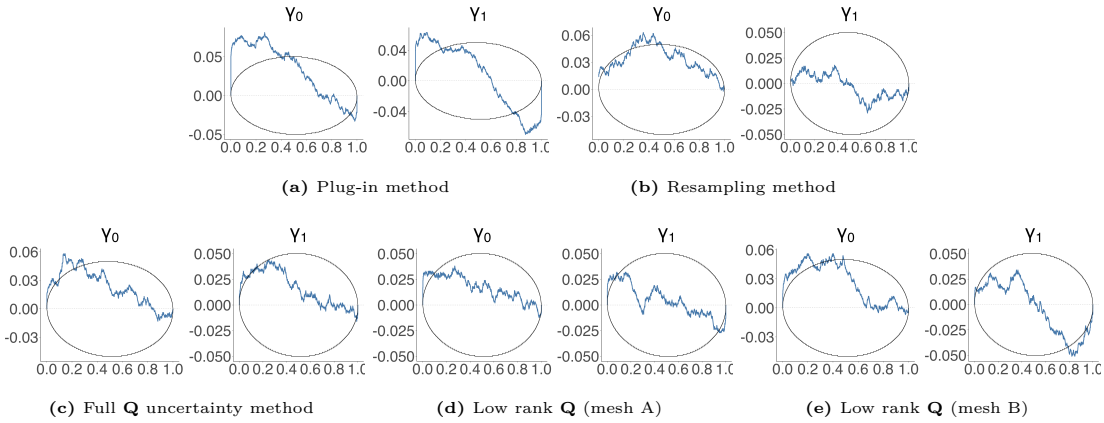


Figure 6.9: ECDF difference plot of p_k for γ_0 and γ_1 using Algorithm 6.2 out of 1000 data replicates for the new specification of the two-stage Poisson spatial model (Section 6.3.2.2) and using different approaches: (a) plug-in method (b) resampling method (c) full $\mathbf{Q}$ method (d) low rank $\mathbf{Q}$ (mesh A) method (e) low rank $\mathbf{Q}$ (mesh B) method

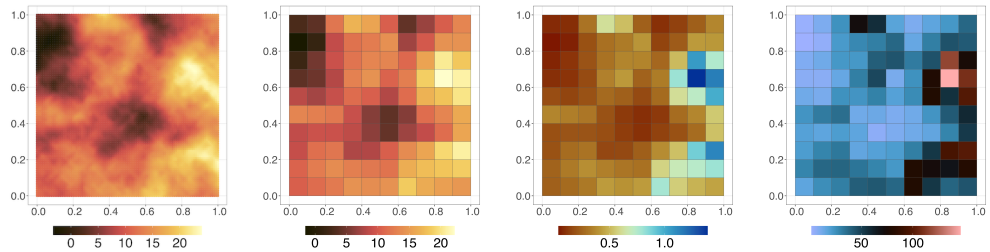
6.3.2.3 Illustration with simulated data

To illustrate the previous insights from the SBC, I simulate a data from both model specifications. I set the true values of the parameters as follows: $\beta_0 = 10, \beta_1 = 3, \gamma_0 = -3, \gamma_1 = 0.15, \sigma^2_{e_1} = 1, \sigma_\xi = 0.6, \rho_\xi = 4$. Moreover, I keep the spatial locations of the first-stage observations as in Section 6.3.1 (Figure 6.2a), and set $\mathrm{E}(B) = 100$ for all blocks.

Figure 6.10a shows the simulated $\mu(\mathbf{s})$. The classical specification aggregates $\mu(\mathbf{s})$

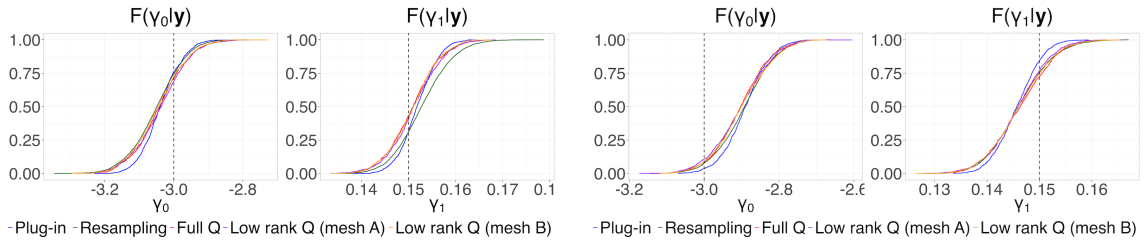
over the blocks, which is shown in Figure 6.10b. The corresponding $\lambda(B)$ are in Figure 6.10c, while the simulated Poisson outcomes are in Figure 6.10d. For the new specification, I first compute the latent intensity field $\lambda(\mathbf{s})$, which is then aggregated over the blocks to yield $\lambda(B)$. The simulated $\lambda(\mathbf{s}), \lambda(B)$, and $\mathrm{y}(B)$ for the new specification are shown in Figure C.21 in Appendix C.

Figures 6.11a and 6.11b show the estimated marginal posterior CDFs of γ_0 and γ_1 for the classical specification and new specification, respectively. For both model specifications, the plug-in method evidently has the smallest posterior uncertainty among the four approaches. The resampling method and the **Q**-based methods have very similar posterior results. The posterior median for γ_0 using the new model specification is slightly overestimated, but is well within the 95% credible interval. The results from this specific simulated data are consistent with the SBC results.



(a) $\mu(\mathbf{s})$ (b) $\mu(B)$ (c) $\lambda(B)$ (d) $\mathrm{y}(B)$

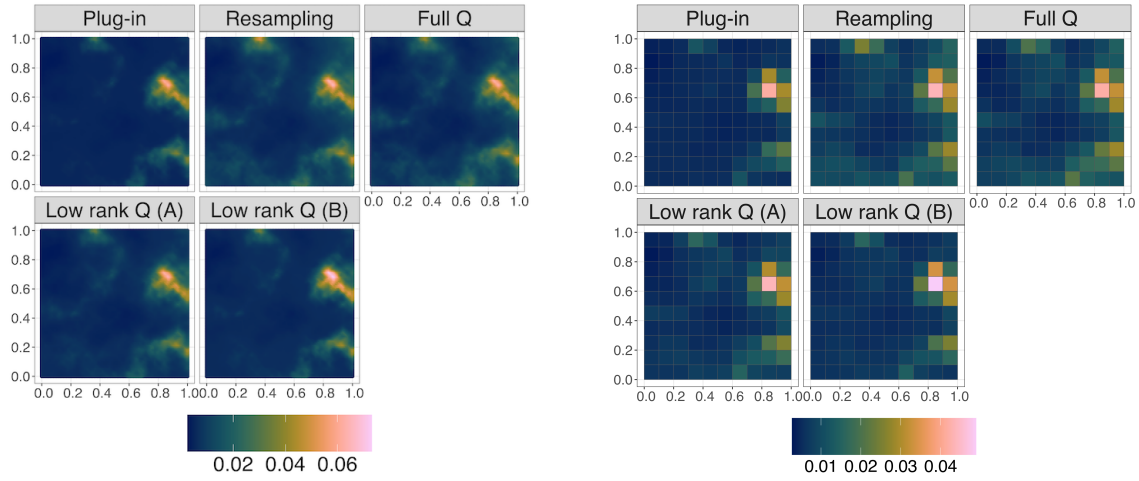
Figure 6.10: Simulated quantities from the classical model specification of the two-stage Poisson model in Section 6.3.2.1



(a) Classical specification (b) New specification

Figure 6.11: Marginal posterior CDFs of γ_0 and γ_1 for a simulated dataset from the two-stage Poisson spatial model: (a) classical specification and (b) new specification; and using different estimation approaches: plug-in, resampling method, full **Q** method, low rank **Q** (mesh A) method, and low rank **Q** (mesh B) method

Figures 6.12a and 6.12b show the posterior standard deviations of $\lambda(\mathbf{s})$ and $\lambda(B)$, respectively, using the new model specification and for the different uncertainty propagation approaches. It is evident that the plug-in approach generally has the smallest posterior uncertainty. The resampling method and the **Q**-based methods have quite



(a) SD of $\lambda(\mathbf{s})$ **(b)** SD of $\lambda(B)$

Figure 6.12: Comparison of the posterior uncertainty in (a) $\lambda(\mathbf{s})$ and (b) $\lambda(B)$ from a simulated data of the two-stage Poisson spatial model (new specification) using different approaches: plug-in method, resampling method, full $\mathbf{Q}$ method, low rank $\mathbf{Q}$ (mesh A) method, and low rank $\mathbf{Q}$ (mesh B) method

similar results, although the low rank $\mathbf{Q}$ method with a very coarse mesh has slightly smaller uncertainty estimates in some areas. The corresponding posterior means of $\lambda(\mathbf{s})$ and $\lambda(B)$ are shown in Figure C.23 in Appendix C. The posterior means of both $\lambda(\mathbf{s})$ and $\lambda(B)$ from the four uncertainty propagation approaches are very similar to the simulated truth. Moreover, the results from the classical model specification are shown in Figure C.22 in Appendix C. The results also show the same insights as the new model specification, i.e., the plug-in method has the smallest posterior uncertainty, while the resampling method and the $\mathbf{Q}$-based method have similar results.

Table 6.1 shows the computational time (in seconds) for the different estimation approaches on the simulated data examples. For both the classical and new model specification, the plug-in method has the fastest computing time. The resampling method took the longest time for the classical model, while the full $\mathbf{Q}$ approach had a significant reduction in the computational time. In addition, the low rank $\mathbf{Q}$ (mesh A) took longer to run than the full $\mathbf{Q}$ method, but the low rank $\mathbf{Q}$ (mesh B) was faster than the previous two. This is consistent with the results from the simulated data example in Section 6.3.1, which show that the coarseness of the mesh for the error component is crucial in terms of the reduction in the computational time. On the other hand, for the new model specification, the full $\mathbf{Q}$ method took the longest computational time. The new model specification is a highly non-linear model, and introducing an error component all the more increases the model complexity; hence,

making it plausible for the model fitting to even take longer. On the other hand, the low rank **Q** approach (both mesh A and mesh B) significantly reduced the computational time.

Method	Classical specification	New specification
Plug-in	3.04	4.64
Resampling	44.11	95.18
Full **Q**	14.28	110.76
Low rank **Q** (mesh A)	25.20	76.43
Low rank **Q** (mesh B)	6.78	20.38

Table 6.1: Summary of computational time (in seconds) for the different approaches on the data illustration for the two-stage Poisson model

The results for the two-stage Poisson models show that the plug-in method is expected to underestimate the posterior uncertainty in γ_0 and γ_1. On the other hand, the resampling approach is expected to be correct. However, there is a potential bias for the intercept γ_0 with the new model specification. The **Q**-based methods provide a middle ground between the plug-in method and the resampling method, but the gain in the computational time depends on the coarseness of the mesh for the error component. For the new model specification, which is a highly non-linear model, using a very fine mesh for the error component may not be recommended since doing model fitting could potentially take a longer time than the resampling approach.

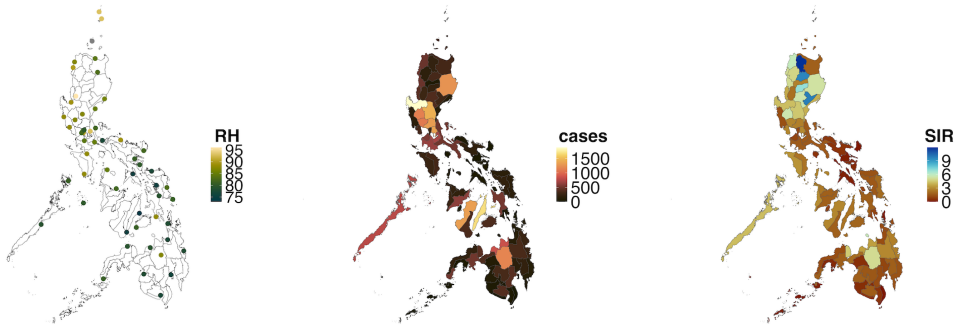
6.4 Real data application

This section illustrates the proposed method in a real data application, which aims to link relative humidity (RH) and dengue cases in the Philippines for August 2018. Relative humidity is known to increase the risks of dengue, since high humidity enhances reproduction and breeding, and increases survival and lifespan of mosquitoes (Murray et al., 2013; Naish et al., 2014; Thu et al., 1998). The emphasis in this section is a comparison of the different uncertainty propagation approaches.

6.4.1 Data

For this application, I used climate data from weather synoptic stations in the Philippines, as discussed in Chapter 4. Shown in Figure 6.13a are the spatial locations of

the weather stations in the Philippines as maintained by PAGASA. Data on dengue cases are obtained from the UN Office for the Coordination of Humanitarian Affairs, as discussed in Chapter 5. Shown in Figure 6.13b is a plot of dengue cases in the country by province for August 2018. The standardized incidence ratios (SIR) are shown in Figure 6.13c. The expected cases are derived via internal standardization (Waller and Carlin, 2010). The SIR indicates the relative excess in the incidence of the disease with respect to what might have been expected based on the reference national rates (Schoenbach and Rosamond, 2000).



(a) Weather stations **(b)** Dengue cases **(c)** Dengue SIRs

Figure 6.13: (a) weather stations in the Philippines (b) plot of dengue cases by province for August 2018 (c) plot of the standardized incidence ratios (SIR) of dengue by province for August 2018

I perform the inference in a two-stage modelling framework. The first stage models RH, while the second stage models the dengue health counts using information from the first-stage model as an input.

- **First-stage model** – Suppose $\mu(\mathbf{s})$ is the true relative humidity level at an arbitrary spatial location $\mathbf{s}$. I assume the following latent process:

$$\mu(\mathbf{s}) = \beta_0 + \beta_1 \text{Elevation}(\mathbf{s}) + \beta_2 \text{Temperature}(\mathbf{s}) + \beta_3 \big(\text{Temperature}(\mathbf{s})\big)^2 + \xi(\mathbf{s}),$$

where $\xi(\mathbf{s})$ is a Matérn field and $\boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{pmatrix}^\intercal$ are fixed effects. I assume that the observed values at the weather stations follow the classical error model, i.e., $\text{w}(\mathbf{s}_i) = \mu(\mathbf{s}_i) + e(\mathbf{s}_i)$ and $e(\mathbf{s}_i) \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_e^2), i = 1, \ldots, 57$. The temperature field is assumed to be known using the predicted values from the climate data fusion models in Chapter 4.

- **Second-stage model** – I consider both the classical and new specification in Section 6.3.2. Suppose $\mathrm{y}(B)$ and $\mathrm{E}(B)$ are the observed and expected dengue cases in province B, respectively. I assume that $\mathrm{y}(B) \sim \text{Poisson}\big(\mu_{\mathrm{y}}(B)\big), \mu_{\mathrm{y}}(B) = \mathbb{E}[\mathrm{y}(B)] = \mathrm{E}(B) \times \lambda(B)$. This implies that $\lambda(B) = \dfrac{\mathrm{y}(B)}{\mathrm{E}(B)}$, so that the disease risk $\lambda(B)$ is also interpreted as the model-based estimate of the SIR. For the classical specification, I assume that

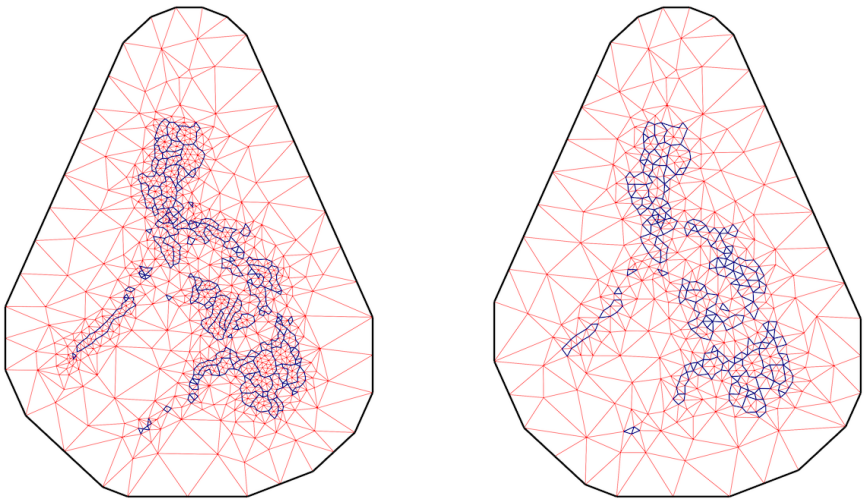
$$\log\Big(\lambda(B)\Big) = \gamma_0 + \gamma_1 \frac{1}{|B|}\int_B \mu(\mathbf{s})d\mathbf{s} + \phi(B),$$

where $\phi(B)$ is an area-specific effect, which I model as $\phi(B) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma_B^2\big)$. For the new specification, the model is given by:

$$\log\Big(\lambda(B)\Big) = \log\left(\frac{1}{|B|}\int_B \lambda(\mathbf{s})d\mathbf{s} + \phi(B)\right), \quad \lambda(\mathbf{s}) = \exp\big\{\gamma_0 + \gamma_1\mu(\mathbf{s})\big\}.$$

I used the INLA-SPDE approach to fit the models. The mesh, with 1077 nodes, used to estimate the the Matérn field $\xi(\mathbf{s})$ is shown in Figure 6.14a. I use vague priors for the fixed effects and σ_e, and a joint normal prior (Equations (6.20) and (6.21)) for the Matérn parameters. In particular, I use $\begin{bmatrix}\log(\tau)\\ \log(\kappa)\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}2.71\\ -4.66\end{bmatrix}, \begin{bmatrix}4 & 0\\ 0 & 1\end{bmatrix}\right)$. This implies that a plausible range of values for the range parameter is from 110 km to 800 km, which are consistent with the estimates in Chapter 4. Moreover, I set the plausible range of values for the marginal standard deviation as 0.2683 to 14.65, based on the empirical standard deviation of RH which is 5.37.

I compared the posterior estimates of γ_0 and γ_1 from the four uncertainty propagation approaches under consideration: plug-in, resampling, full $\mathbf{Q}$, and low rank $\mathbf{Q}$ approach. In this case study, there is no strong motivation for the low rank $\mathbf{Q}$ approach, since the dimension of the latent parameters is not large. Nonetheless, I also explored the low rank $\mathbf{Q}$ approach in order to have a full comparison with the other approaches. Figure 6.14b shows the mesh for the error component of the low rank $\mathbf{Q}$ approach, which has 546 nodes.

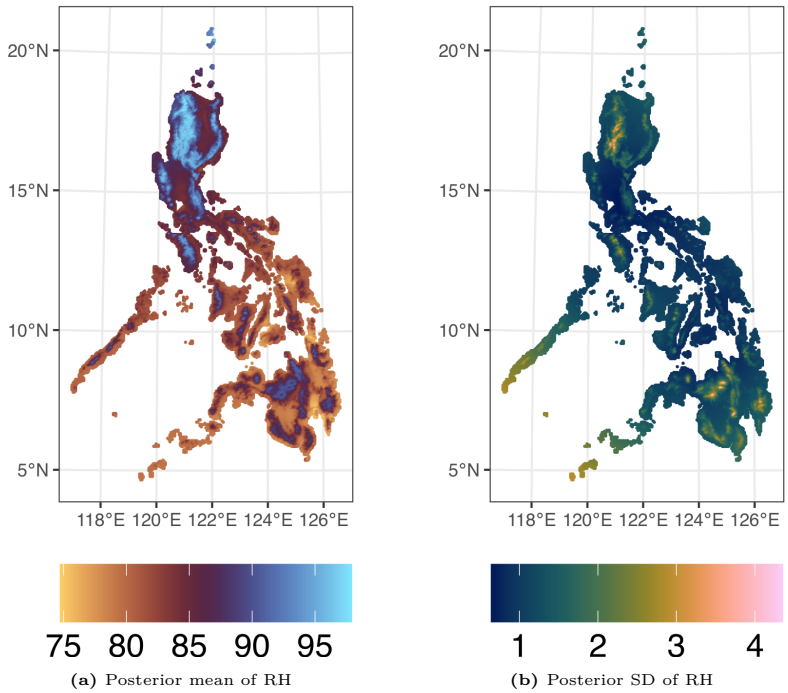


(a) Full **Q** mesh **(b)** Low rank **Q** mesh

Figure 6.14: (a) mesh for the full **Q** method (b) mesh for the low rank **Q** method

6.4.2 Results

Figures 6.15c and 6.15d show the estimated posterior mean and standard deviation of the relative humidity field. The posterior estimates of the parameters are reported in Table C.1 in Appendix C.



(a) Posterior mean of RH **(b)** Posterior SD of RH

Figure 6.15: (a) estimated RH field (b) posterior uncertainty of RH field

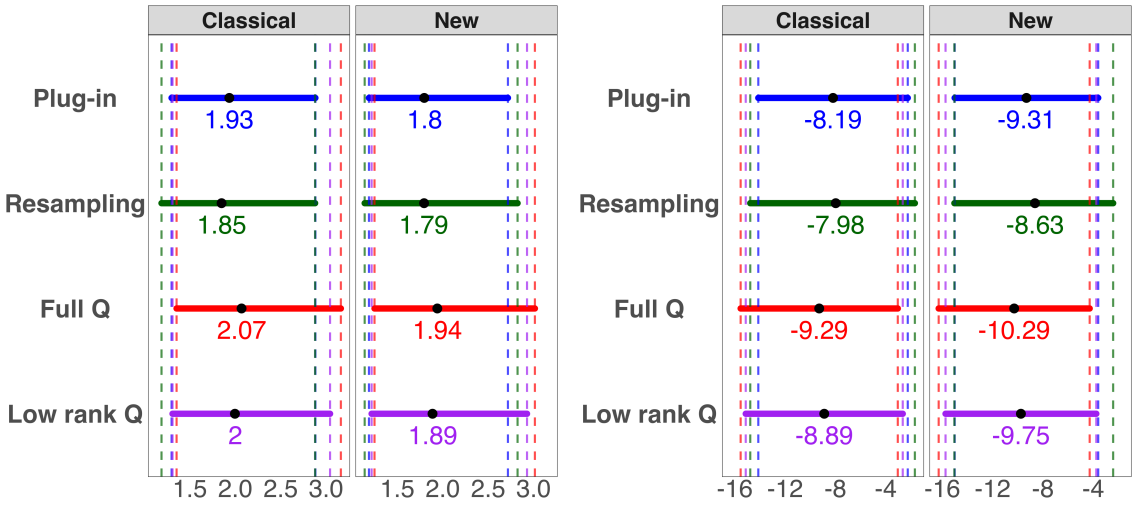
Figure 6.16a shows the estimated relative risks (RR) and the 95% CI (in dashed

lines) associated with one standard deviation (SD) change in RH. The results show that for a one SD change in RH, the risk of dengue approximately doubles. This is consistent with Chapter 5 and several studies which have shown a positive association between RH and the risk of dengue (Murray et al., 2013; Naish et al., 2014; Thu et al., 1998). The width of the CIs from Figure 6.16 does not seem to be too different, but the linear predictor here is in the log risk scale, which makes these differences not negligible (see also Chapter 5). Moreover, this is the typical length of the CIs in the literature (Blangiardo et al., 2016; Lee et al., 2017; Liu et al., 2017).

We also computed the RR for other ω-units change in RH. In particular, we considered $\omega = \{1, 2, 3, 4, 5.4\}$, where 5.4 corresponds to one SD in RH. The results are shown in Figure 6.17a. The figure shows that the CI widths in the RR are not too different when the magnitude of the change in the RH is small, but the difference in the CIs become more apparent when the change in RH is large. The posterior estimates for γ_0 and the 95% CI widths for the different uncertainty propagation approaches are shown in Figure 6.16b and 6.17b, respectively. The actual 95% CIs are shown in Figure C.35 of Appendix C.

Figures C.34a and C.34b in Appendix C show the estimated marginal CDFs of γ_0 and γ_1 for the classical and new specification, respectively, while the point estimates and 95% CI are shown in Tables C.2 and C.3 of Appendix C. The resampling method and the two proposed methods have slightly larger posterior uncertainty than the plugin method for γ_1. The differences in the posterior uncertainty for γ_0 are more apparent.

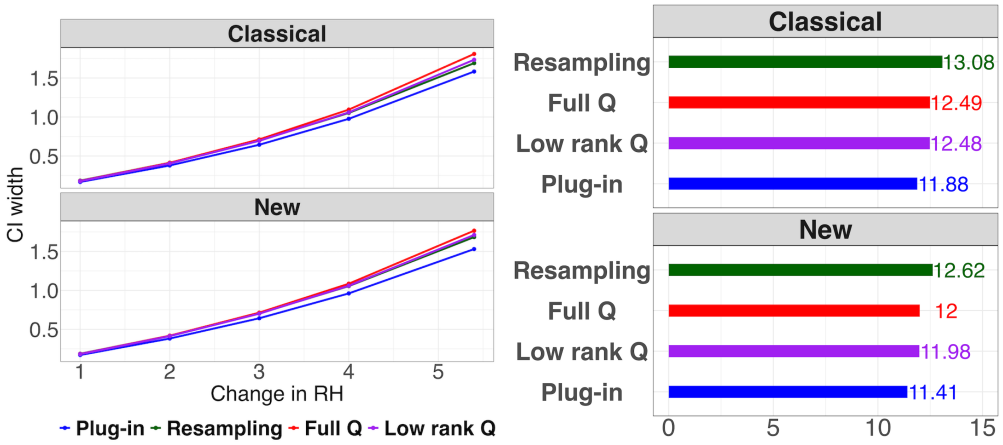
Figure 6.16 shows some differences in the posterior results among the four uncertainty propagation approaches. The posterior mean and the lower limit of the 95% CI of the RR for the resampling method is the lowest among the four uncertainty propagation approaches. This attenuation to the null risk of one is also observed in Lee et al. (2017) and Liu et al. (2017), where they argue that it is due to the posterior predictive distribution of the first-stage model outweighing the spatial (or spatio-temporal) variation in the data, which results in the estimated effects being washed away. Even so, we also see that the posterior mean for γ_0 from the resampling method is the highest, so that both parameters balance each other out when calculat-



(a) 95% CI of RR with 1 SD change in RH

(b) 95% CI for γ_0

Figure 6.16: (a) 95% CI of RR associated with 1 standard deviation change in relative humidity (b) 95% CI for γ_0. Shown in broken lines (- - -) are the lower and upper limit of the 95% CI. The black dot (·) is the posterior mean



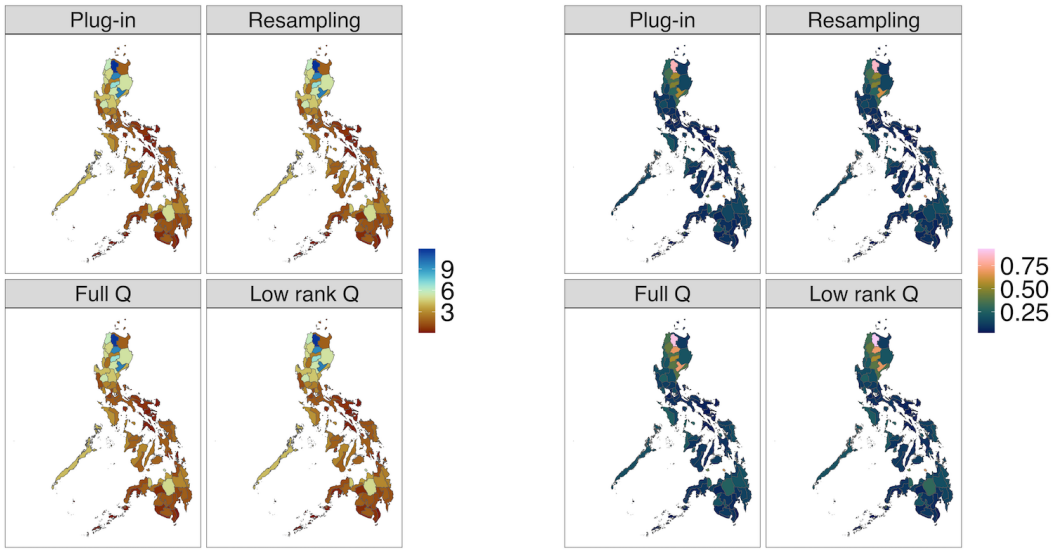
(a) 95% CI widths of RR for different ω-units

(b) 95% CI width for γ_0

Figure 6.17: (a) 95% CI width of RR associated with ω-units change in RH (b) 95% CI width for γ_0

ing the log risks, $\log\Big(\lambda(B)\Big)$. This is the same observation, although in the opposite direction, from the results of the **Q**-based methods, where the posterior mean of γ_1 is relatively high, but the posterior mean for γ_0 is relatively low. This observed push and pull between the two parameters explains why the estimated posterior means of $\lambda(B)$ for the four uncertainty propagation methods are very similar, as shown in Figure 6.18a for the classical specification and Figure C.36 in Appendix C for the new specification. Moreover, the estimated disease risks look very similar to the computed SIRs in Figure 6.13c. The corresponding posterior SDs are shown in Figure 6.18b for the classical specification and Figure C.37 in Appendix C for the new specification. The posterior SDs from the plug-in and resampling method are very similar, which

is the same result as Chapter 5. The **Q**-based methods have slightly higher posterior SDs for some areas in the northern part of the country.



(a) Posterior mean of $\lambda(B)$ **(b)** Posterior SD of $\lambda(B)$

Figure 6.18: (a) Posterior mean of $\lambda(B)$ from the classical model specification (b) Posterior SD of $\lambda(B)$ from the classical model specification

6.5 Conclusions

This chapter formally addresses the problem of uncertainty propagation in two-stage Bayesian models. This approach is appropriate for scenarios when there is a clear one-directional physical relationship between the two models. Also, it is a practical approach when the first-stage model is already complex in itself; for example, it might involve fitting a complex data fusion model, such as the data fusion model in Chapter 4. In addition, a two-stage modeling framework avoids potential unwanted feedback effects that could occur in fully Bayesian approaches (Gryparis et al., 2009; Shaddick and Wakefield, 2002; Wakefield and Shaddick, 2006). The drawback of the two-stage modeling framework is that uncertainty is not automatically propagated between the two models.

In this chapter, I validated different uncertainty propagation approaches for two-stage models by testing the self-consistency property of Bayesian models using the simulation-based calibration (SBC) method of Talts et al. (2018). In particular, I investigated the correctness of two commonly used methods for two-stage modeling:

the plug-in method and the resampling method. In addition, I also explored a new method called the **Q** uncertainty method. This introduces a new model component, called an *error component*, in the second-stage model. The error component is given a Gaussian prior with mean zero and precision matrix **Q**, derived from the Gaussian approximation of the latent parameters of the first-stage model. The **Q** matrix can be of high dimension (for example, in large spatio-temporal applications); hence, I also proposed a low rank approximation of the **Q** matrix. Thus, there are two versions of the proposed method: the full **Q** method and the low rank **Q** method. The **Q** uncertainty method is implemented using the INLA methodology.

Another contribution of this chapter is the proposed modification of the SBC method of Talts et al. (2018) to address challenges specific to two-stage Bayesian models. The proposed SBC variant is implemented conditional on fixed values of certain first-stage hyperparameters. This approach ensures that the evaluation focuses on the second-stage model parameters, avoiding the influence of first-stage parameters that may violate the self-consistency property. I also referred to this proposed variant as the conditional SBC.

Results from both the original SBC and the conditional SBC in the simulation experiments confirm that the plug-in method underestimates the posterior uncertainty of the second-stage model parameters, while the resampling method provides correct uncertainty estimates. The proposed **Q**-based methods also produce correct posterior uncertainty estimates, although results indicate that the coarseness of the **Q** approximation can affect the accuracy of the approximate posteriors and the computational cost. However, there were some inadequacies when applying the different methods on the new specification of the Poisson model. In this work, I considered the individual parameters as test functions. The use of test functions which are data-dependent can be done in future work.

The computational efficiency of the **Q**-based methods depends on the coarseness of the mesh for the error component. For the new specification of the Poisson model, the full **Q** method required longer computational time compared to the resampling method. This is because the **Q**-based methods introduce an additional model component, significantly increasing model complexity and making fitting more computa-

tionally intensive, particularly for the highly nonlinear Poisson model. Nevertheless, a sufficiently coarse mesh in the low rank $\mathbf{Q}$ method can address this challenge while maintaining correctness. Moreover, the low rank $\mathbf{Q}$ approach may also take longer to run compared to the full $\mathbf{Q}$ method if the resolution of the $\mathbf{Q}$ approximation is not coarse enough. The reason for this is that the predictor expression of the second-stage model which implements the low rank $\mathbf{Q}$ approach involves more matrix operations, which may potentially increase the computational requirements for model estimation.

Some aspects of the $\mathbf{Q}$ uncertainty method require further investigation. Firstly, I fixed the scaling parameter of the $\mathbf{Q}$ matrix to 1, but this choice may not be optimal. If this parameter were estimated rather than fixed, the results showed that the estimated value of the scaling parameter tends to be very large. This behavior implies an increased confidence in the first-stage posterior estimates, effectively reducing the uncertainty in the error component, and consequently producing narrower uncertainty estimates for the second-stage model parameters. The simulation experiments revealed that as the fixed value of the scaling parameter decreases, the posterior uncertainty of the second-stage model parameters widens and deviates more significantly from the crude plug-in method. Conversely, when the scaling parameter increases, the posterior uncertainty narrows, approaching the uncertainty estimates produced by the plug-in method. Despite this, the SBC results indicated that fixing the scaling parameter to 1 appears appropriate, as it did not violate the self-consistency property of the model. Nevertheless, further work is needed to determine whether this choice is indeed optimal.

Secondly, the low rank $\mathbf{Q}$ method requires a more thorough investigation. I hypothesize that the coarser the approximation to the $\mathbf{Q}$ matrix, the more likely it becomes that the self-consistency of the model is violated. I have not properly explored and investigated the breakdown point of the low rank approximation of $\mathbf{Q}$. This breakdown point is likely influenced by several factors, including the smoothness of the random field and the relative proportion of variability in the response variable explained by the fixed covariates versus the random field. A deeper understanding of these dependencies is essential to ensure the robustness of the low rank approximation.

Thirdly, I used the empirical Bayes approach to fit the models in both the simulation experiments and data application. This implies that the **Q** matrix is computed at the mode of the first-stage model hyperparameters. If another numerical integration strategy is chosen by the user when implementing INLA, such as a grid approach, there will be several **Q** matrices, one for each of the integration point for the model hyperparameter. In this scenario, I propose the use of the weighted average of the **Q** matrices, where the weights are the same integration weights from the numerical integration used to compute the approximated posteriors of the latent parameters.

The SBC method is a computational method for testing the self-consistency property of Bayesian models. It is implemented in a specific Bayesian model, prior specification, and Bayesian inference algorithm. In this chapter, I validated simple two-stage spatial models. However, in practice, the models that are investigated are more complex. For the toy models considered in this chapter, results have shown and illustrated that the crude plug-in method indeed underestimates the posterior uncertainty of the second-stage model parameters. For more complex models, this underestimation of the posterior uncertainty will also be highly likely true. Moreover, results also showed that the resampling method is correct. I think that the resampling method should also be able to compute the correct posterior uncertainty for more complex models. However, the only way to exactly know this is to implement the SBC method for every new Bayesian model specification, new prior specification, and new Bayesian algorithm. This aligns with the proposal in Talts et al. (2018) that SBC should be an integral part of a robust Bayesian workflow (Gelman et al., 2020). However, the SBC method is computationally expensive and might not be a practical route in many contexts. Therefore, I propose that a more practical approach to perform a two-stage model analysis is to implement different uncertainty propagation approaches, and compare the obtained posterior uncertainties. The crude plug-in method is definitely the easiest strategy, but the results derived from such should only be taken as an initial understanding of the model. A more comprehensive analysis should involve doing resampling and other approaches, such as the proposed **Q** uncertainty method when the Bayesian inference is done using INLA. Implementing different uncertainty propagation strategies allows an objective comparison of the estimated posterior un-

certainties, which would then help uncover interesting model insights and guide both the statistical and practical interpretation of the results.

Chapter 7

Conclusions and Future Work

7.1 Main contributions of the theses

This thesis tackles a common framework in spatial epidemiology, which performs inference in two stages (Blangiardo et al., 2016; Cameletti et al., 2019; Lee et al., 2017; Liu et al., 2017). The first stage fits the model for the covariate whose effect on the health outcome is of interest. The second stage then fits the health model using the predictions from the first-stage as an input to the model. It is argued in Chapter 1 why a two-stage modelling approach is practical and/or ideal. Firstly, it offers an intuitive physical interpretation, e.g., climate affects dengue incidence, and air pollution affects incidence of respiratory diseases, but not the other way around. Secondly, it is computationally efficient, especially when the first-stage model is already complex in itself. As an example, the data fusion models presented in Chapters 3 and 4 were complex models; hence, it is ideal to focus on developing these models separately from the health model. Doing joint modelling in this context will be very computationally challenging and expensive. Moreover, doing multiple health effect analyses or running multiple candidate epidemiological models requires refitting the first-stage model. A joint modelling framework in this context also is not practical. Thirdly, a two-stage modelling framework avoids potential feedback effects which happens in a joint modelling framework (Gryparis et al., 2009; Shaddick and Wakefield, 2002; Wakefield and Shaddick, 2006).

This work does not perform a comparison between a two-stage modelling approach and a joint modelling approach, but assesses the appropriateness of common approaches used in doing two-stage modelling.

The main contributions of this theses are the following:

1. The first contribution is a proposed model which combines outcomes from multiple data sources with different accuracy levels and sparsity, a process called data fusion. The proposed model is based on the Bayesian melding model (Fuentes and Raftery, 2005), and is an extension of what has been done in the literature. It has a flexible specification for the different biases and measurement errors in the data outcomes, and also can extend to scenarios with more than two data sources with varying spatial support and resolutions. This is tackled in both Chapter 3 and 4. Chapter 3 provides an initial exploration of the problem using the INLA-SPDE approach; but this model specification is not flexible enough to account for the biases in the proxy data. Chapter 4 provides a more flexible model specification. This chapter also shows that the proposed data fusion model outperforms a stations-only model and the regression calibration model. A concrete data application is presented, which is motivated by a meteorological data problem in the Philippines.

2. The second main contribution of this thesis is the validation of two commonly used methods for doing two-stage modelling: a plug-in method and the resampling method. These two algorithms are formally presented in Chapter 6 and are extensively applied in Chapter 5 in linking climate and dengue in the Philippines. I used the simulation-based calibration approach (Talts et al., 2018), which tests the self-consistency property of Bayesian models, to validate the two approaches. Results show that the plug-in method underestimates the uncertainty in the second-stage model parameters, while the resampling method is correct.

3. The third main contribution of this thesis is a new approach to do uncertainty propagation in two-stage models. This approach is called the **Q**-uncertainty approach. It avoids resampling, and is hence potentially faster, and more im-

portantly, is able to compute correct uncertainty estimates. Moreover, this work proposes a low rank approximation of the proposed method, which is beneficial for large spatio-temporal applications. To validate and assess the correctness of the proposed method, I also used the simulation-based calibration (SBC) approach. This is discussed in Chapter 6.

4. The fourth main contribution of this thesis is the proposed variation in the SBC, which is useful in validating two-stage models. The motivation of the proposal is that some parameters in the first-stage model may validate the self-consistency property, but the main parameters of interest are the second-stage model parameters. The SBC variant is therefore implemented by fixing certain parameters in the first-stage model. This variant was also referred to as conditional SBC, since the implementation is conditional on fixed values of certain parameters. This is presented in Chapter 6.

5. The last main contribution of this work is an extensive case study on the association between climate and dengue in the Philippines, which is presented in Chapter 5. In particular, I looked at three climate variables: temperature, relative humidity, and rainfall. Results show that temperature has a non-linear association with dengue. Moreover, rainfall and relative humidity have a spatially varying association with dengue, depending on the climate type of the region.

7.2 On the proposed data fusion models

Data fusion is defined as the process of combining different data sources to estimate a quantity of interest. For example, in meteorological applications, we have data from a network of weather stations which is usually sparse. In addition, there are weather forecast models and remote-sensed data (via satellites as an example) which provide wide spatial coverage but can be heavily biased. It is advantageous to combine these data sources in order to improve predictions, e.g., a more accurate estimate of the temperature field.

The proposed data fusion model in this thesis is based on the Bayesian melding model. The model assumes that the different data sources are outcomes of the same latent process. This allows all data sources to inform about the true unknown process, while considering the inherent biases in each data source. Note that the different data sources can have different spatial supports. In this thesis, instead of treating the proxy data as areal, they are considered as point-referenced at the centroids. Chapter 4 provides a justification for this assumption. More formally, suppose $x(\mathbf{s})$ is the latent process of interest, which is spatially structured, and that there are two data sources: given by $\mathrm{w}_1(\mathbf{s}_i)$ and $\mathrm{w}_2(\mathbf{g}_j)$, where the former denotes the outcome from a station at location $\mathbf{s}_i$, and the latter denotes the outcome of the proxy data at the grid cell with centroid $\mathbf{g}_j$. The data fusion model, in a purely spatial context, is given by:

$$\begin{aligned}
\mathrm{w}_1(\mathbf{s}_i) &= x(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad \epsilon(\mathbf{s}_i) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_\epsilon\big) \\
\mathrm{w}_2(\mathbf{g}_j) &= \alpha_0(\mathbf{g}_j) + \alpha_1(\mathbf{g}_j) x(\mathbf{g}_j) + \delta(\mathbf{g}_j), \quad \delta(\mathbf{g}_j) \overset{\text{iid}}{\sim} \mathcal{N}\big(0, \sigma^2_\delta\big).
\end{aligned}$$

The model components $\alpha_0(\cdot)$ and $\alpha_1(\cdot)$ are referred to as the additive bias and multiplicative bias, respectively. Chapter 3 provides an initial exploration of the model, in a spatio-temporal context, by assuming that both the additive bias and multiplicative bias are constant. The INLA-SPDE method with a data augmentation approach are able to correctly estimate the parameters based on the simulation study. The sparsity of the stations data indeed affects the quality of the parameter estimates, in terms of the bias and RMSE, unless (weakly) informative priors are used. The number of time points also has a potential effect on the quality of parameter estimates, specifically, lower biases and RMSEs with more time points. The main limitation of the proposed (initial) model in Chapter 3 is that it is not flexible enough to account for the biases in the proxy data, since both $\alpha_0(\cdot)$ and $\alpha_1(\cdot)$ are assumed constant.

Chapter 4 proposes a more flexible specification by assuming that the additive bias $\alpha_0(\cdot)$ is spatially varying, and is referred to as an *error field*. In a spatio-temporal context, the error field is assumed to vary in time. However, the model still assumes a constant multiplicative bias α_1. The results from a simulation study show that the proposed model outperforms a stations-only model and a regression calibration model

when looking at the accuracy of the predicted fields and the parameter estimates. The data application also shows that the proposed model outperforms the two benchmark approaches based on the leave-group-out cross validation.

In fitting the data fusion model, I explored both a data augmentation approach (similar to Chapter 3), an iterative linearized INLA approach (discussed in 2.5.4), and the Bayesian model averaging approach (BMA) with INLA (Gómez-Rubio et al., 2020). The data augmentation approach can be numerically unstable, while the iterative linearized INLA had convergence issues. Thus, the BMA with INLA was eventually used, since it was numerically stable, and it removes non-linear model components in the predictor expression when doing model fitting. This approach is intuitive in the data fusion context since sensible values for the bias parameter α_1 can be specified. If $\alpha_1 = 1$, it implies the absence of multiplicative bias, and if $\alpha_1 \ll 1$ or $\alpha_1 \gg 1$, then the more severe the multiplicative bias. Thus, for model estimation, I defined a grid of α_1 values centered on 1, and then estimated the model conditional on α_1. The final estimates are then obtained by Bayesian model averaging.

To emphasize an important point in Chapter 4, the proposed data fusion model offers several advantages: it defines a unified latent process for all data outcomes, accounts for measurement errors for all data sources, provides flexibility in addressing biases, accommodates multiple spatially-misaligned data sources, and gauges the relative quality of the data sources.

7.2.1 Future work

An immediate extension of the model is to assume that $\alpha_1(\mathbf{s})$ is also a random field. Note that if $x(\mathbf{s})$ were known, then the predictor expression for $\mathrm{w}_2(\mathbf{g}_j)$ is a spatially-varying coefficient model. However, if both $x(\mathbf{s})$ and $\alpha(\mathbf{s})$ are random fields or functions of such, then this is a difficult problem, primarily due to identifiability issues. Strategies to improve identifiability include introducing constraints in the model specification or introducing additional structure to reduce model complexity.

Another avenue for future work is to continue investigating the meteorological data problem in the Philippines, particularly by incorporating a third data source, such as remotely sensed data from satellite imagery. The proposed data fusion model

is able to combine more than two data sources of different spatial supports; hence, this extension is straightforward to implement. Another data application of interest is UK pollution data (Forlani et al., 2020), specifically combining data from a monitoring network maintained by Automatic Urban and Rural Network (AURN) and data from two numerical models: Air Quality Unified Model (AQUM) and Pollution Climate Mapping Model (PCM).

7.3 On the validation of approaches for two-stage modelling

Chapter 5 discussed the uncertainty propagation problem in a two-stage modelling framework. Two commonly used methods for doing two-stage modelling are the crude plug-in method and the posterior sampling method. The former ignores the uncertainty in the first-stage model, while the latter accounts for the uncertainty by sampling from the posterior distributions of the first-stage model, and then fitting the second-stage model using each sample. This thesis validates the correctness of the two aforementioned approaches by testing for the self-consistency property of Bayesian models via the simulation-based calibration method (Talts et al., 2018).

I performed the validation of specific Bayesian spatial models, specific prior specification, and using the INLA approach for Bayesian inference. The results show that the plug-in method indeed tends to underestimate the uncertainty in the second-stage model parameters, while the resampling method is correct. This is consistent with the results from Chapter 5 which looks at the link between climate and dengue in the Philippines. There is a difference in the posterior standard deviations of the second-stage model parameters, with the resampling method generally giving higher posterior uncertainty compared to the crude plug-in method.

The SBC method is implemented for a specific Bayesian model, prior specification, and algorithm/inferential approach. Chapter 6 employed the SBC method in toy spatial models, to illustrate the underestimation of posterior uncertainty when using a crude plug-in method and the correct uncertainty estimation when using the posterior sampling method. Ideally, the SBC method should be implemented on every Bayesian

model and algorithm. In fact, Talts et al. (2018) proposed that the SBC should be an integral part of a robust Bayesian workflow, which includes the following key three steps: model building, inference, and model checking/improvement (Gelman et al., 2020). However, in practice, the models we postulate and investigate are complex. This means that performing SBC is computationally expensive and might not be feasible or practical in some applications. Thus, a good strategy is to explore different approaches when doing two-stage modelling, and then compare the results from the different approaches. The crude plug-in method is definitely the easiest approach, but the results derived from such an approach should only be used to provide initial understanding of the model. Implementing other approaches, such as the posterior sampling approach, should also be considered. A comparison of results from different two-stage modelling approaches provides a more extensive and deep understanding of the model.

7.4 On the proposed Q uncertainty method

A main contribution of this thesis is the proposed new approach for doing uncertainty propagation in a two-stage modelling framework, called the **Q** uncertainty method. The main advantage of the method is that it does not do resampling from the first-stage model posteriors; which means that it does not need to fit the second-stage model several times. This is done by incorporating both the posterior mean and posterior variances and covariances of the first-stage latent parameters in the second-stage model predictor expression. The variances and covariances are encoded in the **Q** matrix, and is introduced in the second-stage predictor expression as a fixed hyperparameter value of a latent vector called the *error component*. More formally, this means that the predictor expression in the second-stage model can be generally written as follows:

$$g(\,\cdot\,) = \gamma_0 \mathbf{1} + \gamma_1 \mathbf{h}(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1} + \boldsymbol{\epsilon}, \cdot),$$

where $g(\cdot)$ is the link function, $\mathbf{h}(\cdot)$ is a vector-valued function of $\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1}$, which is the posterior mean of first-stage latent parameters, and $\boldsymbol{\epsilon}$ is the error component,

which encodes the uncertainty in the first-stage model and is given a Gaussian prior with mean $\mathbf{0}$ and precision matrix $\mathbf{Q}$. The unknowns $\{\gamma_0, \gamma_1\}$ are fixed effects. The computational bottleneck of the above expression is that it involves a product of two Gaussian model components, particularly γ_1 and the $\mathbf{h}(\hat{\boldsymbol{\mu}}_{\boldsymbol{x}_1} + \boldsymbol{\epsilon}, \cdot)$. This is not straightforward to fit in the traditional INLA framework, since the model is technically not latent Gaussian. There are several ways to fit the model in the INLA framework, as discussed in Section 6.1.2 and Section 7.5. In this work, I explored the use of the iterative linearized INLA method (see Section 2.5.4), for which the accuracy of the posterior approximations and the speed of computation depends on the degree of non-linearity in the predictor expression (Lindgren et al., 2024).

The SBC results show that the $\mathbf{Q}$ method gives correct posterior uncertainty estimates on the toy spatial models considered in Chapter 6, although there may be some inadequacies when applied on the new specification of the Poisson model. This work also proposed a low rank approximation of the $\mathbf{Q}$ matrix, which can be beneficial for large spatio-temporal applications. This is done by defining the error component on a coarser mesh, and then solving for the best linear mapper between the two meshes with respect to the variance-covariance structure of the weights at the nodes of the finer mesh.

The computational benefits from using the $\mathbf{Q}$ method depend on the coarseness of the mesh and the degree of non-linearity in the predictor expression. For the toy spatial model with a Gaussian likelihood, the low rank $\mathbf{Q}$ method with a not coarse enough mesh took longer to run compared to the full rank $\mathbf{Q}$ method. This is expected since the low rank $\mathbf{Q}$ method involves additional operations in its predictor expression. With the use of an even coarser mesh for the error component, the computational time was faster than the full rank $\mathbf{Q}$ method, without sacrificing the accuracy of the posterior approximations. For the toy spatial model with a Poisson likelihood and a classical model specification, both the full rank and low rank $\mathbf{Q}$ methods were faster compared to the resampling method. Also, similar to the Gaussian case, the low rank $\mathbf{Q}$ approach with a not coarse enough mesh took longer to run compared to the full rank $\mathbf{Q}$ method. For the new specification of the Poisson spatial model, however, the full $\mathbf{Q}$ method took longer to run compared to the resampling method. This is

due to the highly non-linear predictor expression which slows down the convergence of the iterative linearized INLA approach. Thus, there is no guarantee that the **Q** method will always outperform the resampling method in terms of computational efficiency. The **Q** method involves non-linear model components, which does not fit quite conveniently in the INLA framework.

7.4.1 Future work

In the current implementation of the **Q** uncertainty method, I fixed the scaling parameter τ equal to 1. This parameter may need to be optimally determined. When this parameter is fixed to a large value, the posterior uncertainty of the second-stage model coefficient narrows; while if the scaling parameter is fixed to a value smaller than 1, the posterior uncertainty widens. This is expected since increasing the scaling parameter also inflates the precision of the first-stage latent parameters, while decreasing its value inflates the posterior variances and covariances of the first-stage latent parameters. Fixing the value of the scaling parameter to 1 is reasonable, and also shows correct posterior aproximations based on the SBC method. If the scaling parameter were not fixed, it is usually estimated to be very large, which means that the posterior estimates are very similar to the crude plug-in method. Nevertheless, further work is needed to assess the optimality of the choice for the scaling parameter value.

Moreover, a thorough understanding of the low rank **Q** method needs to be done. I postulate that the coarser the approximation to the **Q** matrix is, it becomes more likely that the self-consistency of the model is violated. I have not properly explored and investigated the breakdown point of the low rank approximation of **Q**. As mentioned in Chapter 6, the breakdown point can be influenced by several factors, including the smoothness of the random field and the relative proportion of variability in the response variable explained by the fixed effects versus the random effects.

I used the empirical Bayes approach in the initial development of the method, which implies that the **Q** matrix is computed at the mode of the first-stage model hyperparameters. If another numerical integration strategy is chosen by the user when implementing INLA, such as a grid approach, there will be several **Q** matrices,

one for each of the integration point for the model hyperparameter. In this scenario, I propose computing the weighted average of the $\mathbf{Q}$ matrices, where the weights are the same integration weights from the numerical integration used to compute the posteriors of the first-stage latent parameters.

7.5 On fitting conditional latent Gaussian models

Many of the models mentioned in this thesis are not latent Gaussian, i.e., they do not fall in the class of models in the INLA framework. Examples are the new specification of the Poisson spatial model discussed in Chapter 6, a data fusion model which involves a product of two Gaussian model components, and the proposed $\mathbf{Q}$ method which involves the product of γ_1 and $\mathbf{h}(\cdot)$.

A methodological innovation in this area is to propose a new method for fitting non-latent Gaussian models in the INLA framework. Here, I work on the assumption that INLA is a strong choice for doing Bayesian inference, since it is established as fast and accurate. Other models where this can be useful, in addition to the models mentioned previously, are the following: spatial autoregressive combined (SAC) model (Manski, 1993), zero-inflated Poisson model, mixture models, and Bayesian Lasso.

The following are the existing approaches for fitting conditional INLA models: a hybrid MCMC-INLA approach (Gómez-Rubio and Rue, 2018), model averaging with INLA (Gómez-Rubio et al., 2020), importance sampling (IS) with INLA or an adaptive IS with INLA (Berild et al., 2022), and the linearized INLA (Lindgren et al., 2024). The idea of the first approach is that INLA is used to fit the model conditional on the parameters, say $\boldsymbol{\theta}$, which causes the violation of the latent Gaussian assumption; and then MCMC is used to estimate the posterior distribution of $\boldsymbol{\theta}$. The problem with this approach is that it is very slow. The second approach creates a grid of values for $\boldsymbol{\theta}$ and then fits the model using INLA conditional on each $\boldsymbol{\theta}$ value. All conditional INLA models are then averaged. However, this is only possible when we have an idea of the most plausible values of $\boldsymbol{\theta}$. Moreover, this can be computationally expensive when $\boldsymbol{\theta}$ is of high dimension. The third approach can be faster than the

MCMC-INLA approach because of the non-sequential nature of IS, but it requires good proposal distributions. The common ground among the three approaches is that they perform model averaging of all the conditional INLA models; they only differ on how $\boldsymbol{\theta}$ is sampled or determined. Finally, the fourth approach may suffer from some difficulties in convergence, depending on the degree of non-linearity in the predictor expression.

7.5.1 Future work

Let $\mathbf{y}$ be the observed data; and $\boldsymbol{\theta}$ be the model parameters – including both latent model parameters and hyperparameters. The goal is to obtain the posterior marginals, $\pi(\theta_i|\mathbf{y})$. Suppose that the model violates latent Gaussianity unless some parameters are fixed, say, $\boldsymbol{\theta}_c$. Let $\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_c & \boldsymbol{\theta}_{-c} \end{pmatrix}^\intercal$, where $\boldsymbol{\theta}_c$ is fixed so that INLA can then be used for inference. The posterior marginals of interest are then $\pi(\theta_{c,i}|\mathbf{y})$ and $\pi(\theta_{-c,i}|\mathbf{y})$.

My proposal is to fit the conditional models using a batched sequential importance sampling approach. It is batched since the approach can be easily parallelized. It is sequential since the goal is that subsequent samples should have increasing (model averaging) weights from a random start. It is based on the importance sampling method, since high posterior samples of $\boldsymbol{\theta}$ should be given higher weights when doing model averaging. The implementation of the importance sampling is explained in Section 7.5.1.1. The sequence of proposed values for $\boldsymbol{\theta}$ are determined using stochastic approximation (see Section 7.5.1.2), such as the central difference estimation method, of the posterior distribution of $\boldsymbol{\theta}$. The strength of the proposed method is that, although sampling is done sequentially and which requires the computation of gradients, the different sequences can be parallelized (see Section 7.5.1.3), and that for each sequence, the succeeding posterior samples have increasing weights (see Section 7.5.1.4).

7.5.1.1 Implementation of importance sampling approach

Suppose $g(\boldsymbol{\theta}_c)$ is the importance or proposal distribution for $\boldsymbol{\theta}_c$. Computation of the posterior marginals $\pi(\theta_{-c,i}|\mathbf{y})$ is given as follows:

$$\begin{aligned}\pi(\theta_{-c,i}|\mathbf{y}) &= \int \pi(\theta_{-c,i}|\boldsymbol{\theta}_c,\mathbf{y})\pi(\boldsymbol{\theta}_c|\mathbf{y})d\boldsymbol{\theta}_c \\ &= \int \pi(\theta_{-c,i}|\boldsymbol{\theta}_c,\boldsymbol{y})\frac{\pi(\boldsymbol{\theta}_c|\mathbf{y})}{g(\boldsymbol{\theta}_c)}g(\boldsymbol{\theta}_c)d\boldsymbol{\theta}_c \\ &\approx \sum_j \pi\big(\theta_{-c,i}|\boldsymbol{\theta}_c^{(j)},\mathbf{y}\big)\frac{\pi\big(\boldsymbol{\theta}_c^{(j)}|\mathbf{y}\big)}{g(\boldsymbol{\theta}_c^{(j)})}, \quad \boldsymbol{\theta}_c^{(j)} \sim g\big(\boldsymbol{\theta}_c\big).\end{aligned} \tag{7.1}$$

In Equation (7.1), $\pi(\theta_{-c,i}|\boldsymbol{\theta}_c,\mathbf{y})$ is the conditional posterior marginal of $\theta_{-c,i}$, $\pi(\boldsymbol{\theta}_c|\mathbf{y})$ is the posterior marginal of $\boldsymbol{\theta}_c$, and $\pi(\theta_{-c,i}|\mathbf{y})$ is estimated using model averaging. The exact form of $\pi(\boldsymbol{\theta}_c|\mathbf{y})$ is unknown; instead, it is known up to a proportionality constant, given by

$$\pi(\boldsymbol{\theta}_c|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta}_c)\pi(\boldsymbol{\theta}_c).$$

Thus, we have

$$\pi(\theta_{-c,i}|\mathbf{y}) \approx \sum_j \pi\big(\theta_{-c,i}|\boldsymbol{\theta}_c^{(j)},\mathbf{y}\big)w^{*(j)},$$

where $w^{*(j)}$ are the standardised weights $w^{(j)} = \dfrac{\pi\big(\boldsymbol{y}|\boldsymbol{\theta}_c^{(j)}\big)\pi\big(\boldsymbol{\theta}_c^{(j)}\big)}{g\big(\boldsymbol{\theta}_c^{(j)}\big)}$.

7.5.1.2 Stochastic Approximation of $\pi(\boldsymbol{\theta}_c|\mathbf{y})$

Note that we want good samples from $\pi(\boldsymbol{\theta}_c|\mathbf{y})$. This can be done by performing a stochastic approximation on $\pi(\boldsymbol{\theta}_c|\mathbf{y}) \propto \pi(\mathbf{y}|\boldsymbol{\theta}_c)\pi(\boldsymbol{\theta}_c)$, whose steps are presented in Algorithm 7.1.

Algorithm 7.1 Stochastic approximation of $\pi(\boldsymbol{\theta}_c|\mathbf{y})$

Step 1: Initialize $\boldsymbol{\theta}_c = \boldsymbol{\theta}_c^0 \in \mathbb{R}^p, \boldsymbol{\theta}_c^0 \sim g(\boldsymbol{\theta}_c)$.

Step 2: Obtain an estimated gradient of $\pi(\boldsymbol{\theta}_c^0|\mathbf{y}), \widehat{\nabla\pi(\boldsymbol{\theta}_c^0|\mathbf{y})}$

Step 3: Set $\boldsymbol{\theta}_c^1 = \boldsymbol{\theta}_c^0 + \boldsymbol{\beta}\widehat{\nabla\pi(\boldsymbol{\theta}_c^0|\mathbf{y})}$, for a fixed $\boldsymbol{\beta}$ which is the learning rate.

Step 4: For $j = 2, \ldots, N$,

$$\boldsymbol{\theta}_c^j = \boldsymbol{\theta}_c^{j-1} + \boldsymbol{\beta}\widehat{\nabla\pi(\boldsymbol{\theta}_c^{j-1}|\mathbf{y})}$$

Stop if a stopping criterion is met.

The next bottleneck is how to obtain $\widehat{\nabla\pi(\boldsymbol{\theta}_c^j|\mathbf{y})}$. This can be done using the central difference estimation approach, which is presented in Algorithm 7.2. Since $\pi(\boldsymbol{\theta}_c^j|\mathbf{y})$ is known only up to a constant, the computed $\widehat{\nabla\pi(\boldsymbol{\theta}_c^j|\mathbf{y})}$ should be normalized.

Algorithm 7.2 Central difference estimation approach to obtain $\widehat{\nabla\pi(\boldsymbol{\theta}_c^j|\mathbf{y})}$

Suppose p is the dimension of $\boldsymbol{\theta}_c$. Let $\boldsymbol{e}_i$ be a p-dimensional vector with 1 in the ith position and 0 elsewhere.

Step 1: Do for $i = 1, \ldots, p$:

$$\text{Compute } L = \pi\left(\boldsymbol{\theta}_c^j - \boldsymbol{e}_i\frac{\delta}{2}\right) \text{ and } R = \pi\left(\boldsymbol{\theta}_c^j + \boldsymbol{e}_i\frac{\delta}{2}\right)$$

$$\frac{\hat{\partial}\pi(\boldsymbol{\theta}_c^j)}{\partial\theta_{c,i}^j} = \frac{R-L}{\delta}$$

Step 2: $\widehat{\nabla\pi(\boldsymbol{\theta}_c^j|\mathbf{y})} = \left(\frac{\hat{\partial}\pi(\boldsymbol{\theta}_c^j)}{\partial\theta_{c,1}^j} \quad \frac{\hat{\partial}\pi(\boldsymbol{\theta}_c^j)}{\partial\theta_{c,2}^j} \quad \ldots \quad \frac{\hat{\partial}\pi(\boldsymbol{\theta}_c^j)}{\partial\theta_{c,p}^j}\right)^{\intercal}$

The learning rate $\boldsymbol{\beta}$ in Algorithm 7.1 is pre-determined. But it could also be adaptively learned. The determination of the value $\boldsymbol{\beta}$ should be carefully chosen. If $\boldsymbol{\beta}$ is too small, it might take very long for the stochastic approximation to converge. If $\boldsymbol{\beta}$ is too large, we might not have done a good search over the $\boldsymbol{\theta}_c-$space. Moreover, the constant δ is fixed at a small number. A sensible stopping criterion is if there is a decline in the value of the weight $w^{(j)}$, i.e., stop if

$$\frac{\pi\left(\mathbf{y}|\boldsymbol{\theta}_c^{j+1}\right)\pi\left(\boldsymbol{\theta}_c^{j+1}\right)}{g\left(\boldsymbol{\theta}_c^{j+1}\right)} < \frac{\pi\left(\mathbf{y}|\boldsymbol{\theta}_c^{j}\right)\pi\left(\boldsymbol{\theta}_c^{j}\right)}{g(\boldsymbol{\theta}_c^{j})}.$$

The algorithm is not quite efficient since it computes the gradient for every sample/particle. With the assumption that $\pi(\boldsymbol{\theta}_c|\mathbf{y})$ is unimodal, then a potential solution is to compute the gradient only at the random start. However, this is still not quite efficient because of the sequential updates and a random stopping time. Since we are assured that the weights are increasing, then it might be smart to specify a reasonable number of steps to take from the random start in a single sequence. A solution then is to identify the equidistant samples/particles in a sequence given the fixed number of steps. The benefit of doing this is that we can perform parallel (INLA) computation (for each sample), and there is no need to specify a stopping rule.

If the random start is far from the mode $\pi(\boldsymbol{\theta}_c|\mathbf{y})$, the obtained samples for a single sequence will eventually have negligible weights. Thus, a waste of memory and computing time. A solution here is to define a (line) segment from a random start to our estimated mode, say $\boldsymbol{\theta}_c^*$. The steps are presented in Algorithm 7.3.

Algorithm 7.3 Modified algorithm for obtaining $\boldsymbol{\theta}_c$ samples

Step 1: Initialize $\boldsymbol{\theta}_c = \boldsymbol{\theta}_c^0 \in \mathbb{R}^p, \boldsymbol{\theta}_c^0 \sim g(\boldsymbol{\theta}_c)$.

Step 2: Obtain an estimated (normalized) gradient of $\pi(\boldsymbol{\theta}_c^0|\boldsymbol{y}), \widehat{\nabla\pi(\boldsymbol{\theta}_c^0|\mathbf{y})}$

Step 3: Compute the distance, d, between $\boldsymbol{\theta}_c^0$ and $\boldsymbol{\theta}_c^*$.

Step 4: Compute N equidistant points from 0 to d:

$$\boldsymbol{\beta} = \begin{pmatrix} \beta_1 & \beta_2 & \dots & \beta_j & \dots & \beta_N \end{pmatrix}$$

Step 5: For $j = 1, \dots, N$: $\boldsymbol{\theta}_c^j = \boldsymbol{\theta}_c^0 + \beta_j \widehat{\nabla\pi(\boldsymbol{\theta}_c^0|\mathbf{y})}$

7.5.1.3 Batch processing

The stochastic approximation of $\pi(\boldsymbol{\theta}|\mathbf{y})$ can be easily parallelized. Suppose that we have the following:

$\begin{pmatrix} \boldsymbol{\theta}_c^{0,1} & \boldsymbol{\theta}_c^{1,1} & \dots & \boldsymbol{\theta}_c^{t,1} & \dots & \boldsymbol{\theta}_c^{N_1,1} \end{pmatrix}$ are the N_1 samples from the 1st batch.

$\vdots$

$\begin{pmatrix} \boldsymbol{\theta}_c^{0,b} & \boldsymbol{\theta}_c^{1,b} & \dots & \boldsymbol{\theta}_c^{j,b} & \dots & \boldsymbol{\theta}_c^{N_b,b} \end{pmatrix}$ are the N_b samples from the bth batch.

$\vdots$

$\begin{pmatrix} \boldsymbol{\theta}_c^{0,B} & \boldsymbol{\theta}_c^{1,B} & \dots & \boldsymbol{\theta}_c^{t,B} & \dots & \boldsymbol{\theta}_c^{N_B,B} \end{pmatrix}$ are the N_B samples from the Bth batch.

Given the samples $\boldsymbol{\theta}_c^{(t,b)}, \quad t = 1, \dots, N_b; \quad b = 1, \dots, B$, how do I use these to obtain the posterior estimates of $\boldsymbol{\theta}$?

Estimating $\pi(\boldsymbol{\theta}_{\text{-}c}|\mathbf{y})$:

All samples $\begin{pmatrix} \boldsymbol{\theta}_c^{0,b} & \boldsymbol{\theta}_c^{1,b} & \dots & \boldsymbol{\theta}_c^{t,b} & \dots & \boldsymbol{\theta}_c^{N_b,b} \end{pmatrix}, b = 1, 2, \dots, B$ can be combined

via model averaging to estimate $\pi(\theta_{\text{-}c,i}|\mathbf{y})$. This is given as follows:

$$\pi(\theta_{\text{-}c,i}|\mathbf{y}) \approx \sum_{b=1}^{B}\sum_{j=0}^{N_b} \pi(\theta_{\text{-}c,i}|\boldsymbol{\theta}_c^{j,b}, \mathbf{y}) \frac{w^{j,b}}{\sum_{b=1}^{B}\sum_{j=0}^{N_b} w^{j,b}}$$
$$w^{j,b} = \frac{\pi(\mathbf{y}|\boldsymbol{\theta}_c^{j,b})\pi(\boldsymbol{\theta}_c^{j,b})}{g(\boldsymbol{\theta}_c^{j,b})}$$

Estimating $\pi(\boldsymbol{\theta}_c|\mathbf{y})$:

The joint posterior of $\boldsymbol{\theta}_c$ can be estimated as

$$\pi(\boldsymbol{\theta}_c|\mathbf{y}) \approx \sum_{b=1}^{B}\sum_{j=0}^{N_b} w^{j,b}\delta(\boldsymbol{\theta}_c - \boldsymbol{\theta}_c^{j,b}),$$

where δ is the Dirac delta function. Another approach is weighted nonparametric kernel density estimation. The posterior marginals of $\theta_{c,i}$ are estimated using similar approaches.

7.5.1.4 Computation of weights

A critical component of the proposed solution is the computation of the weights, given by $w^{j,b}, j = 0, \ldots, N_b; b = 1, \ldots, B$. To do this, I use the fact that for an arbitrary function $h(\cdot)$ of $\boldsymbol{\theta}_c$, the Monte Carlo estimator for $h(\boldsymbol{\theta}_c)$ is given by

$$\begin{aligned}\mathbb{E}\big[h(\boldsymbol{\theta_c})|\mathbf{y}\big] = \int h(\boldsymbol{\theta}_c)\pi(\boldsymbol{\theta}_c|\mathbf{y})d\boldsymbol{\theta}_c &\approx \frac{1}{\sum_{b=1}^{B}(N_b+1)}\sum_{b=1}^{B}\sum_{j=0}^{N_b} h(\boldsymbol{\theta}_c^{j,b})w^{j,b} \\ &= \frac{1}{B}\sum_{b=1}^{B}\left(\frac{1}{N_b+1}\sum_{j=0}^{N_b} h(\boldsymbol{\theta}_c)^{j,b}w^{j,b}\right),\end{aligned}$$

which can be written as

$$\mathbb{E}\big[h(\boldsymbol{\theta_c})|\mathbf{y}\big] \approx \frac{1}{B}\left[\left(\frac{h(\boldsymbol{\theta}_c^{0,1})w^{0,1}}{N_1+1} + \ldots + \frac{h(\boldsymbol{\theta}_c^{0,B})w^{0,B}}{N_B+1}\right) + \sum_{b=1}^{B}\sum_{j=1}^{N_b}\frac{h(\boldsymbol{\theta}_c^{j,b})w^{j,b}}{N_b+1}\right].$$

Thus, we have

$$\mathbb{E}\big[h(\boldsymbol{\theta_c}|\mathbf{y})\big] \approx \frac{1}{B}\left[\sum_{b=1}^{B}\frac{h(\boldsymbol{\theta}_c^{0,b})w^{0,b}}{N_b+1} + \sum_{b=1}^{B}\sum_{j=1}^{N_b}\frac{h(\boldsymbol{\theta}_c^{j,b})w^{j,b}}{N_b+1}\right].$$

7.5.1.5 Next steps

The next steps from here involve the actual computation of the weights, as discussed in Section 7.5.1.4. The benchmark approaches here would be the importance sampling approach (Berild et al., 2022), INLA-with-MCMC approach (Gómez-Rubio and Rue, 2018), and the linearized INLA approach (Lindgren et al., 2024).

7.6 On the problem of time misalignment

Another feature of the data, which I did not consider but is related to the change of support problem, is time misalignment, i.e., the time dimension of the data are observed at different frequencies. In this thesis, the dengue cases are observed at the weekly level, which I aggregated to the monthly level. Moreover, it is possible to request from PAGASA for daily outcomes of the climate variables. Another example is in econometric applications, where Gross Domestic Product is commonly calculated quarterly, while currency exchange rates are constantly fluctuating. In the aforementioned examples, the dependent variable is low frequency, while the predictor is high frequency. I propose to address this problem using Mixed Data Sampling (MIDAS) models (Ghysels et al., 2020).

7.6.1 Mixed Data Sampling

Let $t \in \mathbb{N}$ index the low frequency observations of a dependent variable $y_t \in \mathbb{R}$.

Let $\tau \in \mathbb{N}$ index the high frequency observations.

Let $x_\tau \in \mathbb{R}$ be the single high-frequency observation.

Let m_t denote the number of high-frequency observations pertaining to the t^{th} low-frequency observation.

Let $s(t) = \sum_{j=1}^{t} m_j$ denote the total number of high-frequency periods available up till (and including) the t^{th} low frequency observation.

The stylized MIDAS regression model is given by

$$y_t = g\left(\sum_{i=0}^{k} w_i x_{s(t)-i}; \boldsymbol{\beta}\right) + \varepsilon_t \tag{7.2}$$

$$\forall i, w_i = h(\boldsymbol{\gamma}, i) \text{ and } \sum_{i=0}^{k} w_i = 1 \tag{7.3}$$

The function $g : \mathbb{R} \to \mathbb{R}$ can be parametric, in which case $\boldsymbol{\beta} \in \mathbb{R}^b$ is its low-dimensional parameter vector. When $g(z; \boldsymbol{\beta}) = g(z) = z$, $\forall z \in \mathbb{R}$, then we have the so-called distributed lag MIDAS (DL-MIDAS) regression. Furthermore, the function $g : \mathbb{R} \to \mathbb{R}$ can be non-parametric, in that case $\boldsymbol{\beta} = 1$ is imposed. The zero mean error term ε_t is independent of x_t, and also identically distributed. The normalization condition $\sum_{i=0}^{k} w_i = 1$ is often required for the identification of $\boldsymbol{\beta}$ and/or g. The function $h(\cdot)$ is a constraint function which aligns the low-frequency and high-frequency observations. In particular, we have $h(\boldsymbol{\gamma}, i) : \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}$, where $\boldsymbol{\gamma} \in \mathbb{R}^d$ and i is the lag index $i \in \{0, 1, \ldots, k\}$. The values w_i are constrained to sum to 1 so that $\boldsymbol{\beta}$ or g$(\cdot)$ are identifiable.

7.6.1.1 Constraint functions

Since $\{w_i\}_{i=0,\ldots,k}$ are restricted to add to one, it is convenient to represent the functional constraint $h(\cdot)$ in the following form:

$$w_i = h(\boldsymbol{\gamma}, i) = \frac{\psi(\boldsymbol{\gamma}, i)}{\sum_{j=0}^{k} \psi(\boldsymbol{\gamma}, j)}. \tag{7.4}$$

The choice of the $\psi : \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}$ determines the shape of h. Here, d is the dimension of γ. Three widely used parametric forms of ψ are the following:

1. exponential Almon polynomials: $\psi(\boldsymbol{\gamma}, i) = \exp\left(\sum_{j=1}^{d} \gamma_j i^j\right)$
2. beta polynomial: $\psi(\boldsymbol{\gamma}, i) = x_i^{\gamma_1 - 1}(1 - x_i)^{\gamma_2 - 1}$, where $x_i = \xi + (1 - \xi)\dfrac{i-1}{k-1}$ and marginally small quantity $\xi > 0$. A special case is to fix $\gamma_1 = 1$, so that $d = 1$, and which still provide a flexible constraint (Ghysels and Qian, 2019).
3. hyperbolic scheme polynomial: $\psi(\boldsymbol{\gamma}, i) = \dfrac{\Gamma(i + \gamma)}{\Gamma(i+1)\Gamma(\gamma)}$, where $\Gamma(\cdot)$ is the gamma function.

Other constraint functions are provided in Ghysels et al. (2016). Figure 7.1 illustrates the values of the weights for the three constraint functions. For the exponential Almon polynomial (see Figure 7.1a), declining weights with respect to the lag

are guaranteed as long as $\gamma_2 \leq 0$. Weights within the vicinity of zero seem to be reasonable values for the $\boldsymbol{\gamma}$ parameter (Ghysels et al., 2007). Figure 7.1b illustrates the weights for the beta polynomial where γ_1 is fixed at 1. When $\gamma_1 = 1$, the weights are declining when $\gamma_2 > 1$, and the rate of decline depends on its magnitude, particularly, higher values of γ_2 give rapidly declining weights. Figure 7.1c shows weights for the hyperbolic scheme polynomial. The reasonable values of γ for this constraint function is $0 < \gamma < 1$.

For the constrained MIDAS models, the selection of the constraint function h, the dimension of $\boldsymbol{\gamma}$, and the lag order k has to be carefully considered. For the unconstrained MIDAS, there is no constraint on the parameters; thus, the selection of k corresponds directly to the selection of number of parameters in this regression.

7.6.2 Proposed estimation approach

The MIDAS model, as introduced in Section 7.6.1, cannot be directly implemented in INLA. However, model reparameterization can be the strategy in order to use the INLA framework with the MIDAS model. This involves writing `generic models` in the `INLA` library. A `generic model` is a way to define latent model components that are not yet available in the `INLA` library.

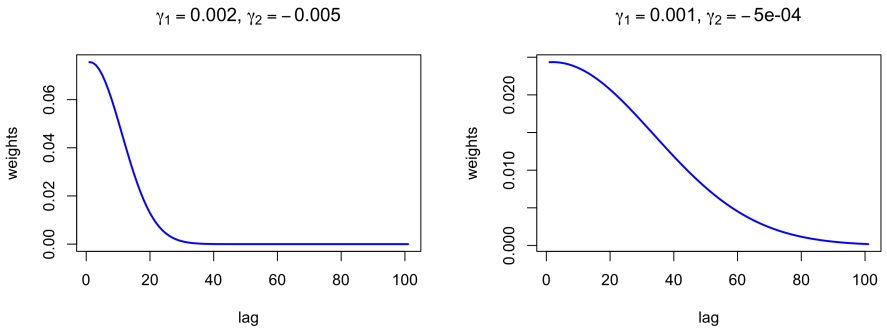
7.6.3 Toy examples

This section presents initial results from simulated toy datasets. Sections 7.6.3.1 and 7.6.3.2 present some results for a constrained MIDAS model with exponential Almon polynomial constraint and a hyperbolic scheme polynomial constraint, respectively.

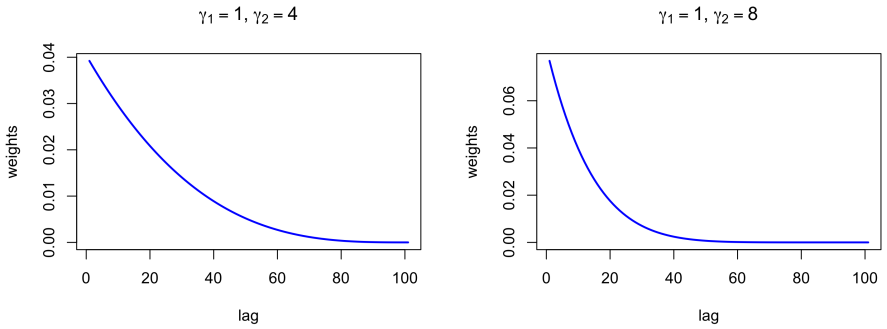
7.6.3.1 Constrained MIDAS with exponential Almon polynomials

Here, I simulate a data from a constrained MIDAS model with the exponential Almon polynomial as the constraint function. I set $d = 2, \gamma_1 = 0.006$, and $\gamma_2 = -0.0005$. Moreover, I assume $s(t) = 5, k = 4, \beta_0 = 3$, $\beta_1 = 1$, and $\sigma_e = 1$. Shown in Figure 7.2 is a simulated data from the model, for a time series data of length $T = 3000$.

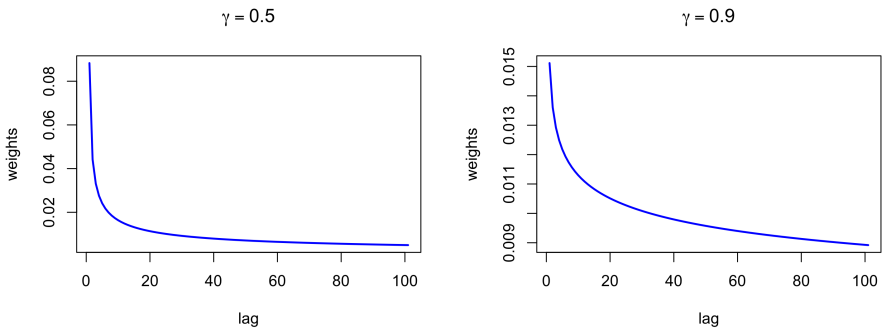
Figure 7.3 shows the posterior estimates of three parameters: β_0, β_1, and $\tau_e = 1/\sigma_e^2$. The results shows that these parameters are correctly estimated. All the



(a) exponential Almon polynomial weights

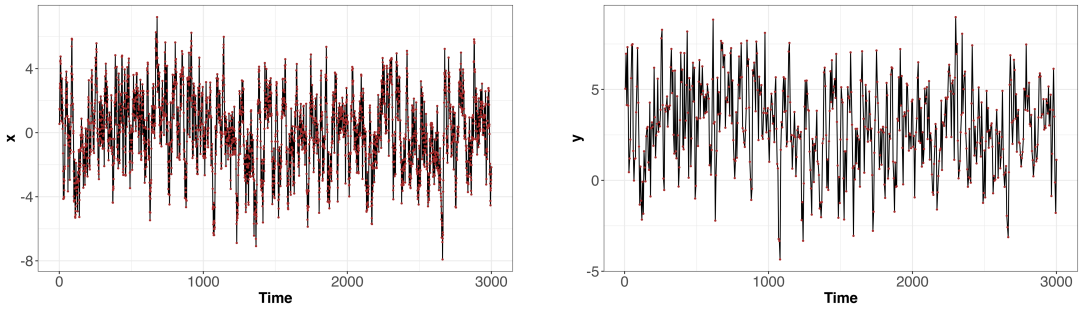


(b) Beta polynomial weights



(c) Hyperbolic scheme polynomial weights

Figure 7.1: Illustration of the weights values for the three constraint functions



(a) High frequency predictor

(b) Low frequency response variable

Figure 7.2: Simulated data from a constrained MIDAS model with the exponential Almon polynomial as constraint function

posterior means are inside the respective 95% credible intervals. Moreover, Figure 7.4 shows the estimated posterior mean and 95% credible intervals for the weights $w_i, i = 0, 1, \ldots, 4$. This shows that the model (lag) weights are also correctly estimated. Finally, Figure 7.5 shows a close correspondence between the true values and predicted values of the response varible y.

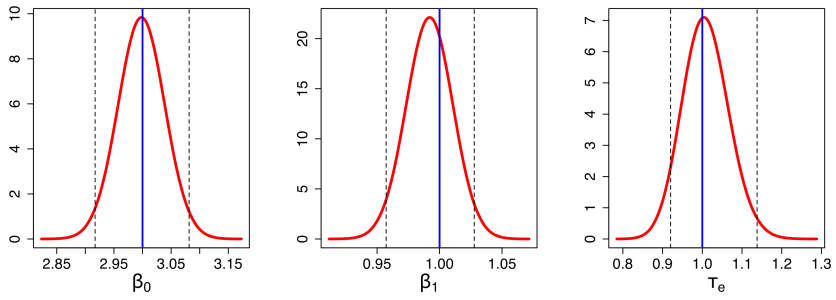


Figure 7.3: Posterior estimates of model parameters. Shown in blue line is the true value, while the shaded lines show the 95% credible intervals.

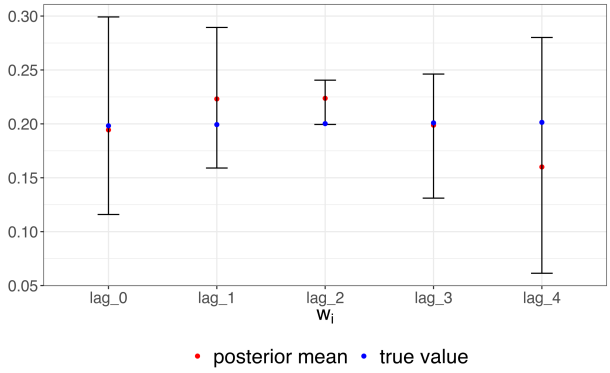


Figure 7.4: Posterior estimates of the lag weights, w_i. The line segment is the 95% credible interval

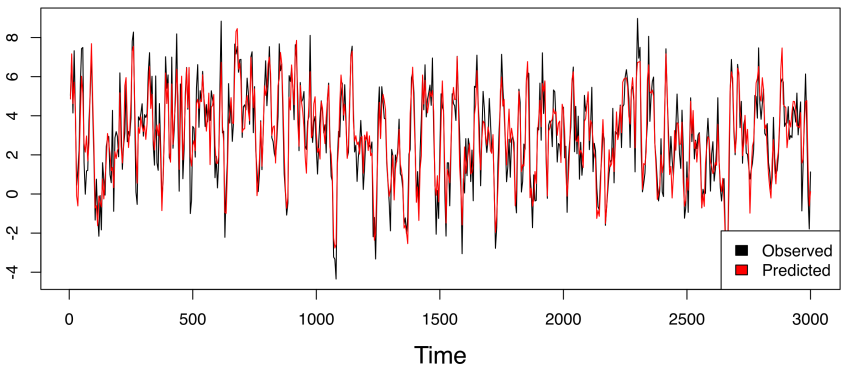


Figure 7.5: Observed versus predicted values of the response variable y

7.6.3.2 Constrained MIDAS with hyperbolic scheme polynomial

Here, I simulate a data from a constrained MIDAS model with the hyperbolic scheme polynomial as the constraint function. I also assume $k = 4, \beta_0 = 3$, $\beta_1 = 1$, and $\sigma_e = 1$. The constraint function parameter is $\gamma = 0.5$.

Shown in Figure 7.6 are the estimated posterior marginals for the parameters $\beta_0, \beta_1, \tau_e = 1/\sigma_e^2$, and γ. The results show that the parameters are correctly estimated. Moreover, Figure 7.7 shows the estimated weights $w_i, i = 0, 1, \ldots, 4$. All the (lag) weights are also correctly estimated. Finally, Figure 7.8 shows a close correspondence between the true values and predicted values of the response variable.

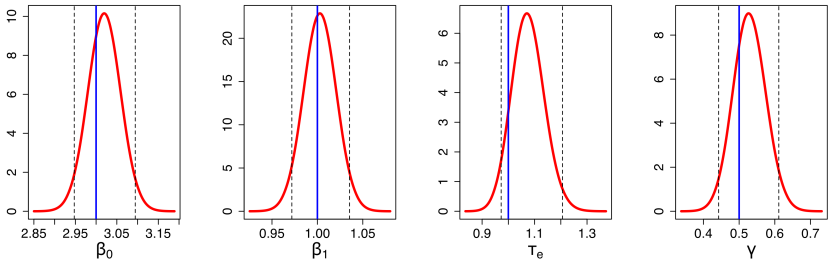


Figure 7.6: Posterior estimates of model parameters. Shown in blue line is the true value, while the shaded lines show the 95% credible intervals.

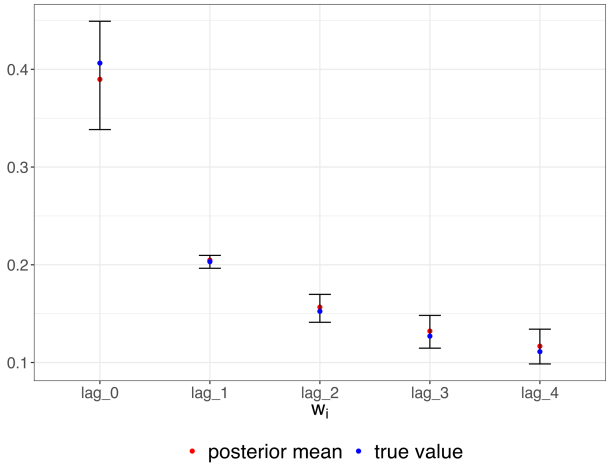


Figure 7.7: Posterior estimates of the lag weights, w_i. The line segment is the 95% credible interval

7.6.4 Next steps

The next step is to investigate further the capabilities of INLA in fitting other types of MIDAS regression models, such as models which include a trend component. The

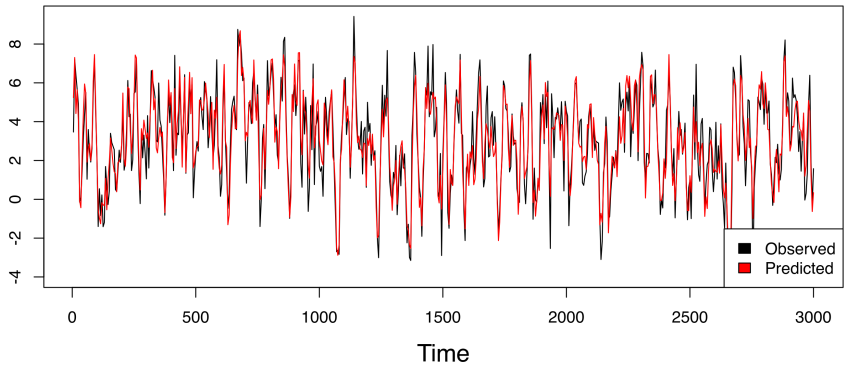


Figure 7.8: Observed versus predicted values of the response variable y

defined `rgeneric` functions worked in toy examples. These functions can be translated into `cgeneric` functions, i.e., to define these new latent models in `C` in order to get the correct speed. Further next steps would be to apply the models in real-life data, and to build an `R` package to make these functions accessible to practitioners and applied users.

7.7 On the link between climate and dengue in the Philippines

Finally, this thesis extensively looked at the association between climate and dengue in the Philippines. This is done in a two-stage modelling framework, wherein the first stage involves fitting climate models, as presented in Chapter 4, and the second stage involves fitting the health model using the climate predictions as an input and the dengue incidence as the response variable, as presented in Chapter 5. The entire process of doing the statistical analysis exactly follows the two-stage Bayesian modelling framework illustrated in Figure 1.6.

This work postulates a spatio-temporal model with several random effects to capture extra variation in the data unexplained by the climate covariates and other fixed effects. The random effects are specified using both structured and unstructured effects in space and time, and their interaction. The novelty in the application is the use of a methodological approach which is new and different compared to existing

published work using Philippines data, as emphasized in Chapter 5. The results agree with existing studies on climate and dengue. Temperature is shown to have a non-linear relationship with dengue. An increase in temperature implies an increase in dengue incidence and risks. However, too high temperature is shown to have a negative association with dengue. This result is expected since studies have shown that excessively high temperature can shorten the lifespan of mosquitoes and cause a reduction in their population size (Myer et al., 2020). Moreover, the results show that the association between dengue and rainfall and relative humidity varies in space. For the eastern section of the country, which is relatively wet all year round and with low variation in the amount of rainfall, the effect of rainfall is negative. On the other hand, for most western section of the country, which has a pronounced wet and dry season, the relationship is the opposite, i.e., dengue and rainfall are positively related. These are exactly the same results shown in Cawiding et al. (2025), and this is explained by the fact that consistent and low variation in the amount of rainfall tends to wash away breeding sites of mosquitoes, while intermittent rainfall, especially during dry season, tends to create more breeding sites.

The extensive analysis in Chapter 6 performed a comparison between a crude plug-in method and the resampling method. The results show that the posterior uncertainty in the second-stage mode parameters are generaly larger with the resampling method.

7.7.1 Future work

7.7.1.1 Inclusion of social, economic, and other factors

One extension of the work is to account for social and economic factors of dengue incidence, such as health infrastructure, poverty incidence, housing, and level of urbanization. Other factors include human behavior such as travel, water storage practices, and land use. Another extension is to consider lagged effects of climate variables and to include covariates at the pathogen-levels, which includes mosquito abundance and biting rates.

7.7.1.2 Variance partitioning approach

An innovation in the model specification is to use the variance partitioning approach proposed in Franco-Villoria et al. (2022). The proposed models are basically reparametrized versions of the spatio-temporal models in Knorr-Held (2000), which are based on Kronecker product of intrinsic Gaussian Markov random fields (IGMRF). This approach introduces a mixing parameter that distributes the total variability (generalized variance) between the main and interaction effects.

More formally, suppose that η_{ij} is the linear predictor for the i^{th} time point and j^{th} spatial location. Moreover, suppose that β_{1_i} is the temporal main effect, β_{2_j} is the spatial main effect, and δ_{ij} is the interaction effect which is modelled as a Kronecker product IGMRF. The variance partitioning approach specifies η_{ij} as follows:

$$\eta_{ij} = \alpha + \sqrt{\tau^{-1}}\left[\sqrt{1-\gamma}\left(\sqrt{1-\phi}\beta_{1_i} + \sqrt{\phi}\beta_{2_j}\right) + \sqrt{\gamma}\delta_{ij}\right] \tag{7.5}$$

In Equation (7.5), τ is the overall (generalized) variance, which is distributed between the main effects and the interaction effect via the mixing parameter $\gamma \in [0, 1]$. The main effect is further distributed between the temporal effect β_{1_i} and the spatial effect β_{2_j} via the mixing parameter $\phi \in [0, 1]$. This model specification makes it easy to quantify the relative contribution of the different random effects components to the total variance. The advantage of the variance partitioning approach is that it provides a way of eliciting the prior in a very intuitive way, via the use of the penalized complexity (PC) prior on γ. This construction guarantees parsimony and avoids overfitting, since the prior is based on the assumption that the interaction term shrinks to zero. A Dirichlet prior can also be used for ϕ (Fuglstad et al., 2020), whose base model assumes equal weights to both the temporal and spatial terms, i.e., the prior assumes ignorance about how the variance is distributed.

Equation 7.5 can be further extended to the case when the spatial and temporal main effects are defined with both unstructured and structured effects, such as the random walk model and the intrinsic CAR model. Suppose that ϵ_{1_i} is the unstructured temporal effect, and ϵ_{2_j} is the unstructured spatial effect. The linear predictor

is now specified as

$$\eta_{ij} = \alpha + \sqrt{\tau^{-1}}\Big[\sqrt{1-\gamma}\Big(\sqrt{1-\phi}\Big(\sqrt{1-\psi_1}\beta_{1_i} + \sqrt{\psi_1}\epsilon_{1_i}\Big) + \sqrt{\phi}\Big(\sqrt{1-\psi_2}\beta_{2_j} + \sqrt{\psi_2}\epsilon_{2_j}\Big)\Big) + \sqrt{\gamma}\delta_{ij}\Big]. \tag{7.6}$$

In Equation 7.6, $\psi_1 \in [0,1]$ and $\psi_2 \in [0,1]$ are mixing parameters which distributes the variability of the temporal and spatial main effects, respectively, to the structured and unstructured effects. A PC prior can also be used on both ψ_1 and ψ_2 (Riebler et al., 2016), where the base model is the model without structured effects.

7.7.1.3 New specification of the Poisson model

Another area of future work is the application of the new specification of the Poisson model, as presented in Section 6.3.2, in a spatio-temporal context, such as Chapter 5. This is given as follows:

$$\begin{aligned} \mathrm{y}(B,t) &\sim \text{Poisson}\Big(\mu(B,t)\Big) \\ \mathbb{E}\big[\mathrm{y}(B,t)\big] &= \mu(B,t) = \lambda(B,t) \times \mathrm{E}(B,t) \\ \log\Big(\lambda(B,t)\Big) &= \log\left(\frac{1}{|B|}\int_B \lambda(\mathbf{s},t)d\mathbf{s}\right) \\ \lambda(\mathbf{s},t) &= \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t) + \varphi(\mathbf{s},t)\Big\}, \end{aligned} \tag{7.7}$$

where $\mathrm{y}(B,t)$ is the observed cases, $\mathrm{E}(B,t)$ is the expected number of cases, $\lambda(B,t)$ is the disease risk, $x(\mathbf{s},t)$ is the covariate of interest, say the temperature field, and $\varphi(\mathbf{s},t)$ is the spatio-temporal random effect. Equation 7.7 assumes that the disease risk at a block B and time t is an aggregation of a continuously-indexed risk/intensity function $\lambda(\mathbf{s},t)$ over B at time t. The risk function is then linked to the covariate field $x(\mathbf{s},t)$ in a non-linear fashion.

Interpreting γ_1

Since the specification of the joint model for the health outcomes models the Poisson outcomes at the intensity level, then it is of interest to look at how we interpret γ_1 under this non-traditional specification. I first consider the case without

spatio-temporal effects, $\varphi(\boldsymbol{s}, t)$, for simplicity, i.e,

$$\lambda(B,t) = \frac{1}{|B|}\int_B \lambda(\mathbf{s},t)d\mathbf{s} = \frac{1}{|B|}\int_B \exp\left\{\gamma_0 + \gamma_1 x(\mathbf{s},t)\right\}d\mathbf{s}.$$

Suppose $x(\mathbf{s},t)$ increases by one unit, then

$$\begin{aligned}
\lambda^*(B,t) &= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1\big(x(\mathbf{s},t)+1\big)\Big\}d\mathbf{s} \\
&= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t)\Big\}\exp\{\gamma_1\}d\mathbf{s} \\
&= \exp\{\gamma_1\}\frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t)\Big\}d\mathbf{s} \\
&= \exp\{\gamma_1\}\lambda(B,t).
\end{aligned}$$

We interpret $\exp\{\gamma_1\}$ as the multiplicative change in $\lambda(B,t)$ for a one unit increase in $x(\mathbf{s},t)$ for all $\mathbf{s} \in B$. Now, suppose $x(\mathbf{s},t)$ is log-transformed. If $x(\mathbf{s},t)$ increases by a factor of q, then

$$\begin{aligned}
\lambda^*(B,t) &= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 \log\big(x(\mathbf{s},t)\times q\big)\Big\}d\mathbf{s} \\
&= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 \log\big(x(\mathbf{s},t)\big) + \gamma_1\log(q)\Big\}d\mathbf{s} \\
&= \exp\Big\{\gamma_1\log(q)\Big\}\frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 \log\big(x(\mathbf{s},t)\big)\Big\}d\mathbf{s} \\
&= \exp\Big\{\gamma_1\log(q)\Big\}\lambda(B,t).
\end{aligned}$$

Thus, we say that multiplying $x(\mathbf{s},t)$ by a factor q is associated with a multiplicative change in $\lambda(B)$ of $\exp\Big\{\gamma_1\log(q)\Big\}$.

Specifying the spatio-temporal effect

In specifying the spatio-temporal effects $\varphi(\mathbf{s},t)$, we can decompose it as follows:

$$\varphi(\mathbf{s},t) = \varphi^{(c)}(\mathbf{s},t) + \varphi^{(v)}(\mathbf{s},t), \tag{7.8}$$

where $\varphi^{(c)}(\mathbf{s},t)$ is the part which is constant within B and $\varphi^{(v)}(\mathbf{s},t)$ is the part which varies within. For the notation, we say that $\varphi^{(c)}(\mathbf{s},t) = \varphi(B,t)$ when $\mathbf{s} \in B$. Examples of the effects which falls under $\varphi^{(c)}(\mathbf{s},t)$ are iCAR random effects and area-specific

unstructured effects. Effects which falls under $\varphi^{(v)}(\mathbf{s}, t)$ are any continuously-indexed spatial processes, such as those which are derived using SPDE approaches.

The form of $\lambda(B, t)$ is then as follows:

$$\begin{aligned}
\lambda(B,t) &= \frac{1}{|B|}\int_B \lambda(\mathbf{s},t)d\mathbf{s} \\
&= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t) + \varphi(\mathbf{s},t)\Big\}d\mathbf{s} \\
&= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t) + \varphi^{(c)}(\mathbf{s},t) + \varphi^{(v)}(\mathbf{s},t)\Big\}d\mathbf{s} \\
&= \frac{1}{|B|}\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t) + \varphi(B,t) + \varphi^{(v)}(\mathbf{s},t)\Big\}d\mathbf{s} \\
&= \frac{1}{|B|}\exp\Big\{\varphi(B,t)\Big\}\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\boldsymbol{s},t) + \varphi^{(v)}(\mathbf{s},t)\Big\}d\mathbf{s}
\end{aligned}$$

This implies that the predictor expression is of the following form:

$$\log\Big(\lambda(B,t)\Big) = \frac{1}{|B|}\log\left(\int_B \exp\Big\{\gamma_0 + \gamma_1 x(\mathbf{s},t) + \varphi^{(v)}(\mathbf{s},t)\Big\}d\mathbf{s}\right) + \varphi(B,t)$$

This has an implication when doing model fitting, since the effects which are constant within a block B can either be included inside the $\exp(\cdot)$ term and be included in the integration scheme, or be excluded from the integral.

7.8 Final Summary

This PhD thesis proposes and applies a framework for two-stage modelling in spatial epidemiology. In the first stage, I propose the use of data fusion models to improve model predictions of exposures and covariates whose association with the health outcome is of interest, such as climate variables and air pollution concentration. Chapter 3 illustrated the use of a data augmentation approach with INLA-SPDE to fit a data fusion model within a two-stage framework, while Chapter 4 proposed a flexible data fusion model and demonstrated the benefits of adopting such an approach in terms of improved prediction accuracy and parameter estimation. However, a key limitation of these models is that they can be complex and computationally demanding. There remains considerable scope for improving model specification, enhancing Bayesian algorithms, and accelerating computation – especially in today's era of abundant data

to inform physical and environmental processes.

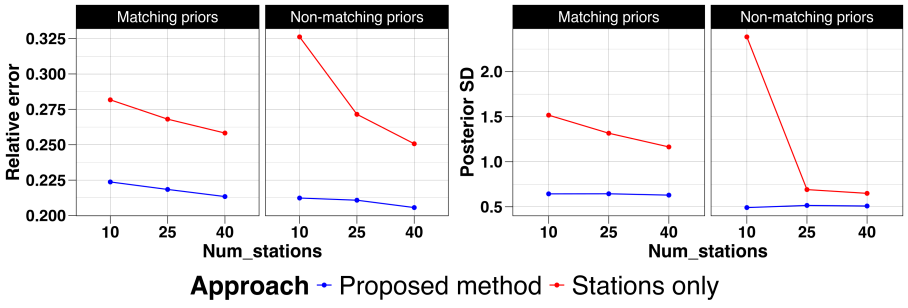
In the second stage, I fit the health model using the first-stage model results/predictions as inputs. In fitting the second-stage model, the uncertainty from the first-stage model needs to be correctly propagated; otherwise, the posterior standard deviations of second-stage model parameters will be underestimated. These were formally discussed in Chapter 6. The validation of Bayesian algorithms in terms of correct uncertainty propagation was done using the simulation-based calibration method, which tests for the self-consistency property of Bayesian models. The results showed that the plug-in method, which ignores the uncertainty, gives incorrect posterior estimates, while the posterior sampling approach is correct. This PhD work also proposed a new method for doing uncertainty propagation, called the **Q** uncertainty method, and, in addition, a low rank approximation of the method. The advantage of the proposed method is that it does not require fitting the second-stage model several times; hence, can be more computationally efficient. However, the benefits from the method depend on the dimension of **Q** and the non-linearity inherent in the predictor expression of the second-stage model. In this work, the primary criterion used to validate the Bayesian algorithms was the self-consistency property of Bayesian models. Exploring alternative validation strategies, such as employing different test functions within an SBC framework, is an exciting direction for future research.

Although the two-stage framework proposed in this work was applied in the context of spatial epidemiology – specifically in linking climate and dengue in the Philippines – it is equally relevant in other applications, such as survival analysis, where longitudinal biological characteristics and biomarkers are first modelled, and the resulting predictions are then used as inputs in a survival model. This PhD thesis formalizes these ideas, with particular emphasis on uncertainty propagation, which is of paramount importance. Substantial scopre remains for future research, particularly in the development of more efficient methods for uncertainty propagation in two-stage models, and in multi-stage models more generally.

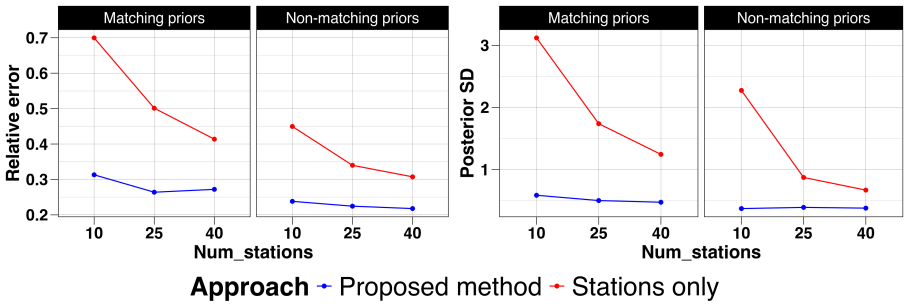
Appendix A

Appendix for Chapter 4

A.1 Simulation study

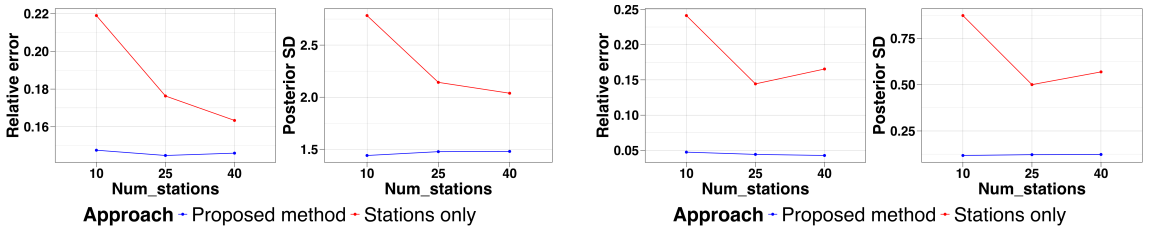


(a) σ_ξ



(b) ρ_ξ

Figure A.1: Plot of average relative errors and average posterior uncertainty from 500 simulated datasets for two hyperparameters: (a) marginal standard deviation σ_ξ of the spatial field and (b) range parameter ρ_ξ of the spatial field.



(a) Average relative errors and average posterior uncertainty in β_0

(b) Average relative errors and average posterior uncertainty in β_1

Figure A.2: Plot of average relative errors and average posterior uncertainty from 500 simulated datasets for the fixed effects: (a) β_0 and (b) β_1.

Table A.1: Posterior estimates of hyperparameters for the temperature model – stations-only model versus proposed data fusion model

	Stations only				Proposed model			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ_{e_1}	0.178	0.011	0.158	0.197	0.243	0.010	0.224	0.264
σ_{e_2}	-	-	-	-	0.022	0.007	0.015	0.043
ρ_1	621.888	50.677	528.743	728.144	764.748	59.322	655.597	888.964
σ_1	5.121	0.612	4.034	6.438	7.690	0.872	6.133	9.555
ϕ_1	0.992	0.002	0.988	0.995	0.998	0.001	0.997	0.999
ρ_2	-	-	-	-	112.768	8.076	97.773	129.554
σ_2	-	-	-	-	0.668	0.066	0.547	0.807
ϕ_2	-	-	-	-	0.937	0.014	0.906	0.960

Table A.2: Posterior estimates of the regression calibration model for temperature

Parameter	Mean	SD	P2.5%	P97.5%
σ_{e_1}	49.893	7.734	36.866	67.233
Range of $\alpha_0(\mathbf{s}, t)$	48.011	13.352	26.779	78.911
SD of $\alpha_0(\mathbf{s}, t)$	0.444	0.065	0.333	0.589
AR parameter of $\alpha_0(\mathbf{s}, t)$	0.777	0.058	0.650	0.875
Range of $\alpha_1(\mathbf{s}, t)$	1103.250	104.709	913.898	1325.815
SD of $\alpha_1(\mathbf{s}, t)$	0.673	0.087	0.517	0.860
AR parameter of of $\alpha_1(\mathbf{s}, t)$	0.999	0.000	0.999	1.000

A.2 Temperature model

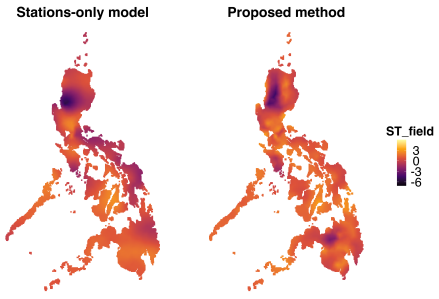


Figure A.3: Comparison of estimated spatial fields $\hat{\xi}(\mathbf{s}, t)$ for August 2019 among the three approaches: stations-only model, regression calibration model, and the proposed data fusion model. The estimated spatial fields are roughly similar.

A.3 Relative humidity model

Table A.3: Marginal log-likelihood values conditional an α_1 and the corresponding BMA weights for the relative humidity data fusion model

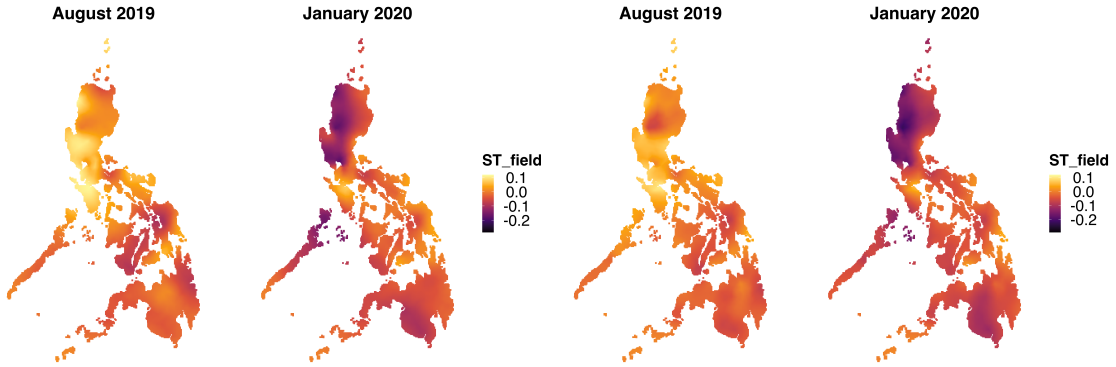
α_1	$\log \pi(\mathbf{Y}\|\alpha_1)$	w_k
0.5	5955.341	0.0000
0.6	5830.252	0.0000
0.7	6076.597	0.0000
0.8	6136.254	0.0000
0.9	5320.351	0.0000
1	6248.282	1
1.1	6116.906	0.0000
1.2	6101.200	0.0000
1.3	6071.814	0.0000
1.4	6044.122	0.0000
1.5	5847.258	0.0000

Table A.4: Posterior estimates of hyperparameters for the relative humidity model – stations-only model versus proposed data fusion model

	Stations only				**Proposed model**			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ_{e_1}	0.012	0.001	0.011	0.014	0.020	0.001	0.018	0.022
σ_{e_2}	-	-	-	-	0.003	0.001	0.002	0.005
ρ_1	287.577	26.952	238.127	344.148	589.113	67.976	468.374	735.571
σ_1	0.087	0.008	0.073	0.103	0.111	0.014	0.087	0.142
ϕ_1	0.929	0.012	0.902	0.949	0.970	0.008	0.952	0.983
ρ_2	-	-	-	-	117.256	8.956	100.403	135.635
σ_2	-	-	-	-	0.040	0.002	0.037	0.045
ϕ_2	-	-	-	-	0.855	0.015	0.824	0.883

Table A.5: Posterior estimates of the regression calibration model for relative humidity

Parameter	Mean	Sd	P2.5%	P97.5%
σ_{e_1}	5769.559	673.476	4563.267	7211.298
Range of $\alpha_0(\mathbf{s}, \mathbf{t})$	1.663	1.542	0.299	5.758
SD of $\alpha_0(\mathbf{s}, \mathbf{t})$	1.935	1.579	0.319	6.099
AR parameter of $\alpha_0(\mathbf{s}, \mathbf{t})$	0.914	0.017	0.877	0.943
Range of $\alpha_1(\mathbf{s}, \mathbf{t})$	2765.158	443.936	2001.709	3745.046
Range of $\alpha_1(\mathbf{s}, \mathbf{t})$	0.105	0.015	0.079	0.137
AR parameter of $\alpha_1(\mathbf{s}, \mathbf{t})$	0.993	0.004	0.983	0.998



(a) Stations-only model **(b)** Proposed data fusion model

Figure A.4: Estimated spatial fields $\hat{\xi}(\mathbf{s}, t)$ for log relative humidity, August 2019 and January 2020, for two approaches: (a) stations-only model, (b) proposed data fusion model.

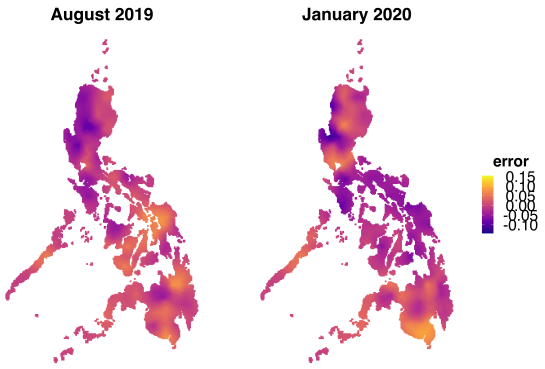
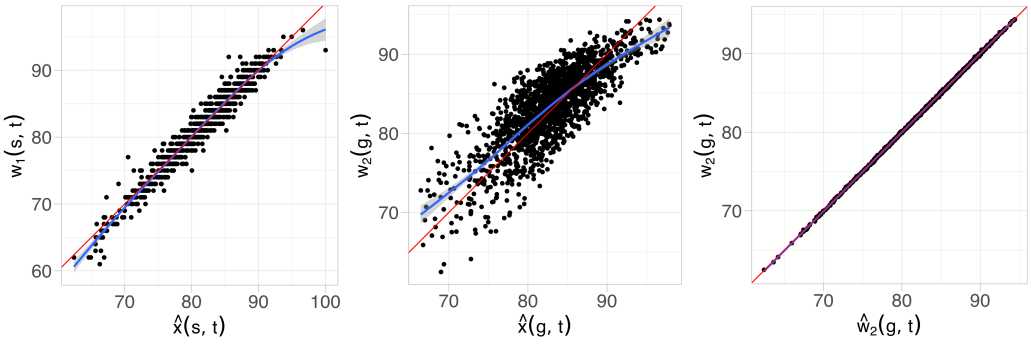


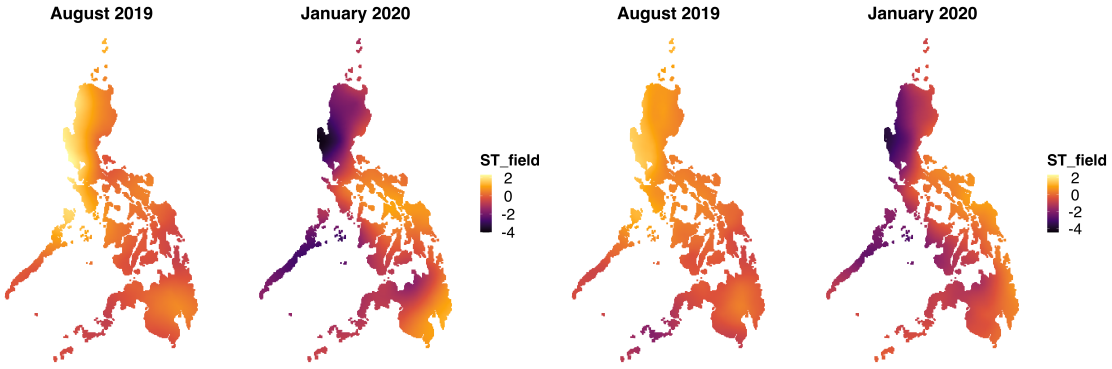
Figure A.5: Estimated error fields for the GSM log relative humidity data, August 2019 and January 2020, using the proposed data fusion model.



(a) $\mathrm{w}_1(\mathbf{s}, t)$ vs $\hat{x}(\mathbf{s}, t)$ **(b)** $\mathrm{w}_2(\mathbf{g}, t)$ vs $\hat{x}(\mathbf{g}, t)$ **(c)** $\mathrm{w}_2(\mathbf{g}, t)$ vs $\hat{\mathrm{w}}_2(\mathbf{g}, t)$

Figure A.6: Plot of observed relative humidity values versus predicted values using the proposed data fusion model for (a) weather stations and (b) GSM data, and (c) calibrated GSM data. The blue line is the smooth local regression curve, while the red line is the identity line.

A.4 Rainfall model



(a) Stations-only model **(b)** Proposed data fusion model

Figure A.7: Estimated spatial fields $\hat{\xi}(\mathbf{s}, t)$ for log rainfall, August 2019 and January 2020, for two approaches: (a) stations-only model and (b) proposed data fusion model.

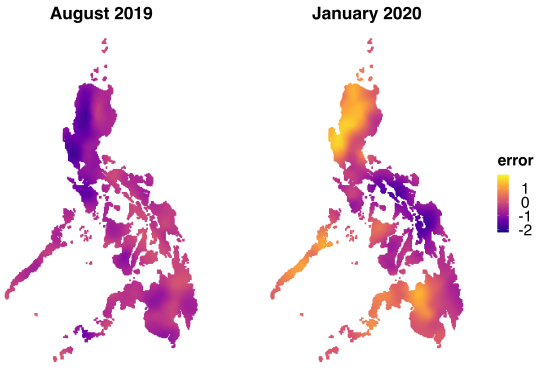


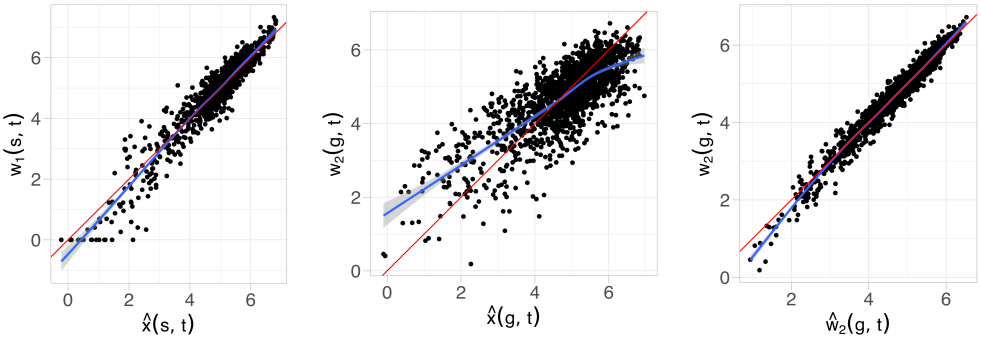
Figure A.8: Estimated error fields for the GSM log rainfall data for August 2019 and January 2020 using the proposed data fusion model.

Table A.6: Posterior estimates of hyperparameters for the log rainfall model – stations-only model versus proposed data fusion model

	Stations only				Proposed model			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ_{e_1}	0.482	0.019	0.446	0.521	0.501	0.022	0.466	0.549
σ_{e_2}	-	-	-	-	0.256	0.014	0.229	0.284
ρ_1	614.617	62.498	501.641	747.464	584.259	59.817	480.042	711.353
σ_1	1.113	0.074	0.976	1.266	1.107	0.065	0.978	1.227
ϕ_1	0.601	0.048	0.501	0.692	0.691	0.037	0.611	0.754
ρ_2					434.935	66.349	323.066	588.186
σ_2	-	-	-	-	0.942	0.095	0.764	1.116
ϕ_2	-	-	-	-	0.870	0.023	0.820	0.908

Table A.7: Posterior estimates of the regression calibration model for the log rainfall

Parameter	Mean	SD	P2.5%	P97.5%
σ_{e_1}	4.467	0.502	3.543	5.515
Range for $\alpha_0(\mathbf{s},t)$	9.012	12.217	1.398	39.452
SD for $\alpha_0(\mathbf{s},t)$	2.291	1.704	0.292	6.588
AR parameter for $\alpha_0(\mathbf{s},t)$	0.257	0.240	-0.276	0.647
Range for $\alpha_1(\mathbf{s},t)$	1083.233	107.018	887.844	1308.788
SD for $\alpha_1(\mathbf{s},t)$	0.410	0.030	0.355	0.473
AR parameter for $\alpha_1(\mathbf{s},t)$	0.820	0.031	0.753	0.874



(a) $w_1(\mathbf{s},t)$ vs $\hat{x}(\mathbf{s},t)$ **(b)** $w_2(\mathbf{g},t)$ vs $\hat{x}(\mathbf{g},t)$ **(c)** $w_2(\mathbf{g},t)$ vs $\hat{w}_2(\mathbf{g},t)$

Figure A.9: Plot of observed log rainfall values versus predicted values using the proposed data fusion model: (a) weather stations, (b) GSM data, (c) calibrated GSM data. The blue line is the smooth local regression curve, while the red line is the identity line.

A.5 LGOCV

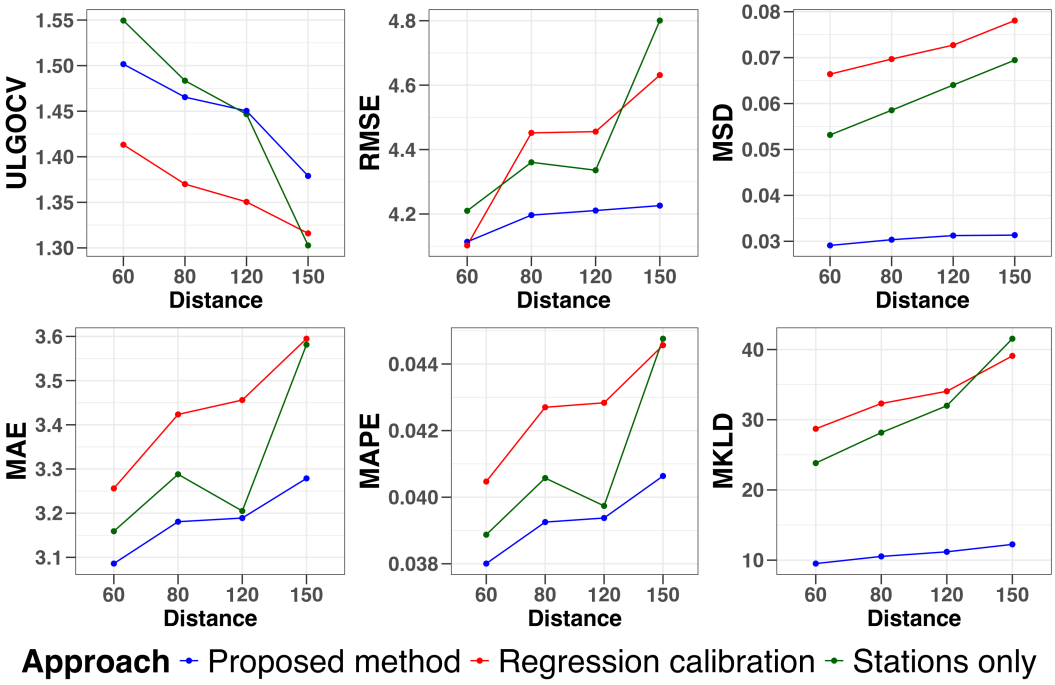


Figure A.10: Comparison of LGOCV results for relative humidity from three models: stations-only model, regression calibration model, and the proposed data fusion model

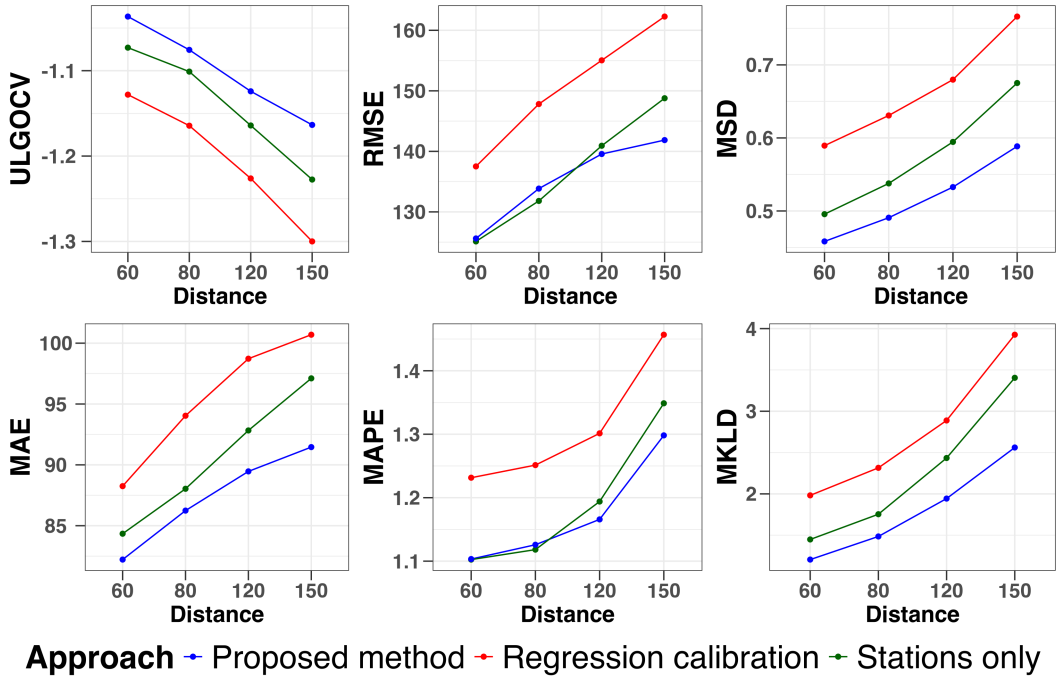


Figure A.11: Comparison of LGOCV results for log rainfall from three models: stations-only model, regression calibration model, and the proposed data fusion model

Appendix B

Appendix for Chapter 5

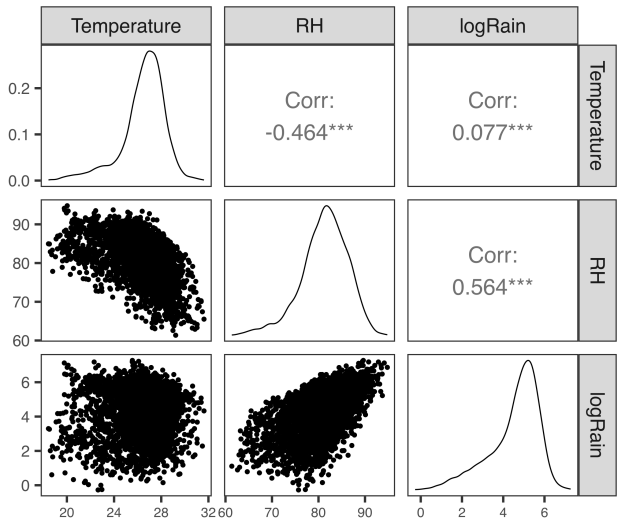


Figure B.1: Pairwise correlation among the block-level estimates of the climate variables: temperature, relative humidity, and log rainfall

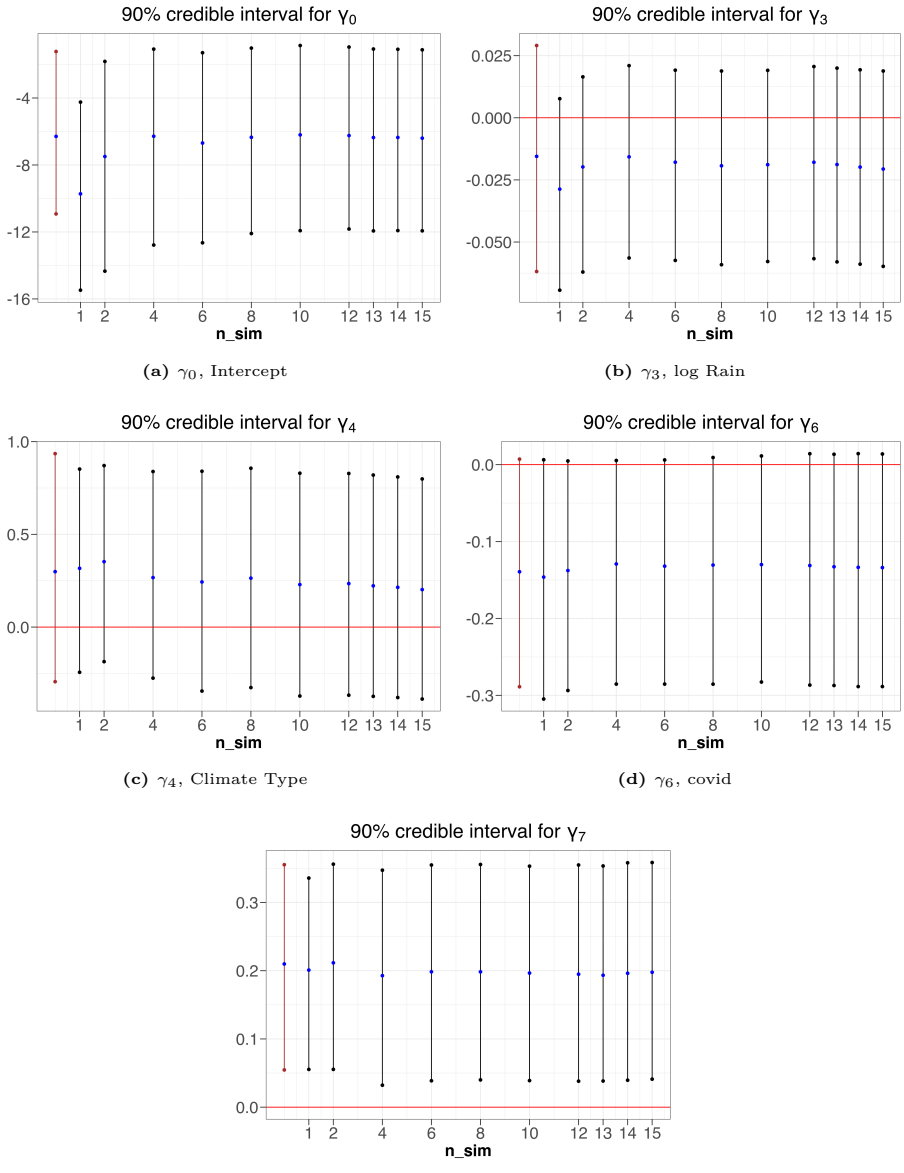
B.1 Input model: data fusion model

	Plug-in method				Resampling method			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ^2_ν, RW2 time	0.0084	0.0091	0.0001	0.0326	0.0726	0.1020	0.0060	0.3063
σ^2_ζ, iid time	0.0005	0.0015	0.0000	0.0032	0.0003	0.0009	0.0000	0.0021
σ^2_ψ, space	0.0357	0.0471	0.0003	0.1657	0.0400	0.1142	0.0000	0.3363
ϕ	0.2420	0.2420	0.0031	0.8415	0.6801	0.4677	0.0072	1.3654
σ^2_υ, interaction	1.1442	0.1304	0.9534	1.4548	0.7028	0.4623	0.0075	1.3630
ρ	0.9070	0.0103	0.8900	0.9289	0.9018	0.0131	0.8751	0.9257

Table B.1: Comparison of hyperparameter estimates between the plug-in method and the resampling method for the dengue model with temperature and log rainfall as the climate covariates

	Plug-in method				Resampling method			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ^2_ν, RW2 time	0.0127	0.0212	0.0007	0.0627	0.0185	0.1266	0.0003	0.0737
σ^2_ζ, iid time	0.0002	0.0003	0.0000	0.0009	0.0003	0.0009	0.0000	0.0016
σ^2_ψ, space	0.0000	0.0000	0.0000	0.0000	0.0608	0.2589	0.0000	0.5053
ϕ	0.1850	0.2130	0.0017	0.7741	0.7091	0.4780	0.0110	1.3783
σ^2_υ, interaction	1.1861	0.1435	0.9487	1.5098	0.7326	0.4779	0.0117	1.3872
ρ	0.9061	0.0116	0.8831	0.9284	0.9040	0.0120	0.8783	0.9253

Table B.2: Comparison of hyperparameter estimates between the plug-in method and the resampling method for the dengue model with relative humidity as the climate covariate



(a) γ_0, Intercept

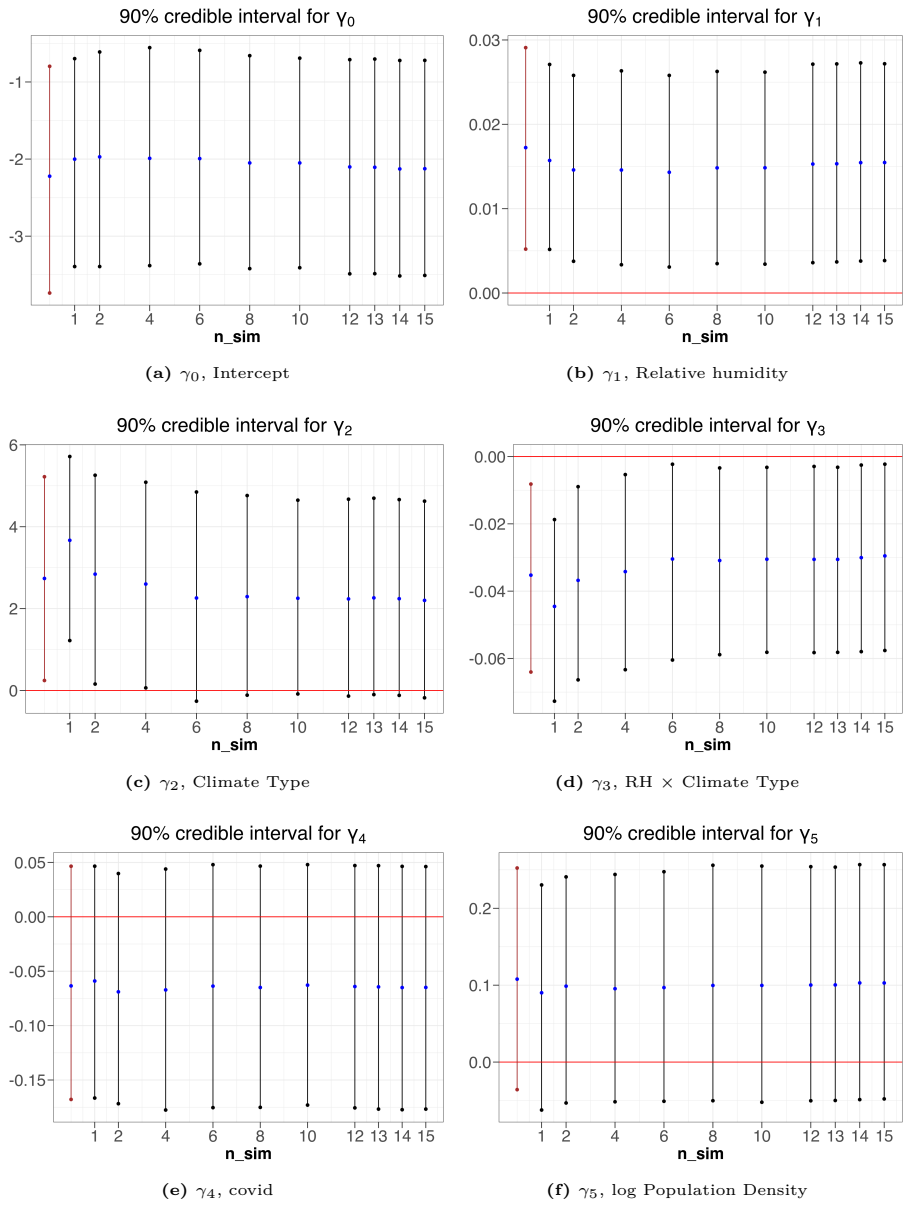
(b) γ_3, log Rain

(c) γ_4, Climate Type

(d) γ_6, covid

(e) γ_7, log Population Density

Figure B.2: Plots showing the posterior means and 90% credible intervals of the fixed effects (except γ_1, γ_2, and γ_5) for the dengue model with temperature and log rainfall as covariates. The first vertical line shows the estimates for the plug-in method, while the rest of the lines show the estimates for the resampling method for different number of resamples, from 1 to 15.



(a) γ_0, Intercept

(b) γ_1, Relative humidity

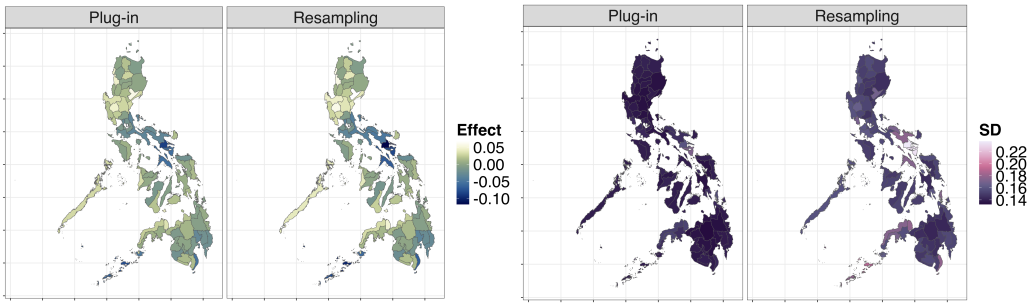
(c) γ_2, Climate Type

(d) γ_3, RH × Climate Type

(e) γ_4, covid

(f) γ_5, log Population Density

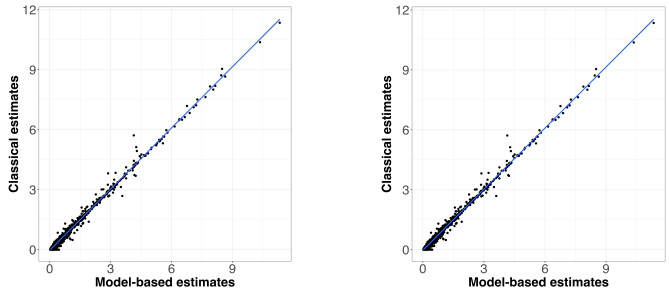
Figure B.3: Plots showing the posterior means and 90% credible intervals of the fixed effects for the dengue model with relative humidity as climate covariate. The first vertical line shows the estimates for the plug-in method, while the rest of the lines show the estimates for the resampling method for different number of resamples, from 1 to 15.



(a) Posterior mean

(b) Posterior standard deviation

Figure B.4: Comparison of (a) posterior mean and (b) posterior standard deviation, of the space effects $\psi(B_i)$ between the plug-in method and resampling method, for the model with relative humidity as climate covariate



(a) Plug-in method **(b)** Resampling method

Figure B.5: Comparison of classical SIR estimates and model-based SIR estimates from the health model with relative humidity as climate covariate

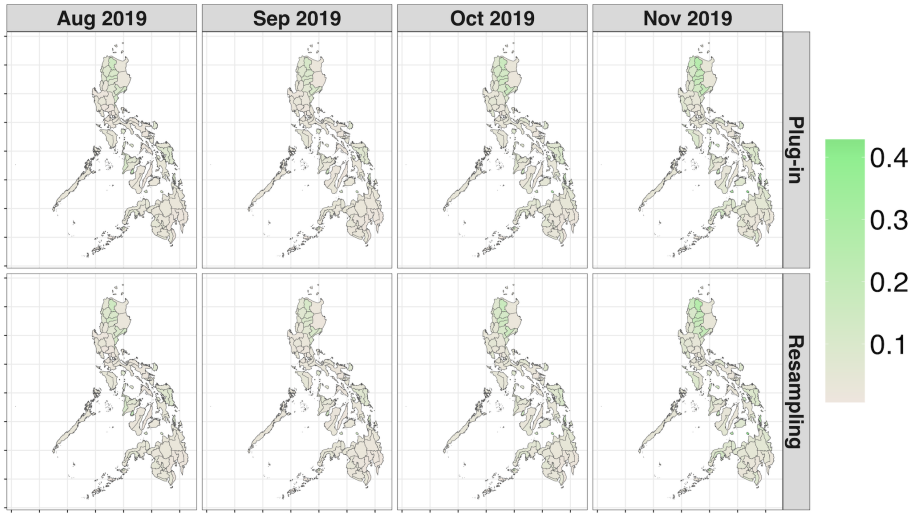


Figure B.6: Posterior uncertainty of model-based estimates of dengue risks from August 2019 to November 2019, for both plug-in method and resampling on the dengue model with temperature and log rainfall as climate covariates

	$\gamma_1 x_1$	$\gamma_2 x_2$	$\gamma_3 x_3$	$\gamma_4 x_4$	$\gamma_5 x_5$	$\gamma_6 x_6$
$\gamma_1 x_1$	1.20517	−0.00111	−0.00725	−0.05684	−0.21671	0.00397
$\gamma_2 x_2$	·	0.00049	−0.00001	−0.00005	0.00076	0.00018
$\gamma_3 x_3$	·	·	0.00018	0.00101	−0.00352	−0.00001
$\gamma_4 x_4$	·	·	·	0.10300	0.02234	0.00311
$\gamma_5 x_5$	·	·	·	·	0.80793	0.00035
$\gamma_6 x_6$	·	·	·	·	·	0.00839

Matrix B.1.1: Variance-covariance matrix across several resamples for the fixed effects components of the linear predictor of the health model with temperature and log rainfall as climate covariates

	$\nu(t)$	$\zeta(t)$	$\psi(B)$	$\upsilon(B,t)$
$\nu(t)$	0.05218	0.00006	0.00000	0.00619
$\zeta(t)$	·	0.00001	−0.00000	−0.00000
$\psi(B)$	·	·	0.00500	0.00107
$\upsilon(B,t)$	·	·	·	0.93642

Matrix B.1.2: Variance-covariance matrix across several resamples for the random effects components of the linear predictor of the health model with temperature and log rainfall as climate covariates

	$\gamma_1 x_1$	$\gamma_2 x_2$	$\gamma_3 x_3$	$\gamma_4 x_4$	$\gamma_5 x_5$	$\gamma_6 x_6$
$\nu(t)$	0.00471	−0.00000	−0.00000	−0.00019	0.00055	0.00064
$\zeta(t)$	−0.00003	0.00000	0.00000	0.00000	−0.00001	−0.00001
$\psi(B)$	−0.01027	0.00012	0.00009	0.00732	0.01425	−0.00000
$\upsilon(B,t)$	0.02381	0.00023	−0.00020	−0.00854	−0.01481	0.00149

Matrix B.1.3: Cross-covariance matrix across several resamples between the fixed effects and random effects components of the linear predictor of the health model with temperature and log rainfall as climate covariates

B.2 Input model: stations-only model

This section presents the results of the two-stage models based on a stations-only (first-stage) model input. The data used here covers a longer time series, specifically from 2016 to 2020.

B.2.1 First-stage model results

The model specification, including the predictor expressions, for the stations-only models is similar to the data fusion models in Chapter 4. I also fitted separate models for the three climate variables.

Tables B.3, B.4, and B.5 show the fixed effects estimates for the temperature model, relative humidity model, and log rainfall model, respectively. These estimates are very similar to the obtained posterior estimates in Chapter 4. The posterior estimates of the model hyperparameters for the temperature model, relative humidity model, and log rainfall model are in Tables B.6, B.7, and B.8, respectively.

Parameter	Mean	SD	P2.5%	P97.5%
β_0	28.2177	3.2672	21.8140	34.6213
β_1, log(Elevation)	-0.6077	0.0928	-0.7896	-0.4259
β_2, Cool	-0.7127	0.1283	-0.9642	-0.4612
β_3, Climate Type	2.1549	0.6904	0.8018	3.5080

Table B.3: Posterior estimates of fixed effects for temperature model

Parameter	Mean	SD	P2.5%	P97.5%
β_0	4.4308	0.0274	4.3771	4.4845
β_1, log(Temperature)	0.6926	0.0278	0.6382	0.7471
β_2, log(Temperature)^2	-0.2106	0.0082	-0.2267	-0.1944
β_3, log(Elevation)	-0.0128	0.0024	-0.0175	-0.0081
β_4, Climate Type	0.0411	0.0159	0.0098	0.0723

Table B.4: Posterior estimates of fixed effects for log relative humidity model

Parameter	Mean	SD	P2.5%	P97.5%
β_0	3.7415	0.2305	3.2898	4.1931
β_1, log(Temperature)	2.9234	0.2925	2.3502	3.4967
β_2, log(Temperature)^2	-0.8718	0.0854	-1.0392	-0.7044
β_3, Season	0.7590	0.1660	0.4337	1.0844
β_4, Climate Type	1.1782	0.0866	1.0085	1.3479
β_5, Season $\times$ Climate Type	-0.7313	0.0955	-0.9185	-0.5441

Table B.5: Posterior estimates of fixed effects for log rainfall model

Parameter	Mean	SD	P2.5%	P97.5%
σ_{e_1}, measurement error SD	0.2238	0.0060	0.2122	0.2358
ρ_{ω_1}, range of ω_1	707.7730	32.0294	647.0915	773.1777
σ_{ω_1}, marginal SD of ω_1	6.0641	0.5621	5.0579	7.2692
ϕ_1, AR parameter of ω_1	0.9937	0.0011	0.9913	0.9957

Table B.6: Posterior estimates of hyperparameters for temperature model

Figure B.7 shows the predicted climate fields, $\hat{x}(\mathbf{s}, t)$ for two months: January 2019 and August 2019. As explained in Section 4.6, the reason for choosing these two specific months is that January is dry and cold, while August is typically hot and wet (PAGASA, 2023). This can be confirmed from Figure B.7a, which shows that it

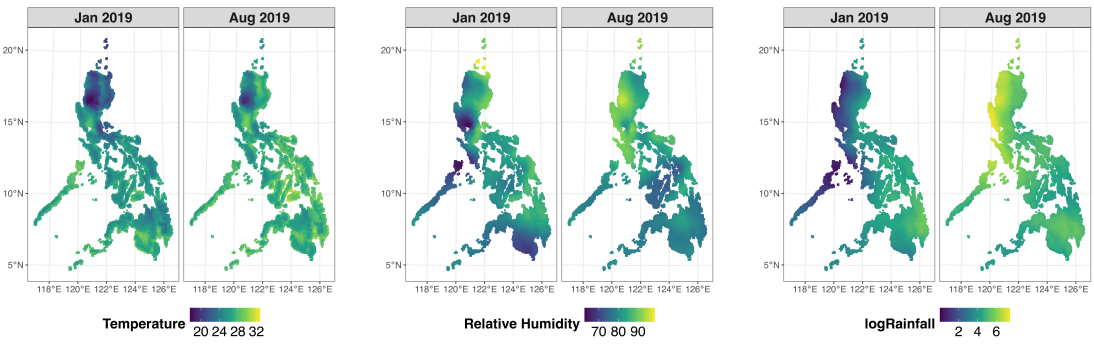
Parameter	Mean	SD	P2.5%	P97.5%
σ_{e_1}, measurement error SD	0.0152	0.0004	0.0144	0.0160
ρ_{ω_1}, range of ω_1	339.2946	19.0730	303.4962	378.5705
σ_{ω_1}, marginal SD of ω_1	0.1070	0.0073	0.0936	0.1224
ϕ_1, AR parameter of ω_1	0.9533	0.0060	0.9407	0.9642

Table B.7: Posterior estimates of hyperparameters for log relative humidity model

Parameter	Mean	SD	P2.5%	P97.5%
σ_{e_1}, measurement error SD	0.5294	0.0123	0.5057	0.5541
ρ_{ω_1}, range of ω_1	569.3168	32.1078	509.1225	635.5036
σ_{ω_1}, marginal SD of ω_1	1.2168	0.0452	1.1309	1.3088
ϕ_1, AR parameter of ω_1	0.6636	0.0249	0.6131	0.7110

Table B.8: Posterior estimates of hyperparameters for log rainfall model

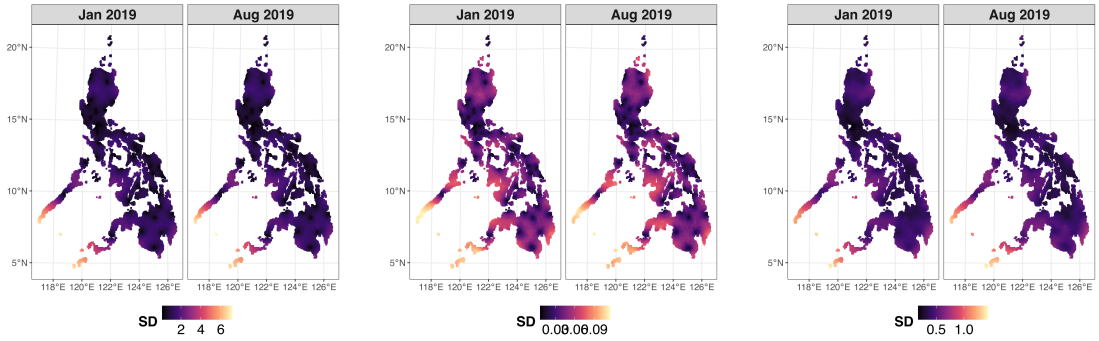
is generally colder during January than August. Moreover, Figure B.7c shows that during January, the western section of the country has low amount of rainfall, but is the opposite during August. On the other hand, the eastern section has relatively high amount of rainfall for both months. This agrees with the results from Chapter 4. Figure B.7b shows the predicted relative humidity fields. The spatial structure in the relative humidity fields is similar to that of the log rainfall fields. The corresponding posterior standard deviations of the predicted climate fields are in Figure B.8.



(a) Temperature fields **(b)** Relative humidity fields **(c)** log Rainfall fields

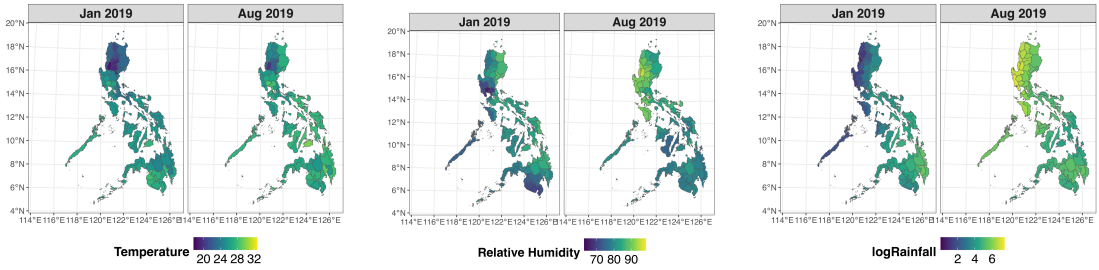
Figure B.7: Predicted climate fields (posterior means), $\hat{x}(\mathbf{s}, t)$, for January and August 2019: (a) temperature (b) relative humidity (c) log rainfall

Finally, Figure B.9 shows the block-level estimates, $\hat{x}(B_i, t)$, of the predicted fields in Figure B.7. These are the values used as covariates in the second-stage model using the plug-in method.



(a) Temperature **(b)** Relative humidity **(c)** log Rainfall

Figure B.8: Posterior standard deviation of the predicted climate fields, $\hat{x}(\mathbf{s}, t)$, for January and August 2019: (a) temperature (b) relative humidity (c) log rainfall



(a) Temperature **(b)** Relative humidity **(c)** log Rainfall

Figure B.9: Predicted block-level climate values, $\hat{x}(B, t)$, for January and August 2019: (a) temperature (b) relative humidity (c) log rainfall

B.2.2 Second-stage model results

B.2.2.1 Temperature and log rainfall

The predictor expression of the dengue model with temperature and log rainfall as climate covariates is the same as Equation (5.13) in Chapter 5. Table B.9 shows the marginal log likelihood (MLik), Watanabe-Akaike Information Criterion (WAIC), and the conditional predictive ordinate (CPO) values for the different models considered. These values are based on the results from the crude plug-in method. Similar to the results in Chapter 5, the Type II interaction model has the highest MLik, the smallest WAIC, and the smallest CPO value as well, for both input models. Hence, the Type II interaction model was considered for further investigation.

Table B.10 shows the fixed effects estimates. Results also suggest a non-linear relationship between temperature and dengue. However, the credible intervals for the resampling approach, which are wider than the plug-in approach, contain the null value of zero for the main effect of temperature. Moreover, the results show

Model	MLik	WAIC	CPO
Type I	-27075.13	38978.99	46047.86
Type II	-24749.85	38467.45	26564.30
Type III	-26973.33	39121.08	42643.34
Type IV	-42936.63	97270.81	52331.64

Table B.9: Marginal log likelihood (MLik), WAIC, and $-\sum \log \text{CPO}_i$ for different dengue models with temperature and log rainfall as climate covariates, using the stations-only climate model as input

that log rainfall is significant and is positively related with dengue. The interaction between log rainfall and climate type is also significant and negative, at least for the plug-in approach. The resampling approach, which has higher posterior uncertainty, gave wider credible intervals containing zero. Moreover, population density is also positively related with dengue. Finally, the `covid` binary variable is not significant in the model, although the intervals cover mostly negative values. The model insights are similar to the results from the dengue models with the data fusion models as input. The results also show that the posterior standard deviations from using the resampling method are generally larger compared to the plug-in method.

	Plug-in method				**Resampling method**			
Parameter	Mean	SD	P5%	P95%	Mean	SD	P5%	P95%
γ_0, Intercept	-3.6612	1.6296	-6.3025	-1.0625	-2.9708	1.7753	-5.9465	-0.0945
γ_1, Temperature	0.2296	0.1174	0.0482	0.4319	0.1775	0.1343	-0.0376	0.4023
γ_2, Temperature^2	-0.0054	0.0022	-0.0091	-0.0016	-0.0041	0.0026	-0.0085	-0.0000
γ_3, log Rain	0.0219	0.0098	0.0042	0.0386	0.0180	0.0100	0.0015	0.0345
γ_4, ClimateType	0.0491	0.1980	-0.3012	0.3479	0.0155	0.2094	-0.3345	0.3530
γ_5, log Rain $\times$ ClimateType	-0.0350	0.0190	-0.0715	-0.0025	-0.0253	0.0201	-0.0579	0.0079
γ_6, covid	-0.0614	0.0729	-0.1789	0.0551	-0.0667	0.0719	-0.1864	0.0493
γ_7, log PopDensity	0.1548	0.0859	0.0210	0.3032	0.1459	0.0888	-0.0001	0.2907

Table B.10: Comparison of estimates of fixed effects between the plug-in method and the resampling method for the dengue model with temperature and log rainfall as climate covariates, using the stations-only climate model as input

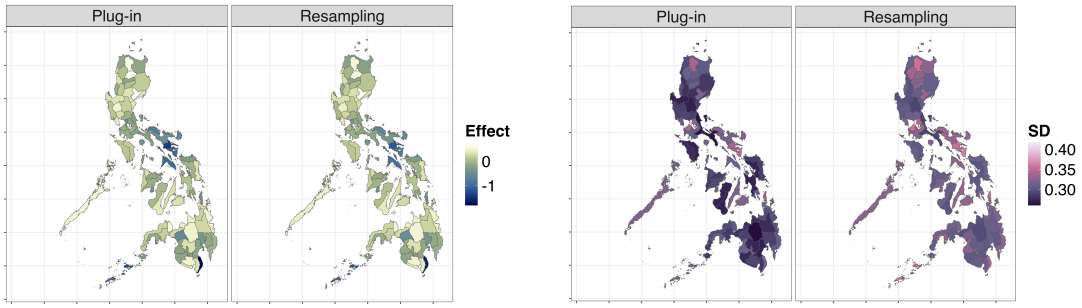
Table B.11 shows the estimates of the hyperparameters. The posterior uncertainty are generally higher for the resampling method compared to the plug-in method. Moreover, the variance explained by the structured effects, both in space and time, are higher compared to the unstructured effects.

Figures B.10a and B.10b show the estimated posterior means and posterior standard deviations, respectively, of the spatial effect $\psi(B_i)$. The posterior means are similar, but the posterior standard deviation from the resampling method is generally higher. Figure B.11a shows the estimated (random walk) time effect. There is not much difference in the posterior uncertainty between the plug-in method and

	Plug-in method				Resampling method			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ^2_ν, RW2 time	1.3002	0.6353	0.5375	2.9652	1.3114	1.8176	0.1397	4.1670
σ^2_ζ, iid time	0.0001	0.0001	0.0000	0.0005	0.0001	0.0001	0.0000	0.0003
σ^2_ψ, space	0.3812	0.0839	0.2400	0.5681	0.3839	0.1102	0.2225	0.6361
ϕ	0.8253	0.1582	0.4086	0.9917	0.6652	0.2162	0.1249	0.9998
σ^2_υ, interaction	0.6797	0.0469	0.6079	0.7885	0.6658	0.2119	0.1277	0.9998
ρ	0.8632	0.0095	0.8471	0.8836	0.8617	0.0098	0.8418	0.8803

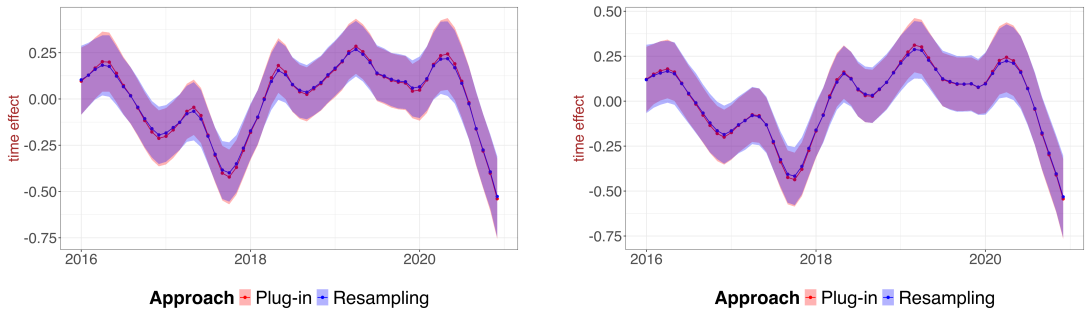
Table B.11: Comparison of hyperparameter estimates between the plug-in method and the resampling method for the dengue model with temperature and log rainfall as the climate covariate, using the stations-only climate model as input

resampling method. Moreover, there is a sharp decline in the time effects during the COVID-19 episode. As argued in Chapter 5, this is potentially the reason why the `covid` binary variable is not significant in the model, since the information on the decline of dengue risks is already accounted for by the temporal random effect.



(a) Posterior mean **(b)** Posterior standard deviation

Figure B.10: Comparison of (a) posterior mean and (b) posterior standard deviation, of the space effects $\psi(B_i)$ between the plug-in method and resampling method, for the dengue model with temperature and log rainfall as climate covariate, using the stations-only climate model as input



(a) Temperature and log rainfall as covariates **(b)** Relative humidity as covariate

Figure B.11: Plot of the estimated structured time effects $\nu(t)$ with the 95% credible intervals between the plug-in method and resampling method (using the stations-only climate model as input): (a) temperature and log rainfall as climate covariates (b) relative humidity as covariate

B.2.2.2 Relative humidity

The linear predictor of the dengue model with relative humidity as climate covariate is the same as Equation (5.14) in Chapter 5. Table B.12 shows the marginal log likelihood, WAIC, and CPO values for the different dengue models, according to the type of space-time interaction. Similar to the model with temperature and log rainfall as climate covariates, the model with Type II interaction also has the highest marginal likelihood, smallest WAIC, and smallest CPO.

Model	MLik	WAIC	CPO
Type I	-27092.76	38981.69	46160.60
Type II	-24747.84	38458.72	26605.39
Type III	-26981.77	39103.29	42802.41
Type IV	-44263.58	102804.50	Inf

Table B.12: Marginal log likelihood, WAIC, and $-\sum \log \text{CPO}_i$ for different dengue models with relative humidity as the climate covariate, using the stations-only climate model as input

Table B.13 shows the estimates of the fixed effects. The coefficient of relative humidity is significant and positive. The results show that the posterior standard deviations of γ_0 and γ_1 are higher for the resampling method compared to the plug-in method. The rest of the parameters are not significant.

Table B.14 shows the estimates of the hyperparameters. The posterior standard deviations from the resampling method are significantly higher for the resampling method compared to the plug-in method. The variation in the data explained by the structured effect in time is higher compared to the unstructured effect.

Figures B.12a and B.12b show a comparison of the posterior mean and posterior standard deviation, respectively, for the spatial effect $\psi(B_i)$ between the plug-in and resampling method. The posterior means are very similar, but the posterior uncertainty from the resampling method is larger compared to the plug-in method.

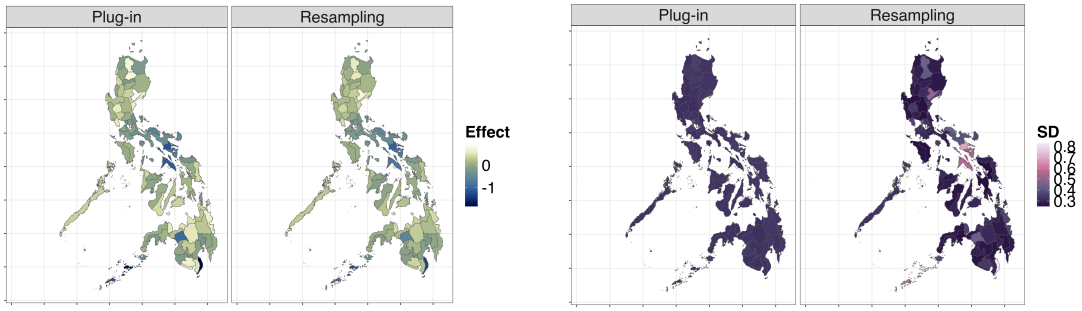
	Plug-in method				Resampling method			
Parameter	Mean	SD	P5%	P95%	Mean	SD	P5%	P95%
γ_0, Intercept	-2.0783	0.5411	-2.9765	-1.1513	-2.0074	0.6038	-3.0343	-1.0631
γ_1, RH	0.0158	0.0036	0.0106	0.0221	0.0125	0.0038	0.0062	0.0188
γ_2, ClimateType	0.2789	0.6090	-0.6775	1.3490	0.3241	0.5944	-0.6638	1.3018
γ_3, RH × ClimateType	-0.0064	0.0071	-0.0166	0.0047	-0.0063	0.0068	-0.0173	0.0049
γ_4, covid	-0.0628	0.0770	-0.1864	0.0470	-0.0669	0.0719	-0.1832	0.0512
γ_5, log PopDensity	0.0810	0.0837	-0.0421	0.2131	0.1210	0.0860	-0.0123	0.2689

Table B.13: Comparison of estimates of fixed effects between the plug-in method and the resampling method for the dengue model with relative humidity as climate covariate, using the stations-only climate model as input

	Plug-in method				Resampling method			
Parameter	Mean	SD	P2.5%	P97.5%	Mean	SD	P2.5%	P97.5%
σ^2_ν, RW2 time	0.9660	0.4065	0.3977	1.9708	1.0275	2.0496	0.2895	2.5071
σ^2_ζ, iid time	0.0002	0.0001	0.0000	0.0005	0.0190	0.1166	0.0000	0.562
σ^2_ψ, space	0.3946	0.0899	0.2612	0.6109	0.2779	0.1815	-0.0000	0.5993
ϕ	0.0305	0.0219	0.0070	0.0884	0.6602	0.3141	0.0000	1.1458
σ^2_υ, interaction	0.6863	0.0443	0.6097	0.7830	0.6626	0.3093	0.0000	1.1428
ρ	0.8638	0.0090	0.8468	0.8819	0.8775	0.0236	0.8471	0.9217

Table B.14: Comparison of hyperparameter estimates between the plug-in method and the resampling method for the dengue model with relative humidity as the climate covariate, using the stations-only climate model as input

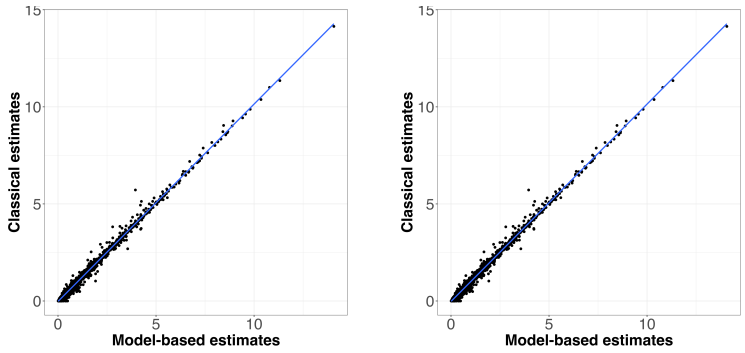
Finally, Figure B.11b shows a comparison of the estimated structured time effects $\nu(t)$ between the plug-in method and resampling method. The posterior means from the two methods are very similar. There is also not much difference in the 95% credible intervals.



(a) Posterior mean **(b)** Posterior standard deviation

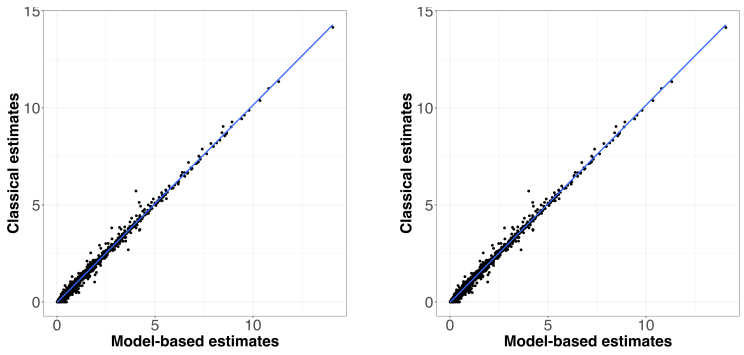
Figure B.12: Comparison of (a) posterior mean and (b) posterior standard deviation, of the space effects $\psi(B_i)$ between the plug-in method and resampling method, for the dengue model with temperature and log rainfall as climate covariate, using the stations-only climate model as input

B.2.2.3 Estimated risks



(a) Plug-in method **(b)** Resampling method

Figure B.13: Comparison of classical SIR estimates and model-based SIR estimates (using the stations-only climate model as input) from the health model with temperature and log rainfall as climate covariates



(a) Plug-in method **(b)** Resampling method

Figure B.14: Comparison of classical SIR estimates and model-based SIR estimates (using the stations-only climate model as input) from the health model with relative humidity as climate covariate

Figure B.13 shows a scatter plot between the classical estimates of SIR and the model-based estimates $\hat{\lambda}(B_i, t)$ from the model with temperature and log rainfall as climate covariates, and both for the plug-in method and the resampling method. Similar to the results in Chapter 5, there is general agreement between the classical estimates and model-based estimates. Figure B.14 shows the same scatter plots; but here, the model-based estimates of SIR are based on the dengue model with relative humidity as climate covariate.

Figure B.15 shows the estimated values $\hat{\lambda}(B_i, t)$ from the model with temperature and log-rainfall as climate covariates, for August 2019 to November 2019, and for both the plug-in method and resampling method. The maps show an agreement between the plug-in method and the resampling method. Moreover, the areas shown to have high estimated risks are the same areas with high number of cases during the dengue epidemic in the country. These maps look very similar to the plot in Figure 5.9.

Figure B.16 shows the posterior uncertainties of the estimated $\hat{\lambda}(B_i, t)$ in Figure B.15. Similar to the results in Chapter 5, there is no difference in the posterior uncertainties between the plug-in method and resampling method. Matrix (B.2.1) shows the average, across posterior samples, of the variance-covariance structure of the fixed effects in the predictor expression. Matrix (B.2.2) shows the same, but for the random effects in the predictor expression, while Matrix (B.2.3) shows the cross-covariance structure between the fixed and random effects. As argued in Section 5.6.3 of Chapter 5, there is a mix of positive and negative correlations among the

model components of the predictor expression, which potentially explains why the uncertainty is not different from the plug-in method. As emphasized in Chapter 5, although the resampling method generally gives higher uncertainty for individual components of the linear predictor, the uncertainty in a linear combination of these components can be washed away because of the latent correlation structure.

Finally, Figure B.17 shows the probability of exceedence $\mathbb{P}\big(\lambda(B_i, t) > 1\big)$ from August 2019 to November 2019, and from using the plug-in method and resampling method, and temperature and log rainfall as climate covariates. Most of the areas with $\mathbb{P}\big(\lambda(B_i, t) > 1\big) = 1$ are the same areas badly hit by dengue during the epidemic.

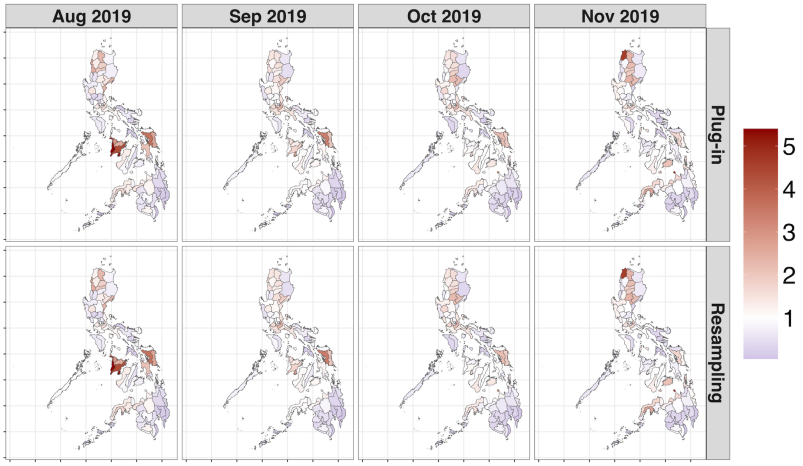


Figure B.15: Model-based estimates of dengue risks from August 2019 to November 2019, and from using the plug-in method and resampling method, and temperature and log rainfall as climate covariates, using the stations-only climate model as input

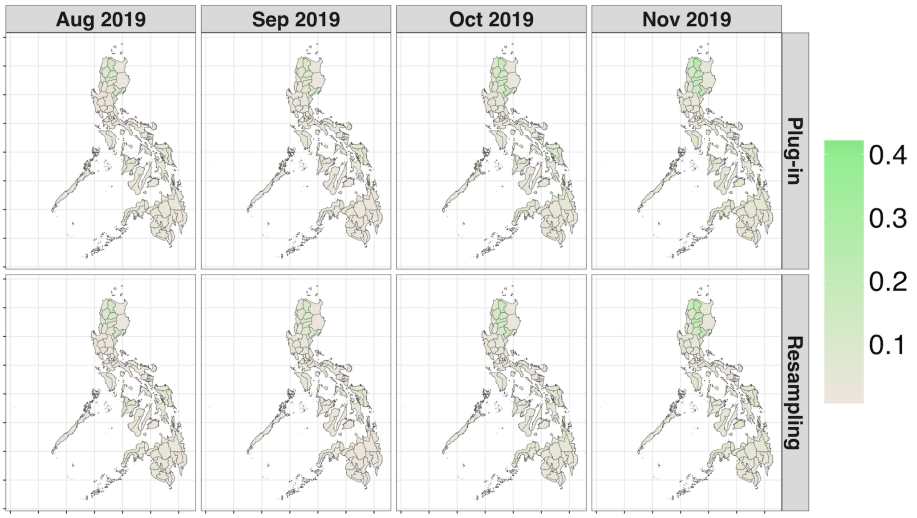


Figure B.16: Posterior standard deviation of the model-based estimates of dengue risks from August 2019 to November 2019, and from using the plug-in method and resampling method, and temperature and log rainfall as climate covariates, using the stations-only climate model as input

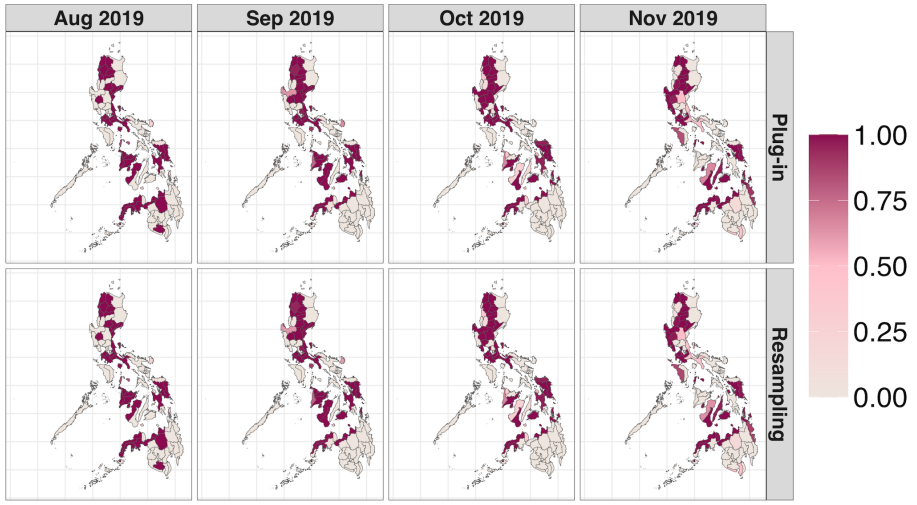


Figure B.17: Probability of exceedence, i.e., $\mathbb{P}(\lambda(B_i, t) > 1)$ from August 2019 to November 2019, and from using the plug-in method and resampling method, and temperature and log rainfall as climate covariates, using the stations-only climate model as input

	$\gamma_1 x_1$	$\gamma_2 x_2$	$\gamma_3 x_3$	$\gamma_4 x_4$	$\gamma_5 x_5$	$\gamma_6 x_6$
$\gamma_1 x_1$	0.12913	0.00053	−0.00053	−0.00081	−0.01790	0.00067
$\gamma_2 x_2$	·	0.00052	−0.00000	−0.00001	−0.00053	0.00003
$\gamma_3 x_3$	·	·	0.00002	0.00002	−0.00057	−0.00000
$\gamma_4 x_4$	·	·	·	0.00309	0.00055	0.00018
$\gamma_5 x_5$	·	·	·	·	0.14575	0.00141
$\gamma_6 x_6$	·	·	·	·	·	0.00347

Matrix B.2.1: Variance-covariance matrix across several resamples for the fixed effects components of the linear predictor of the health model with temperature and log rainfall as climate covariates and using the stations-only model as input

	$\nu(t)$	$\zeta(t)$	$\psi(B)$	$\upsilon(B,t)$
$\nu(t)$	0.03609	0.00003	0.00000	−0.00360
$\zeta(t)$	·	0.00000	0.00000	−0.00000
$\psi(B)$	·	·	0.29588	0.00030
$\upsilon(B,t)$	·	·	·	0.53467

Matrix B.2.2: Variance-covariance matrix across several resamples for the random effects components of the linear predictor of the health model with temperature and log rainfall as climate covariates and using the stations-only model as input

	$\gamma_1 x_1$	$\gamma_2 x_2$	$\gamma_3 x_3$	$\gamma_4 x_4$	$\gamma_5 x_5$	$\gamma_6 x_6$
$\nu(t)$	−0.00005	0.00002	−0.00000	−0.00000	0.00054	0.00020
$\zeta(t)$	0.00000	−0.00000	0.00000	−0.00000	0.00000	−0.00000
$\psi(B)$	−0.03934	−0.00028	0.00027	0.00296	0.06052	0.00000
$\upsilon(B,t)$	−0.00426	−0.00012	0.00004	0.00011	−0.00192	0.00004

Matrix B.2.3: Cross-covariance matrix across several resamples between the fixed effects and random effects components of the linear predictor of the health model with temperature and log rainfall as climate covariates and using the stations-only model as input

Appendix C

Appendix for Chapter 6

Theorem C.1. *Let $\pi(\boldsymbol{\theta}_1)$ be the prior model, $\pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$ be the latent model, and $\pi(\boldsymbol{\mathcal{D}}_1|\boldsymbol{x}_1, \boldsymbol{\theta}_1)$ be the observation density or probability mass function. Let $\boldsymbol{\chi}_1$ be the latent space, and $\boldsymbol{\Theta}_1$ be the $\boldsymbol{\theta}_1$-space, and that both are continuous. Suppose $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1$ is fixed. Let $\tilde{\boldsymbol{x}}_1$ be a sample from the latent field model, i.e., $\tilde{\boldsymbol{x}}_1 \sim \pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$, and $\tilde{\boldsymbol{\mathcal{D}}}_1$ a sample from the observation model, i.e., $\tilde{\boldsymbol{\mathcal{D}}}_1 \sim \pi(\boldsymbol{\mathcal{D}}_1|\tilde{\boldsymbol{x}}_1, \boldsymbol{\theta}_1, \mathbf{Z}_1)$. Suppose that the approximate posterior from applying the Bayesian algorithm is $\hat{\pi}(\boldsymbol{x}_1|\tilde{\boldsymbol{\mathcal{D}}}_1)$. Let $\{\boldsymbol{x}_{1,\ell}\}, \ell = 1, \ldots, L$ be independent samples from the posterior distribution, i.e., $\begin{pmatrix} \boldsymbol{x}_{1,1} & \boldsymbol{x}_{1,2} & \ldots \boldsymbol{x}_{1,L} \end{pmatrix} \overset{iid}{\sim} \hat{\pi}(\boldsymbol{x}_1|\tilde{\boldsymbol{\mathcal{D}}}_1)$. For any unidimensional function $f : \boldsymbol{\chi}_1 \to \mathbb{R}$, the distribution of the rank statistic is given by*

$$r = \sum_{\ell=1}^{L} \mathbb{I}\big[f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1)\big], \quad \mathbb{I}\big[f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1)\big] = \begin{cases} 1 & \text{if } f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1) \\ 0 & \text{if } f(\boldsymbol{x}_{1,\ell}) \geq f(\tilde{\boldsymbol{x}}_1) \end{cases}$$

is $\mathcal{U}(0, 1, \ldots, L)$.

Proof. Let $\pi(\boldsymbol{\theta}_1)$ be the prior model, $\pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$ be the latent model, and $\pi(\boldsymbol{\mathcal{D}}_1|\boldsymbol{x}_1, \boldsymbol{\theta}_1)$ be the observation density or proability mass function. Let $\boldsymbol{\chi}_1$ be the latent space, and $\boldsymbol{\Theta}_1$ be the $\boldsymbol{\theta}_1$-space, and that both are continuous.

Suppose $\boldsymbol{\theta}_1 \in \boldsymbol{\Theta}_1$ is fixed. Let $\tilde{\boldsymbol{x}}_1$ be a sample from the latent field model, i.e., $\tilde{\boldsymbol{x}}_1 \sim \pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$, and $\tilde{\boldsymbol{\mathcal{D}}}_1$ a sample from the observation model, i.e., $\tilde{\boldsymbol{\mathcal{D}}}_1 \sim \pi(\boldsymbol{\mathcal{D}}_1|\tilde{\boldsymbol{x}}_1, \boldsymbol{\theta}_1)$. Note that we hold $\boldsymbol{\theta}_1$ fixed when generating the data replicates, but we fit the model assuming that both $\boldsymbol{\theta}_1$ and $\boldsymbol{x}_1$ are unknown. Let $\{\boldsymbol{x}_{1,\ell}\}, \ell = 1, \ldots, L$ be independent samples from the posterior distribution $\pi(\boldsymbol{x}_1|\boldsymbol{\mathcal{D}}_1)$,

i.e., $\begin{pmatrix} \boldsymbol{x}_{1,1} & \boldsymbol{x}_{1,2} & \dots \boldsymbol{x}_{1,L} \end{pmatrix} \overset{\text{iid}}{\sim} \pi(\boldsymbol{x}_1|\mathfrak{D}_1)$. Let $f : \boldsymbol{\chi}_1 \to \mathbb{R}$. We define the rank statistic for a specific data outcome of the Bayesian model as

$$r = \sum_{\ell=1}^{L} \mathbb{I}\big[f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1)\big], \quad \mathbb{I}\big[f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1)\big] = \begin{cases} 1 & \text{if } f(\boldsymbol{x}_{1,\ell}) < f(\tilde{\boldsymbol{x}}_1) \\ 0 & \text{if } f(\boldsymbol{x}_{1,\ell}) \geq f(\tilde{\boldsymbol{x}}_1) \end{cases}$$

For conciseness, let $f_\ell \equiv f(\boldsymbol{x}_{1,\ell})$ and $f \equiv f(\tilde{\boldsymbol{x}}_1)$. Also, let $\pi(f)$ and $\pi(f|\mathfrak{D}_1)$ be the pushforward probability density function of $\pi(\boldsymbol{x}_1|\boldsymbol{\theta}_1)$ and $\pi(\boldsymbol{x}_1|\mathfrak{D}_1)$, respectively. Suppose $p_\ell = \mathbb{P}(f_\ell < f), \ell = 1, \dots, L$. Also we assume the ordering $f_1 \leq f_2 \leq \cdots \leq f_L$. We then have:

$$\begin{aligned} \pi(r) &= \int df d\mathfrak{D}_1 \pi(\mathfrak{D}_1, f|\boldsymbol{\theta}_1) \binom{L}{r} \prod_{\ell=1}^{r} p_\ell \prod_{\ell=r+1}^{L} (1 - p_\ell) \\ &= \binom{L}{r} \int df d\mathfrak{D}_1 \pi(\mathfrak{D}_1|\boldsymbol{\theta}_1) \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_1) \prod_{\ell=1}^{r} \left[\int_{-\infty}^{f} \pi(f_\ell|\mathfrak{D}_1, f, \boldsymbol{\theta}_1) df_\ell \right] \prod_{\ell=r+1}^{L} \left[1 - \int_{-\infty}^{f} \pi(f_\ell|\mathfrak{D}_1, f, \boldsymbol{\theta}_1) df_\ell \right]. \end{aligned}$$

The probability measure for generating f_ℓ depends only on $\mathfrak{D}_1$ and is independent of the conditioning model configuration. Hence we can write $\pi(f_\ell|\mathfrak{D}_1, f, \boldsymbol{\theta}_1) = \pi(f_\ell|\mathfrak{D}_1) = \pi(f_\ell|\mathfrak{D}_1, \boldsymbol{\theta}_1), \ell = 1, \dots, L$. This implies that

$$\pi(r) = \binom{L}{r} \int df d\mathfrak{D}_1 \pi(\mathfrak{D}_1|\boldsymbol{\theta}_1) \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_1) \prod_{\ell=1}^{r} \left[\int_{-\infty}^{f} \pi(f_\ell|\mathfrak{D}_1, \boldsymbol{\theta}_1) df_\ell \right] \prod_{\ell=r+1}^{L} \left[1 - \int_{-\infty}^{f} \pi(f_\ell|\mathfrak{D}_1, \boldsymbol{\theta}_1) df_\ell \right].$$

Further, since the model used to simulate data and construct posterior distributions is the same, then we have $\pi(f_\ell|\mathfrak{D}_1, \boldsymbol{\theta}_1) = \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_1), \ell = 1, \dots, L$. Consequently, we have

$$\begin{aligned} \pi(r) &= \binom{L}{r} \int df d\mathfrak{D}_1 \pi(\mathfrak{D}_1|\boldsymbol{\theta}_1) \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_1) \prod_{\ell=1}^{r} \left[\int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_1) df' \right] \prod_{\ell=r+1}^{L} \left[1 - \int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_1) df' \right] \\ &= \binom{L}{r} \int d\mathfrak{D}_1 \pi(\mathfrak{D}_1|\boldsymbol{\theta}_1) \int df \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_1) \left[\int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_1) df' \right]^r \left[1 - \int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_1) df' \right]^{L-r}. \end{aligned}$$

Let $u = \int_{-\infty}^{f} \pi(f'|\mathfrak{D}_1, \boldsymbol{\theta}_1) df'$, so that $du = \pi(f|\mathfrak{D}_1, \boldsymbol{\theta}_1) df$. This yields

$$\begin{aligned} \pi(r) &= \binom{L}{r} \int d\mathfrak{D}_1 \pi(\mathfrak{D}_1|\boldsymbol{\theta}_1) \int du (u)^r (1-u)^{L-r} \\ &= \binom{L}{r} B(r+1, L-r+1) = \frac{L!}{r!(L-r)!} \frac{r!(L-r)!}{(L+1)!} = \frac{1}{L+1} \end{aligned}$$

□

□

C.1 SBC results for the Gaussian model (Section 6.3.1)

C.1.1 Results for the first-stage model using Algorithm 6.2

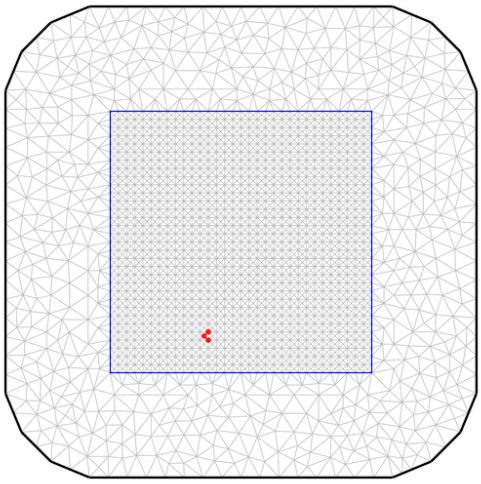


Figure C.1: Results of the KS goodness-of-fit test for uniformity (at 10% significance level) of the normalized ranks p_k of the SPDE (mesh nodes) weights out of 1000 data replicates and using Algorithm 6.2. The red points show the mesh nodes which fail the KS test for uniformity

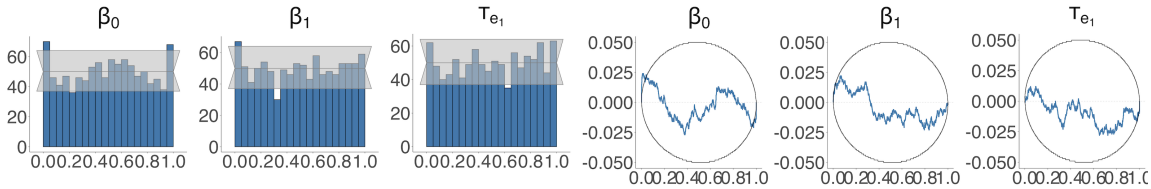


Figure C.2: Histogram and ECDF difference plot of the normalized ranks p_k for β_0, β_1, and $\tau_{e_1} = 1/\sigma^2_{e_1}$ out of 1000 data replicates using Algorithm 6.2

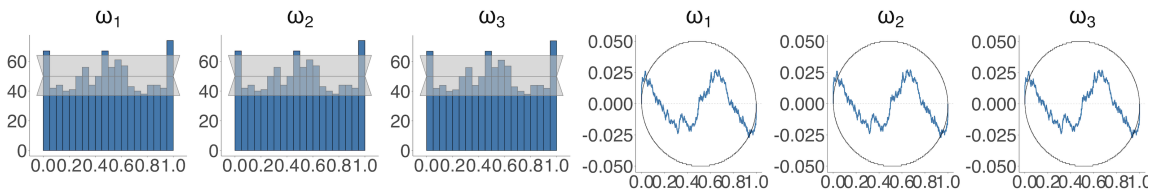


Figure C.3: Histogram and ECDF difference plot of the normalized ranks p_k for ω_1, ω_2, and ω_3 out of 1000 data replicates using Algorithm 6.2

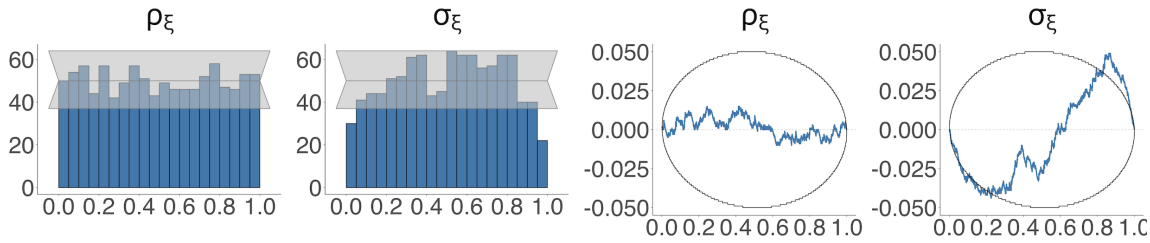


Figure C.4: Histogram and ECDF difference plot of the normalized ranks p_k for ρ_ξ and σ_ξ out of 1000 data replicates using Algorithm 6.2

C.1.2 Results for the first-stage model using Algorithm 6.3

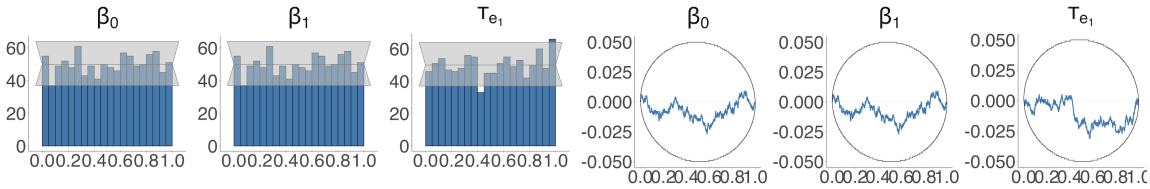


Figure C.5: Histogram and ECDF difference plot of the normalized ranks p_k for β_0, β_1, and $\tau_{e_1} = 1/\sigma^2_{e_1}$ out of 1000 data replicates using Algorithm 6.3 and using PC prior for the Matérn parameters

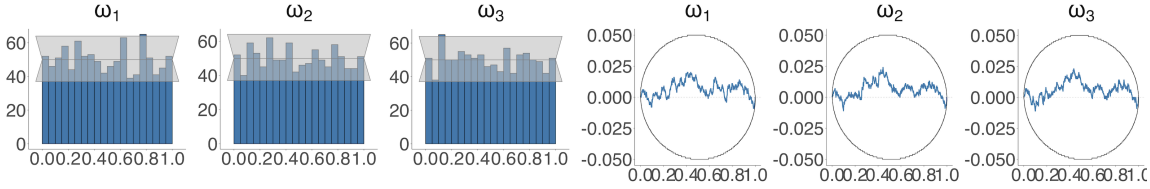
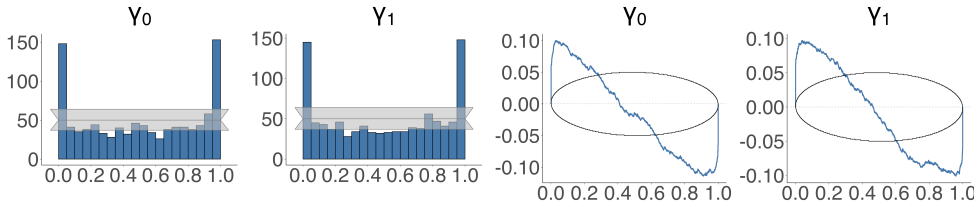
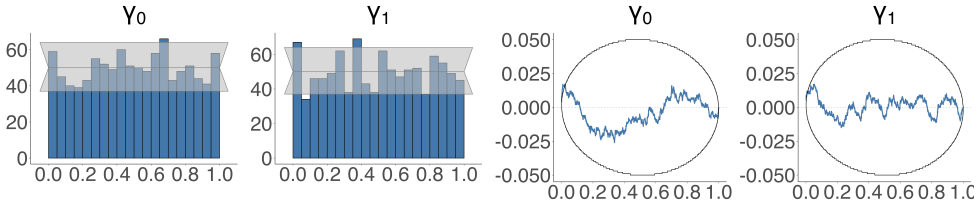


Figure C.6: Histogram and ECDF difference plot of the normalized ranks p_k for ω_1, ω_2, and ω_3 out of 1000 data replicates using Algorithm 6.3 and using PC prior for the Matérn parameters

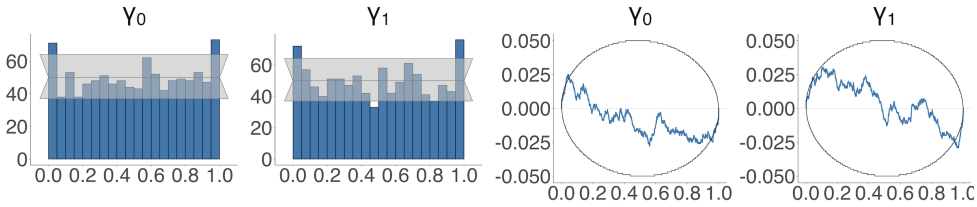
C.1.3 Results for γ_0 and γ_1 using Algorithm 6.2



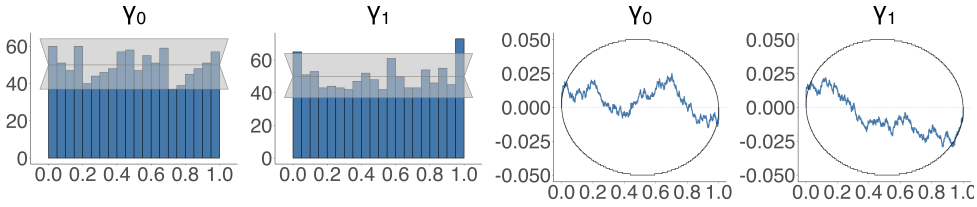
(a) Plug-in method



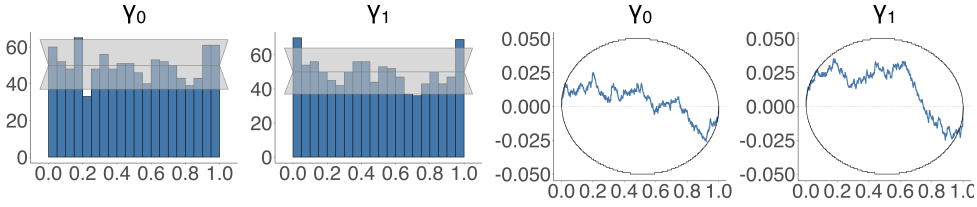
(b) Resampling method



(c) Full **Q** uncertainty method



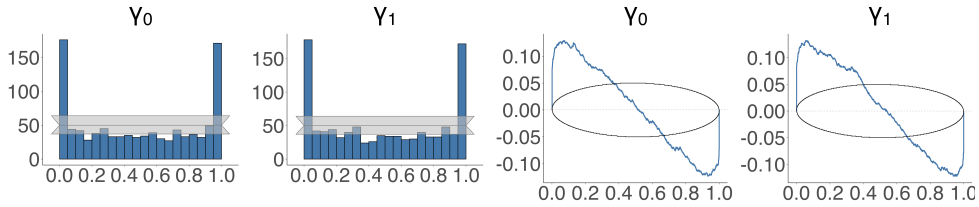
(d) Low rank **Q** uncertainty method (mesh A)



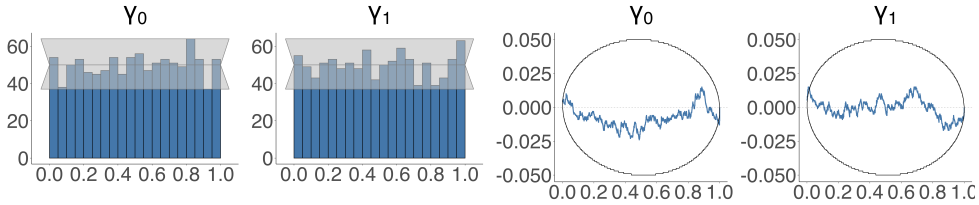
(e) Low rank **Q** uncertainty method (mesh B)

Figure C.7: Histogram and ECDF difference plot of the normalized ranks p_k using Algorithm 6.2 for the second-stage model parameters γ_0 and γ_1 out of 1000 data replicates for the two-stage Gaussian spatial model (Section 6.3.1) using INLA-SPDE and with different approaches: (a) plug-in method (b) resampling method (c) full **Q** method (d) low rank **Q** method (mesh A) (e) low rank **Q** method (mesh B)

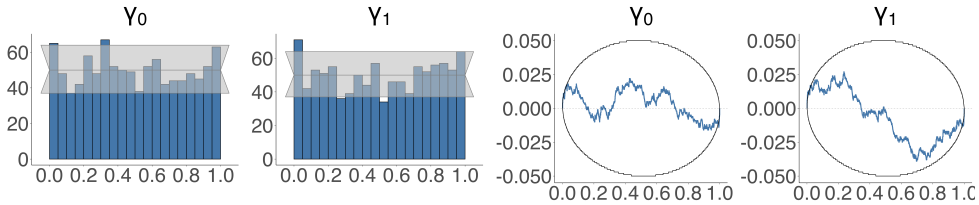
C.1.4 Results for γ_0 and γ_1 using Algorithm 6.3



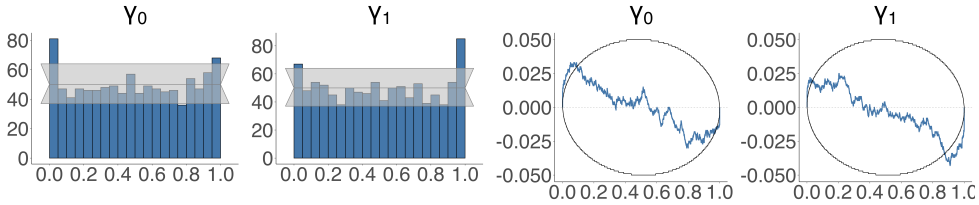
(a) Plug-in method



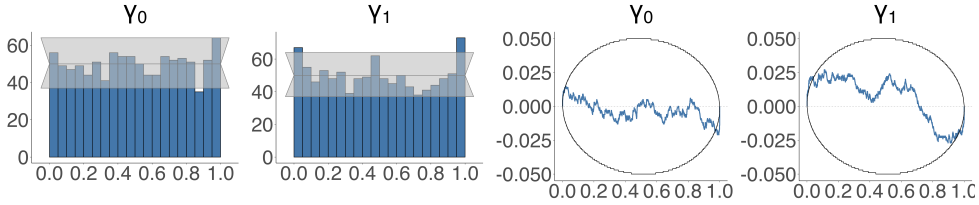
(b) Resampling method



(c) Full **Q** uncertainty method



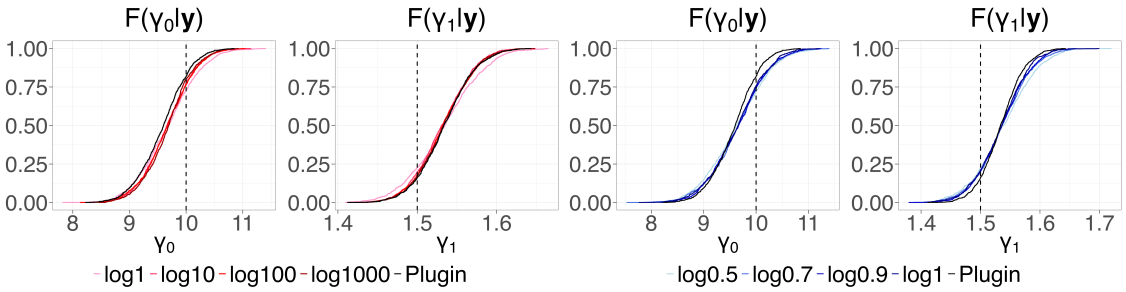
(d) Low rank **Q** uncertainty method (mesh A)



(e) Low rank **Q** uncertainty method (mesh B)

Figure C.8: Histogram and ECDF difference plot of the normalized ranks p_k using Algorithm 6.3 for the second-stage model parameters γ_0 and γ_1 out of 1000 data replicates for the two-stage Gaussian spatial model (Section 6.3.1) using INLA-SPDE and with different approaches: (a) plug-in method (b) resampling method (c) full **Q** method (d) low rank **Q** method (mesh A) (e) low rank **Q** method (mesh B)

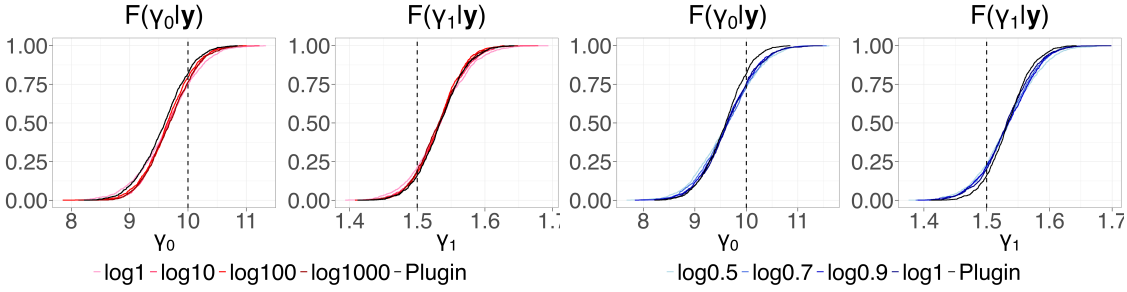
C.1.5 Illustration with a simulated data



(a) Comparison with high log precision values

(b) Comparison with low log precision values

Figure C.9: Comparison of the estimated posterior CDFs of the second-stage model parameters γ_0 and γ_1 for different values of the log precision of the error component in the low rank $\mathbf{Q}$ uncertainty method (mesh A) using the simulated data example in Section 6.3.1



(a) Comparison with high log precision values

(b) Comparison with low log precision values

Figure C.10: Comparison of the estimated posterior CDFs of the second-stage model parameters γ_0 and γ_1 for different values of the log precision of the error component in the low rank $\mathbf{Q}$ uncertainty method (mesh B) using the simulated data example in Section 6.3.1

C.2 SBC results for the Poisson model (Section 6.3.2)

C.2.1 Results for the first-stage model using Algorithm 6.2

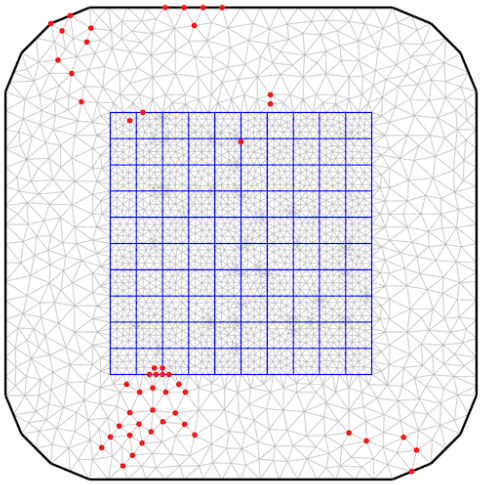


Figure C.11: Results of the KS goodness-of-fit test for uniformity (at 10% significance level) of the normalized ranks p_k of the SPDE (mesh nodes) weights out of 1000 data replicates and using Algorithm 6.2. The red points show the mesh nodes which fail the KS test for uniformity

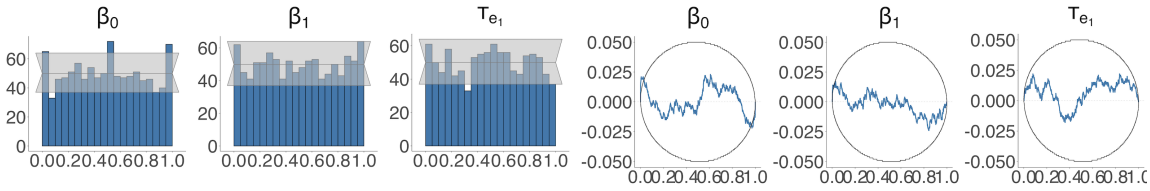


Figure C.12: Histogram and ECDF difference plot of the normalized ranks p_k for β_0, β_1, and $\tau_{e_1} = 1/\sigma^2_{e_1}$ out of 1000 data replicates using Algorithm 6.2

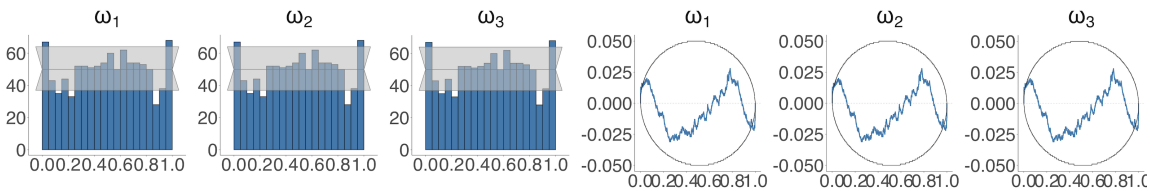


Figure C.13: Histogram and ECDF difference plot of the normalized ranks p_k for ω_1, ω_2, and ω_3 out of 1000 data replicates using Algorithm 6.2

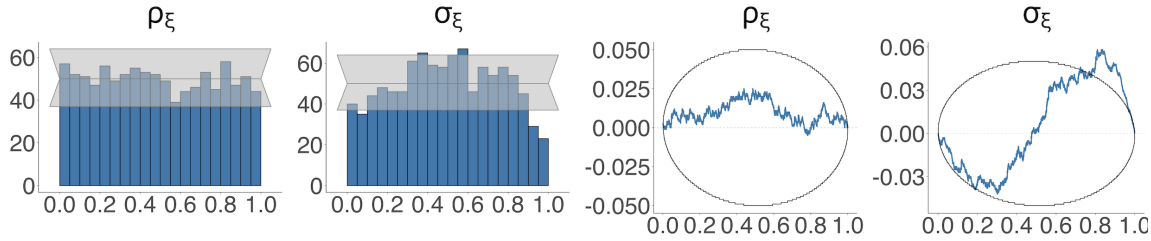


Figure C.14: Histogram and ECDF difference plot of the normalized ranks p_k for ρ_ξ and σ_ξ out of 1000 data replicates using Algorithm 6.2

C.2.2 Results for first-stage model using Algorithm 6.3

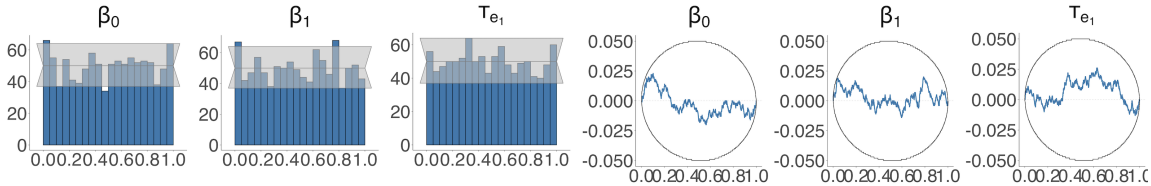


Figure C.15: Histogram and ECDF difference plot of the normalized ranks p_k for β_0, β_1, and $\tau_{e_1} = 1/\sigma_{e_1}^2$ out of 1000 data replicates and using PC prior for the Matérn parameters using Algorithm 6.3

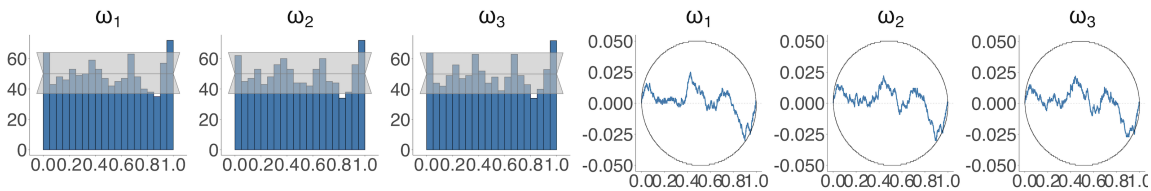
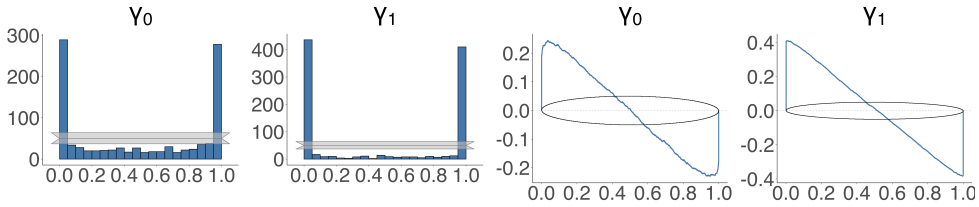
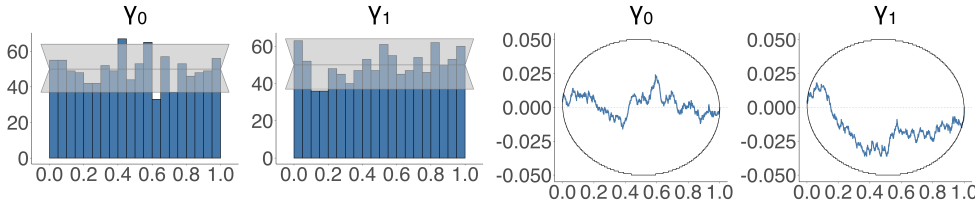


Figure C.16: Histogram and ECDF difference plot of the normalized ranks p_k for ω_1, ω_2, and ω_3 out of 1000 data replicates and using PC prior for the Matérn parameters using Algorithm 6.3

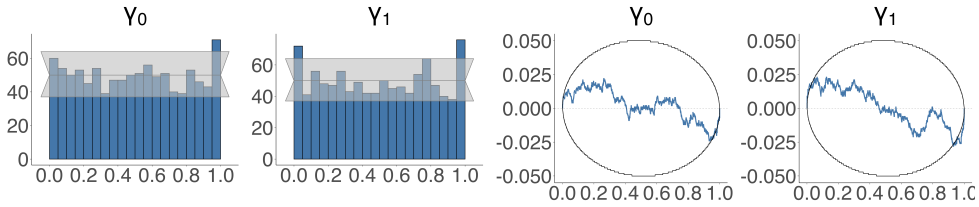
C.2.3 Results for γ_0 and γ_1 of the classical Poisson model specification using Algorithm 6.2



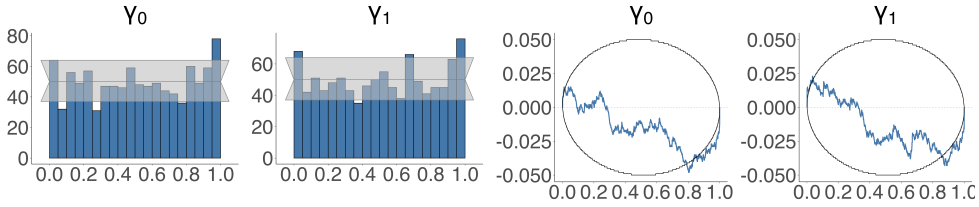
(a) Plug-in method



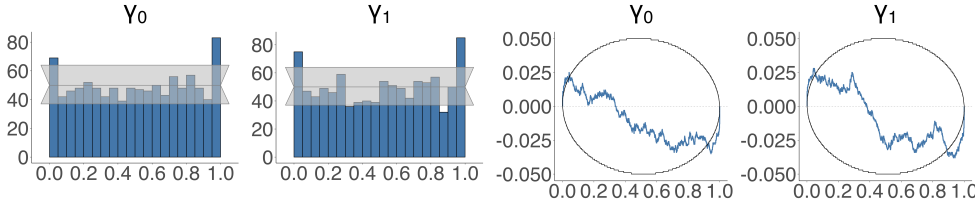
(b) Resampling method



(c) Full **Q** uncertainty method



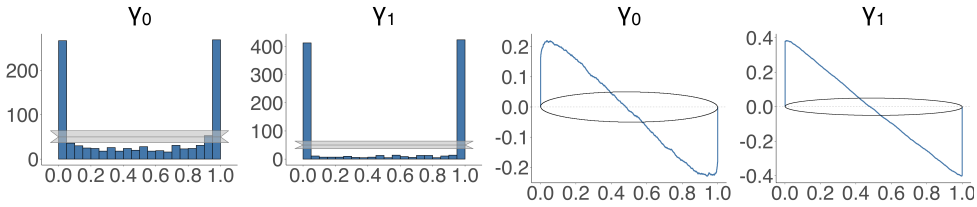
(d) Low rank **Q** uncertainty method (mesh A)



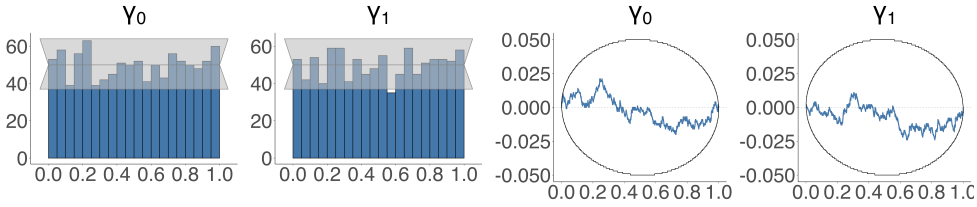
(e) Low rank **Q** uncertainty method (mesh B)

Figure C.17: Histogram and ECDF difference plot of the normalized ranks p_k using Algorithm 6.2 for the second-stage model parameters γ_0 and γ_1 out of 1000 data replicates for the classical specification of the two-stage Poisson spatial model (Section 6.3.2) using INLA-SPDE and with different approaches: (a) plug-in method (b) resampling method (c) full **Q** method (d) low rank **Q** method (mesh A) (e) low rank **Q** method (mesh B)

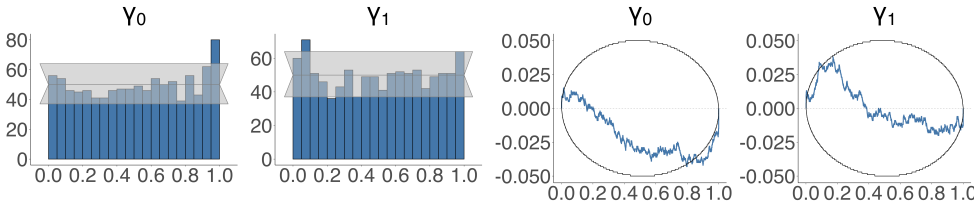
C.2.4 Results for γ_0 and γ_1 of the classical Poisson model specification using Algorithm 6.3



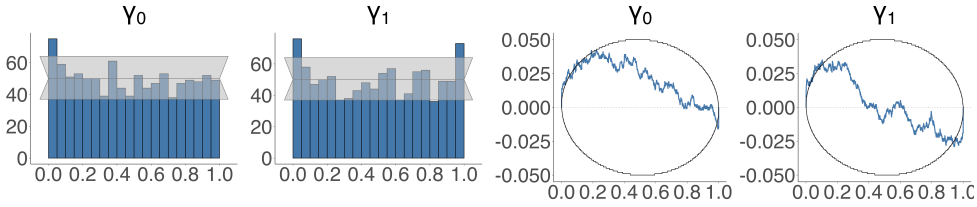
(a) Plug-in method



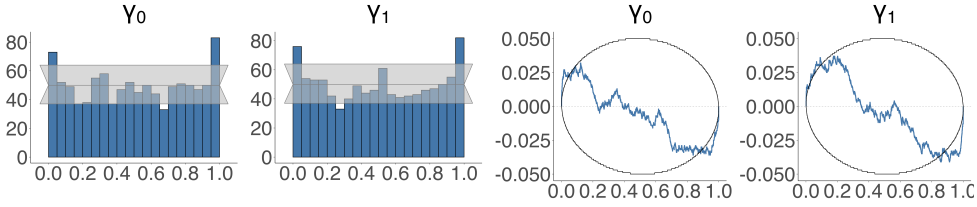
(b) Resampling method



(c) Full **Q** uncertainty method



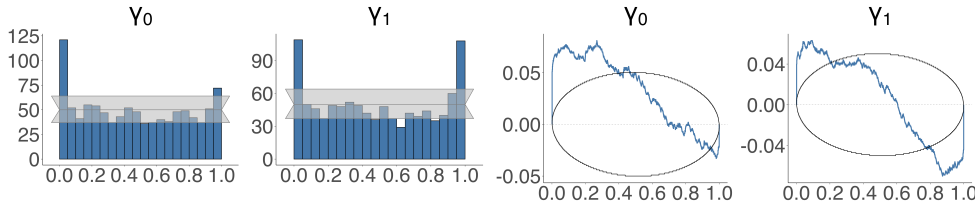
(d) Low rank **Q** uncertainty method (mesh A)



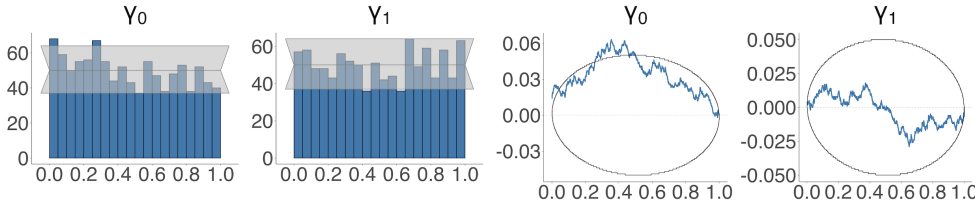
(e) Low rank **Q** uncertainty method (mesh B)

Figure C.18: Histogram and ECDF difference plot of the normalized ranks p_k using Algorithm 6.3 for the second-stage model parameters γ_0 and γ_1 out of 1000 data replicates for the classical specification of the two-stage Poisson spatial model (Section 6.3.2) using INLA-SPDE and with different approaches: (a) plug-in method (b) resampling method (c) full **Q** method (d) low rank **Q** method (mesh A) (e) low rank **Q** method (mesh B)

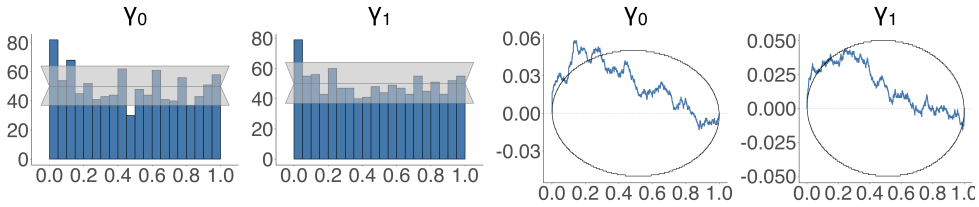
C.2.5 Results for γ_0 and γ_1 of the new Poisson model specification using Algorithm 6.2



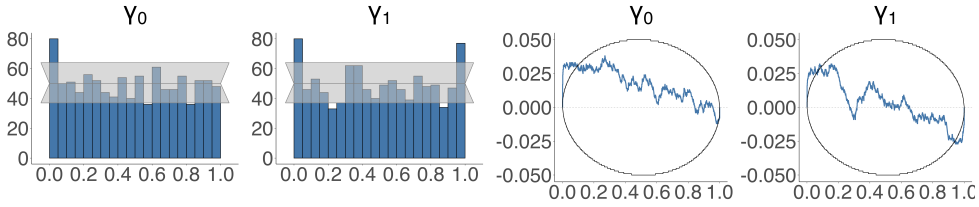
(a) Plug-in method



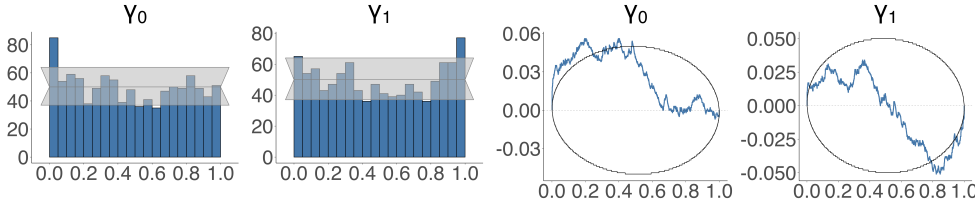
(b) Resampling method



(c) Full **Q** uncertainty method



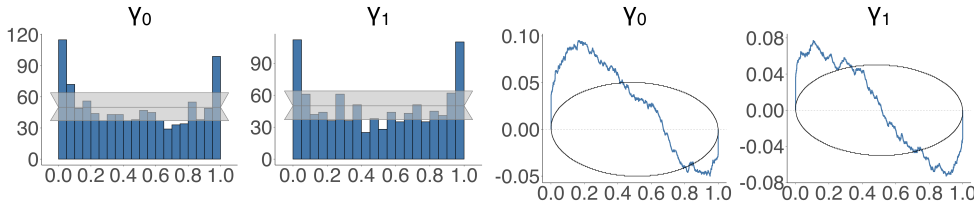
(d) Low rank **Q** uncertainty method (mesh A)



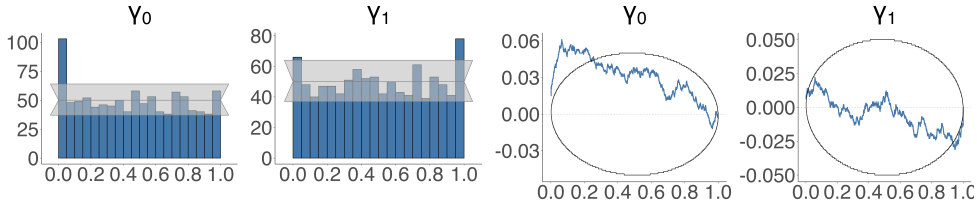
(e) Low rank **Q** uncertainty method (mesh B)

Figure C.19: Histogram and ECDF difference plot of the normalized ranks p_k using Algorithm 6.2 for the second-stage model parameters γ_0 and γ_1 out of 1000 data replicates for the new specification of the two-stage Poisson spatial model (Section 6.3.2) using INLA-SPDE and with different approaches: (a) plug-in method (b) resampling method (c) full **Q** method (d) low rank **Q** method (mesh A) (e) low rank **Q** method (mesh B)

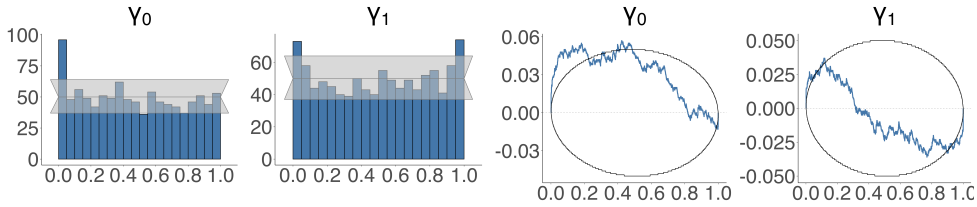
C.2.6 Results for γ_0 and γ_1 of the new Poisson model specification using Algorithm 6.3



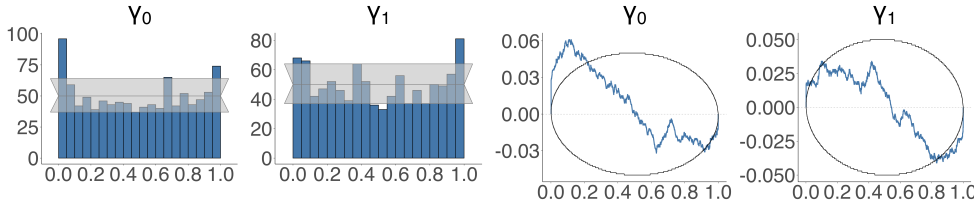
(a) Plug-in method



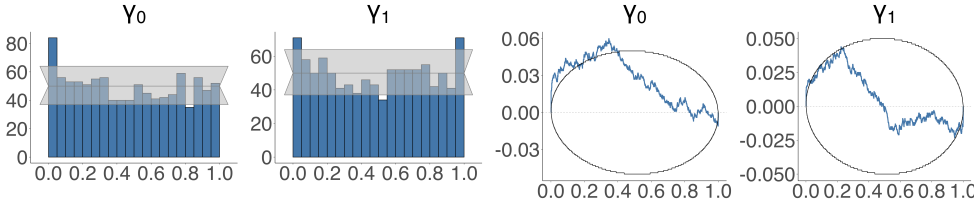
(b) Resampling method



(c) Full **Q** uncertainty method



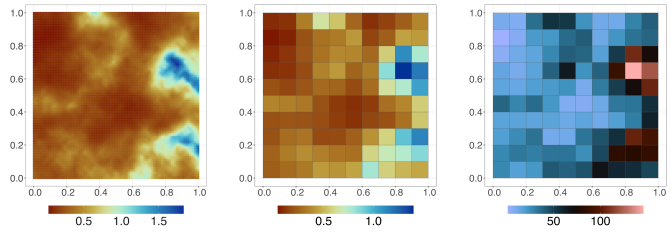
(d) Low rank **Q** uncertainty method (mesh A)



(e) Low rank **Q** uncertainty method (mesh B)

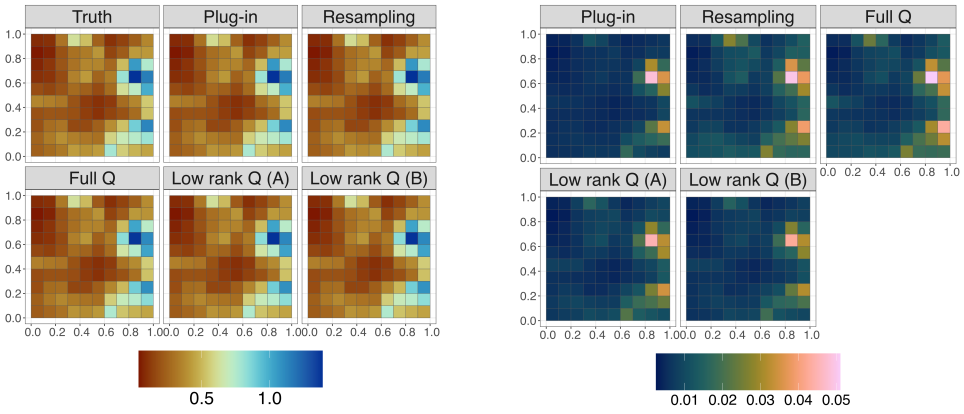
Figure C.20: Histogram and ECDF difference plot of the normalized ranks p_k using Algorithm 6.3 for the second-stage model parameters γ_0 and γ_1 out of 1000 data replicates for the new specification of the two-stage Poisson spatial model (Section 6.3.2) using INLA-SPDE and with different approaches: (a) plug-in method (b) resampling method (c) full **Q** method (d) low rank **Q** method (mesh A) (e) low rank **Q** method (mesh B)

C.2.7 Illustration with a simulated data



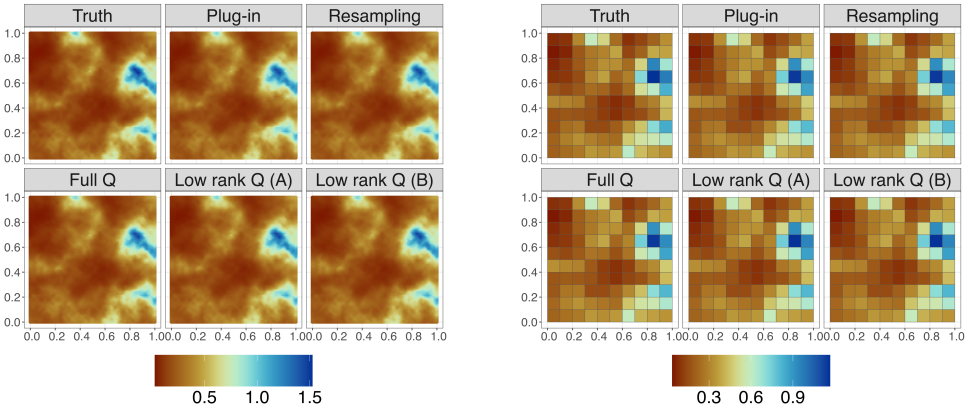
(a) $\lambda(\mathbf{s})$ **(b)** $\lambda(B)$ **(c)** $Y(B)$

Figure C.21: Simulated quantities from the new specification of the two-stage Poisson spatial model in Section 6.3.2



(a) Posterior mean **(b)** Posterior standard deviation

Figure C.22: Comparison of (a) the posterior mean and (b) posterior standard deviation of $\lambda(B)$ from a simulated data of the two-stage Poisson spatial model (classical specification) in Section 6.3.2 using different approaches: the plug-in method, resampling method, full $\mathbf{Q}$ method, low rank $\mathbf{Q}$ (mesh A) method, low rank $\mathbf{Q}$ (mesh B) method



(a) $\lambda(\mathbf{s})$ **(b)** $\lambda(B) = \frac{1}{|B|}\int_B \lambda(\mathbf{s})d\mathbf{s}$

Figure C.23: Comparison of the posterior mean for (a) $\lambda(\mathbf{s})$ and (b) $\lambda(B)$ from a simulated data of the two-stage Poisson model (new specification) in Section 6.3.2 using different approaches: the plug-in method, resampling method, full $\mathbf{Q}$ method, low rank $\mathbf{Q}$ (mesh A) method, and low rank $\mathbf{Q}$ (mesh B) method

C.3 SBC results for non-spatial two-stage models

C.3.1 Gaussian model

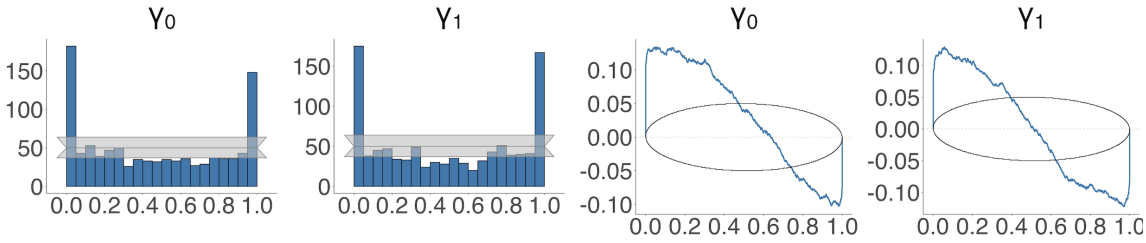


Figure C.24: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using INLA-SPDE and the plug-in method for the two-stage Gaussian model

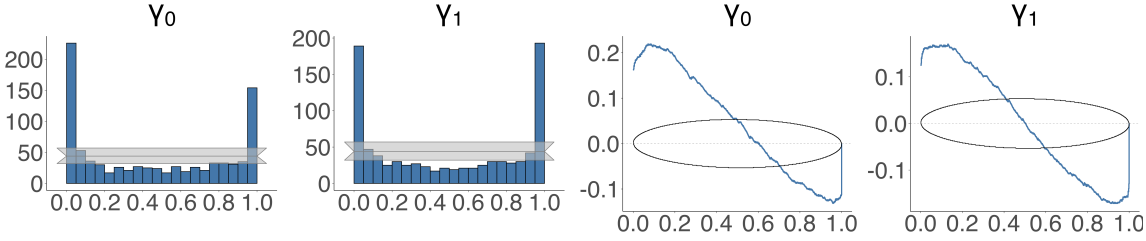


Figure C.25: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using NUTS and the plug-in method for the two-stage Gaussian model

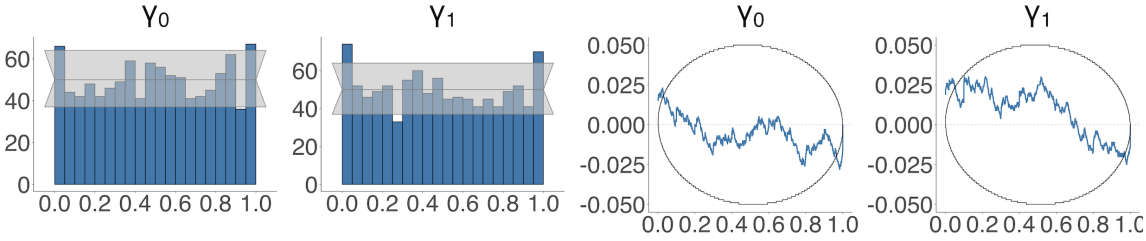


Figure C.26: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using INLA-SPDE and the resampling method for the two-stage Gaussian model

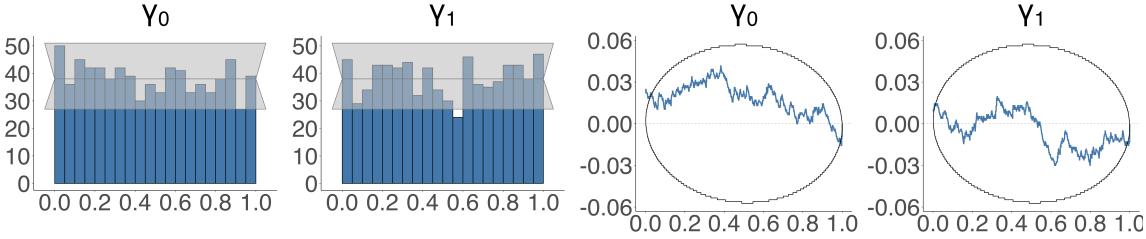


Figure C.27: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using NUTS and the resampling method for the two-stage Gaussian model

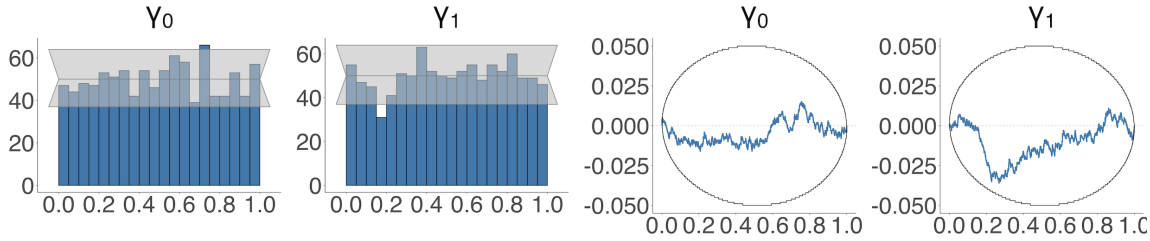


Figure C.28: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using the full **Q** method for the two-stage Gaussian model

C.3.2 Poisson model - classical specification

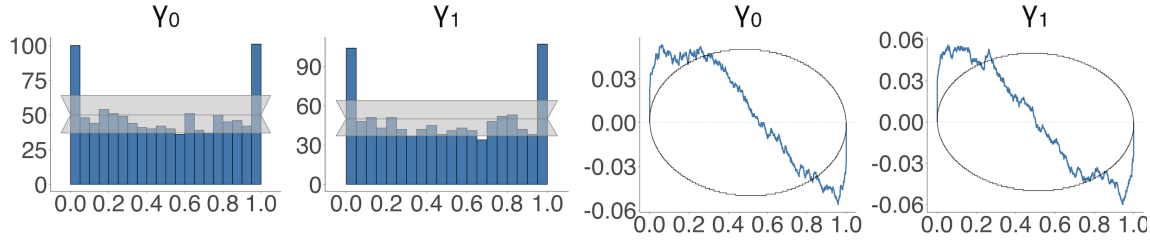


Figure C.29: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using INLA-SPDE and the plug-in method for the two-stage Poisson model (classical specification)

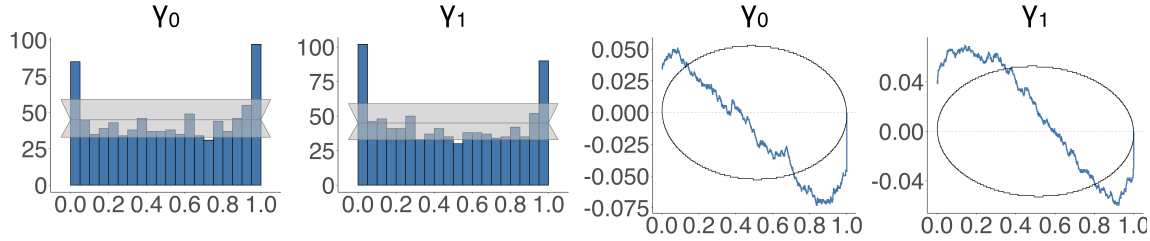


Figure C.30: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using NUTS and the plug-in method (classical specification)

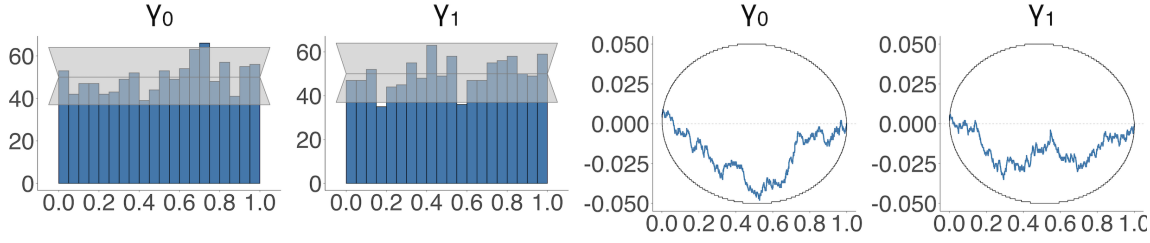


Figure C.31: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using INLA-SPDE and the resampling method for the two-stage Poisson model (classical specification)

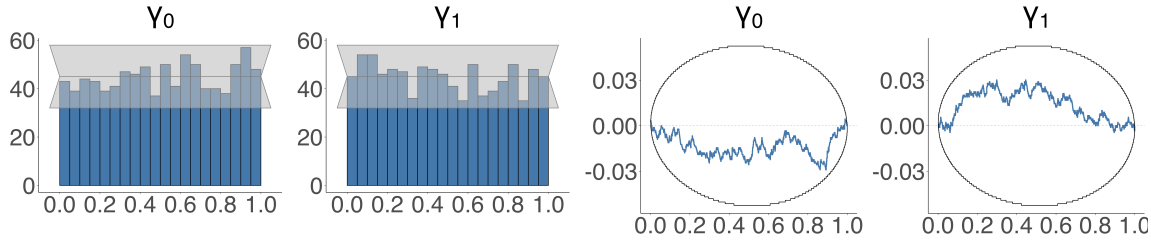


Figure C.32: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using NUTS and the resampling method for the two-stage Poisson model (classical specification)

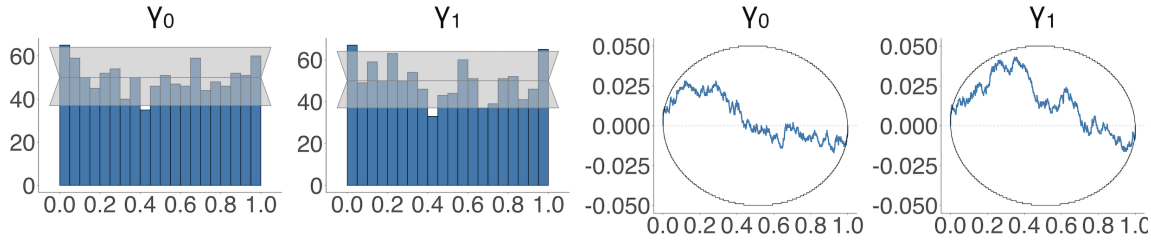


Figure C.33: Histogram and ECDF difference plot of the normalized ranks p_k for γ_0 and γ_1 out of 1000 data replicates using the full **Q** method for the two-stage Poisson model

C.4 Data application

	Mean	SD	P2.5	P97.5
β_0	78.1553	4.1435	70.0341	86.2765
β_1	-0.0058	0.0033	-0.0122	0.0006
β_2	3.2959	0.6142	2.0920	4.4998
β_3	-0.1124	0.0210	-0.1535	-0.0713
$1/\sigma^2_{e_1}$	0.1448	0.0325	0.0910	0.2181
$\log(\tau)$	3.8219	0.6807	2.4272	5.1048
$\log(\kappa)$	-6.1369	0.6244	-7.3248	-4.8675

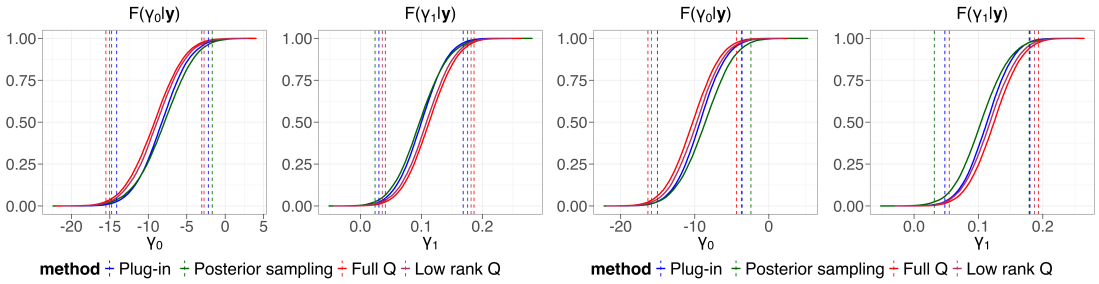
Table C.1: Posterior estimates of first-stage model parameters: posterior mean, posterior standard deviation (SD), and 95% credible intervals

Method		Mean	SD	P2.5	P97.5
Plug-in	γ_0	-8.1913	3.0307	-14.1313	-2.2513
	γ_1	0.1006	0.0353	0.0313	0.1698
Resampling	γ_0	-7.9764	3.3459	-14.7697	-1.6871
	γ_1	0.0981	0.0388	0.0240	0.1758
Full **Q**	γ_0	-9.2889	3.1863	-15.5341	-3.0438
	γ_1	0.1131	0.0371	0.0403	0.1859
Low rank **Q**	γ_0	-8.8892	3.1837	-15.1291	-2.6493
	γ_1	0.1086	0.0371	0.0357	0.1814

Table C.2: Posterior estimates of second-stage model (classical specification)

Method		Mean	SD	P2.5	P97.5
Plug-in	γ_0	-9.3117	2.9108	-15.0168	-3.6067
	γ_1	0.1131	0.0336	0.0473	0.1790
Resampling	γ_0	-8.6326	3.2423	-15.0377	-2.4195
	γ_1	0.1048	0.0372	0.0314	0.1788
Full **Q**	γ_0	-10.2863	3.0618	-16.2873	-4.2852
	γ_1	0.1240	0.0354	0.0546	0.1933
Low rank **Q**	γ_0	-9.7508	3.0567	-15.7418	-3.7597
	γ_1	0.1180	0.0354	0.0487	0.1874

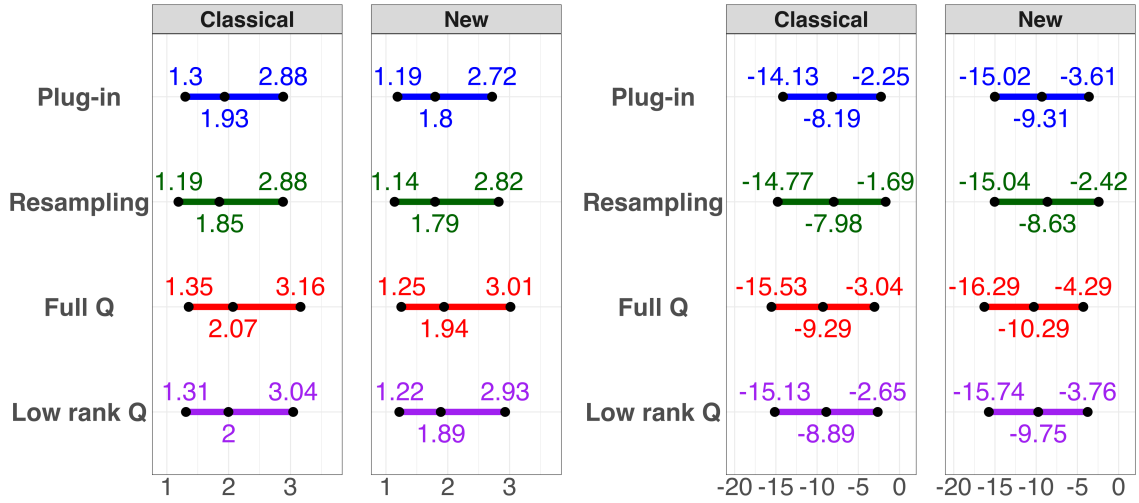
Table C.3: Posterior estimates of second-stage model (new specification)



(a) Classical model specification

(b) New model specification

Figure C.34: Comparison of marginal posteriors of γ_0 and γ_1 using four uncertainty propagation approaches: (a) classical model specification (b) new model specification



(a) 95% CI of RR with 1 SD change in RH

(b) 95% CI for γ_0

Figure C.35: (a) 95% CI of RR associated with 1 SD change in relative humidity (b) 95% CI for γ_0

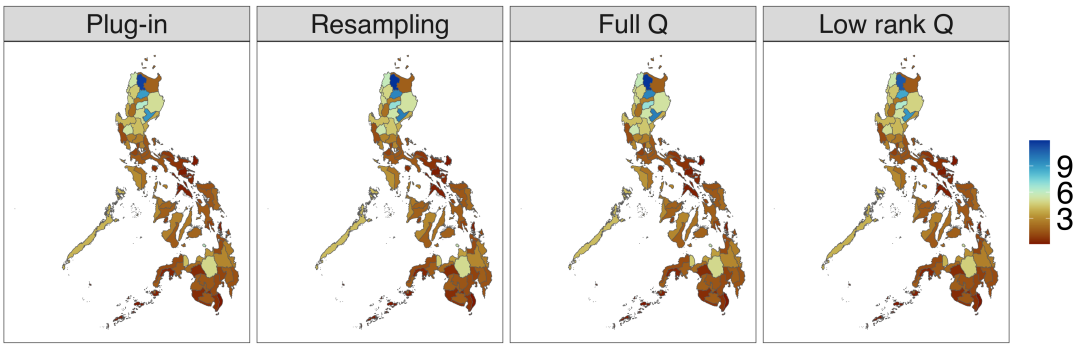


Figure C.36: Posterior means of $\lambda(B)$ using the new specification of the Poisson model

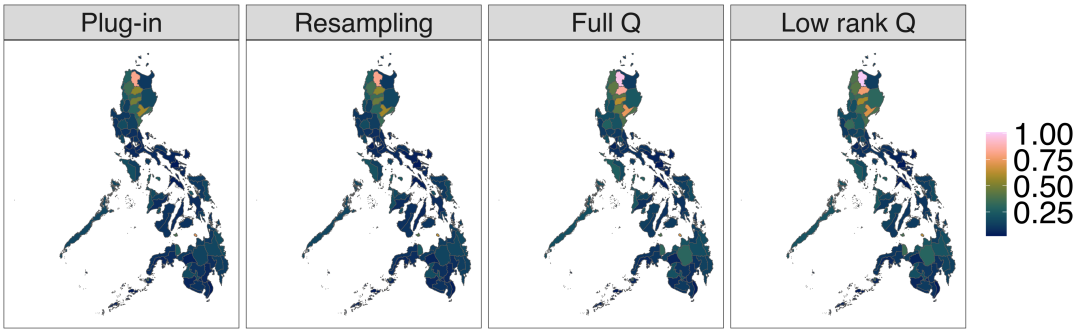


Figure C.37: Posterior standard deviations of $\lambda(B)$ using the new specification of the Poisson model

References

Abdullah, N. A. M. H., Dom, N. C., Salleh, S. A., Salim, H., and Precha, N. (2022). The association between dengue case and climate: A systematic review and meta-analysis. *One Health*, page 100452. 3, 98

Adin, A., Krainski, E., Lenzi, A., Liu, Z., Martínez-Minaya, J., and Rue, H. (2023). Automatic cross-validation in structured models: Is it time to leave out leave-one-out? *arXiv preprint arXiv:2311.17100*. 96, 132

Arab, A., Jackson, M. C., and Kongoli, C. (2014). Modelling the effects of weather and climate on malaria distributions in west africa. *Malaria journal*, 13:1–9. 11

August, T., Harvey, M., Lightfoot, P., Kilbey, D., Papadopoulos, T., and Jepson, P. (2015). Emerging technologies for biological recording. *Biological Journal of the Linnean Society*, 115(3):731–749. 137

Bachl, F. E., Lindgren, F., Borchers, D. L., and Illian, J. B. (2019). inlabru: an R package for Bayesian spatial modelling from ecological survey data. *Methods in Ecology and Evolution*, 10(6):760–766. 138, 167

Bakka, H., Rue, H., Fuglstad, G.-A., Riebler, A., Bolin, D., Illian, J., Krainski, E., Simpson, D., and Lindgren, F. (2018). Spatial modeling with r-inla: A review. *Wiley Interdisciplinary Reviews: Computational Statistics*, 10(6):e1443. 141

Bakka, H., Vanhatalo, J., Illian, J. B., Simpson, D., and Rue, H. (2019). Non-stationary gaussian models with physical barriers. *Spatial statistics*, 29:268–288. 110

Banerjee, S., Carlin, B. P., and Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data*. CRC press. 23, 32, 33, 37

Banerjee, S. and Gelfand, A. (2002). Prediction, interpolation and regression for spatially misaligned data. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 227–245. 185

Bauer, P., Thorpe, A., and Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, 525(7567):47–55. 11, 135

Bayarri, M., Berger, J., and Liu, F. (2009). Modularization in bayesian analysis, with emphasis on analysis of computer models. *Bayesian Analysis*, 4(1):119–150. 9

BBC (2019). Philippines declares dengue epidemic as deaths surge. https://www.bbc.co.uk/news/world-asia-49249144. Accessed: 2025-17-02. 3, 4, 142, 143, 162

Belmont, J., Martino, S., Illian, J., and Rue, H. (2024). Spatio-temporal Occupancy Models with INLA. *arXiv preprint arXiv:2403.10680*. 137

Berild, M. O., Martino, S., Gómez-Rubio, V., and Rue, H. (2022). Importance sampling with the integrated nested laplace approximation. *Journal of Computational and Graphical Statistics*, 31(4):1225–1237. 176, 218, 224

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010a). A bivariate space-time downscaler under space and time misalignment. *The annals of applied statistics*, 4(4):1942. 57

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2010b). A spatio-temporal downscaler for output from numerical models. *Journal of agricultural, biological, and environmental statistics*, 15(2):176–197. 56, 65

Berrocal, V. J., Gelfand, A. E., and Holland, D. M. (2012). Space-time data fusion under error in computer model output: an application to modeling air quality. *Biometrics*, 68(3):837–848. 57

Berry, S. M., Carroll, R. J., and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association*, 97(457):160–169. 138, 185

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, 36(2):192–225. 32, 59

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, 43:1–20. 71, 148

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press. 38

Bivand, R. and Piras, G. (2015). Comparing implementations of estimation methods for spatial econometrics. *Journal of Statistical Software*, 63:1–36. 74, 114

Blangiardo, M. and Cameletti, M. (2015). *Spatial and spatio-temporal Bayesian models with R-INLA*. John Wiley & Sons. 71, 72, 141

Blangiardo, M., Finazzi, F., and Cameletti, M. (2016). Two-stage bayesian model to evaluate the effect of air pollution on chronic respiratory diseases using drug prescriptions. *Spatial and spatio-temporal epidemiology*, 18:1–12. 7, 9, 10, 11, 73, 138, 141, 173, 191, 192, 202, 209

Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. *Journal of the American statistical Association*, 88(421):9–25. 35, 36

Brook, D. (1964). On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems. *Biometrika*, 51(3/4):481–483. 32

Bruno, F., Cameletti, M., Franco-Villoria, M., Greco, F., Ignaccolo, R., Ippoliti, L., Valentini, P., and Ventrucci, M. (2016). A survey on ecological regression for health hazard associated with air pollution. *Spatial statistics*, 18:276–299. 12, 62

Buczak, A. L., Baugher, B., Babin, S. M., Ramac-Thomas, L. C., Guven, E., Elbert, Y., Koshute, P. T., Velasco, J. M. S., Roque Jr, V. G., Tayag, E. A., et al. (2014).

Prediction of high incidence of dengue in the philippines. *PLoS neglected tropical diseases*, 8(4):e2771. 140

Cameletti, M., Gómez-Rubio, V., and Blangiardo, M. (2019). Bayesian modelling for spatially misaligned health and air pollution data through the inla-spde approach. *Spatial Statistics*, 31:100353. 7, 10, 13, 62, 72, 73, 74, 76, 77, 90, 141, 147, 170, 174, 191, 209

Cameletti, M., Lindgren, F., Simpson, D., and Rue, H. (2013). Spatio-temporal modeling of particulate matter concentration through the spde approach. *AStA Advances in Statistical Analysis*, 97(2):109–131. 63, 102, 111, 141

Carlin, B. P. and Louis, T. A. (2008). *Bayesian methods for data analysis*. CRC press. 37

Carvajal, T. M., Viacrusis, K. M., Hernandez, L. F. T., Ho, H. T., Amalin, D. M., and Watanabe, K. (2018). Machine learning methods reveal the temporal pattern of dengue incidence using meteorological factors in metropolitan manila, philippines. *BMC infectious diseases*, 18:1–15. 140, 168

Cawiding, O. R., Jeon, S., Tubera-Panes, D., de los Reyes V, A. A., and Kim, J. K. (2025). Disentangling climate's dual role in dengue dynamics: A multiregion causal analysis study. *Science Advances*, 11(7):eadq1901. 140, 145, 155, 156, 165, 231

CDCP (2025). Areas with risk of dengue. https://www.cdc.gov/dengue/areas-with-risk/index.html. Accessed: 2025-03-12.

Chakraborty, A., Nott, D. J., Drovandi, C. C., Frazier, D. T., and Sisson, S. A. (2023). Modularized bayesian analyses and cutting feedback in likelihood-free inference. *Statistics and Computing*, 33(1):33. 9

Chang, H. H., Peng, R. D., and Dominici, F. (2011). Estimating the acute health effects of coarse particulate matter accounting for exposure measurement error. *Biostatistics*, 12(4):637–652. 170, 174

Chen, C., Chen, Q., Li, G., He, M., Dong, J., Yan, H., Wang, Z., and Duan, Z. (2021). A novel multi-source data fusion method based on Bayesian inference for

accurate estimation of chlorophyll-a concentration over eutrophic lakes. *Environmental Modelling & Software*, 141:105057. 56

Chien, L.-C. and Yu, H.-L. (2014). Impact of meteorological factors on the spatiotemporal patterns of dengue fever incidence. *Environment international*, 73:46–56. 11

Chiles, J.-P. and Delfiner, P. (2012). *Geostatistics: modeling spatial uncertainty*, volume 713. John Wiley & Sons. 27, 29

Christophers, S. R. (1911). Epidemic malaria of the punjab: with a note of a method of predicting epidemic years. *Trans Committee Stud Malaria India*, 2:17–26.

Ciarlet, P. G. (2002). *The finite element method for elliptic problems*. SIAM. 52

Colón-González, F. J., Sewe, M. O., Tompkins, A. M., Sjödin, H., Casallas, A., Rocklöv, J., Caminade, C., and Lowe, R. (2021). Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study. *The Lancet Planetary Health*, 5(7):e404–e414. 3, 144

Cook, S. R., Gelman, A., and Rubin, D. B. (2006). Validation of software for bayesian models using posterior quantiles. *Journal of Computational and Graphical Statistics*, 15(3):675–692. 178

Coronas, J. (1920). *The Climate and Weather of the Philippines, 1903-1918, by Rev. José Coronas. SJ, Chief, Meteorological Division, Weather Bureau, Manila Observatory*. Manila,: Bureau of Printing. 96, 128, 130, 155

Couper, L. I., Farner, J. E., Caldwell, J. M., Childs, M. L., Harris, M. J., Kirk, D. G., Nova, N., Shocket, M., Skinner, E. B., Uricchio, L. H., et al. (2021). How will mosquitoes adapt to climate warming? *Elife*, 10:e69630. 3

Cressie, N. (1985). Fitting variogram models by weighted least squares. *Journal of the international Association for mathematical Geology*, 17:563–586. 24

Cressie, N. (1988). Spatial prediction and ordinary kriging. *Mathematical geology*, 20:405–421. 20

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons. 21, 27, 30

Cruz, E. I., Salazar, F. V., Aguila, A. M. A., Villaruel-Jagmis, M. V., Ramos, J., and Paul, R. E. (2024). Current and lagged associations of meteorological variables and aedes mosquito indices with dengue incidence in the philippines. *PLOS Neglected Tropical Diseases*, 18(7):e0011603. 140, 168

Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. *Annals of Statistics*, pages 65–81. 117

DEFRA (2024). Uk Air Information Resource. https://uk-air.defra.gov.uk/research/air-quality-modelling?view=modelling. Accessed: 2024-04-04. 137

Diggle, P. J. (2013). *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press.

Dulay, A. V. S., Bautista, J. R., and Teves, F. G. (2013). Climate change and incidence of dengue fever (df) and dengue hemorrhagic fever (dhf) in iligan city, lanao del norte, philippines. *Internasional Research Journal of Biological Sciences*, 2(7):37–41. 140

Duque-Lee, C. D., Yu, A. K. D., Ytienza, S. I. E., Yu, A. M. D., Yu, V. C. S., Wangkay, K. A. K., Wong, M. A. R., Zhang, E. M. T., Yumul, W. D., Zipagan, Z. M. R., et al. (2020). Correlation between incidence of dengue and climatic factors in the philippines: An ecological study. *Health Sciences Journal*, 9(2):1–1. 140

ECDC (2023). Factsheet for health professionals about dengue. https://www.ecdc.europa.eu/en/dengue-fever/facts. Accessed: 2025-02-27. 2

ECDC (2024). An emerging threat: mosquito-borne diseases in europe. 146

eClinicalMedicine (2024). Dengue as a growing global health concern. *eClinicalMedicine*, 77. 2

Edillo, F., Ymbong, R. R., Bolneo, A. A., Hernandez, R. J., Fuentes, B. L., Cortes, G., Cabrera, J., Lazaro, J. E., and Sakuntabhai, A. (2022). Temperature, season, and latitude influence development-related phenotypes of philippine aedes aegypti

(linnaeus): Implications for dengue control amidst global warming. *Parasites & Vectors*, 15(1):74. 140, 145

Edillo, F., Ymbong, R. R., Navarro, A. O., Cabahug, M. M., and Saavedra, K. (2024). Detecting the impacts of humidity, rainfall, temperature, and season on chikungunya, dengue and zika viruses in aedes albopictus mosquitoes from selected sites in cebu city, philippines. *Virology Journal*, 21(1):42. 140

Edillo, F. E., Halasa, Y. A., Largo, F. M., Erasmo, J. N. V., Amoin, N. B., Alera, M. T. P., Yoon, I.-K., Alcantara, A. C., and Shepard, D. S. (2015). Economic cost and burden of dengue in the philippines. *The American journal of tropical medicine and hygiene*, 92(2):360. 139

Ewing, D. A., Cobbold, C. A., Purse, B., Nunn, M., and White, S. M. (2016). Modelling the effect of temperature on the seasonal population dynamics of temperate mosquitoes. *Journal of theoretical biology*, 400:65–79. 144, 145

Flores, J. and Balagot, V. (1969). Climate of the philippines. in: Arakawa, h. (ed.). *Climate of Northern and Eastern Asia, World Survey of Climatology*, 8.

Focks, D. A., Daniels, E., Haile, D. G., Keesling, J. E., et al. (1995). A simulation model of the epidemiology of urban dengue fever: literature analysis, model development, preliminary validation, and samples of simulation results. *American journal of tropical medicine and hygiene*, 53(5):489–506. 145

Forlani, C., Bhatt, S., Cameletti, M., Krainski, E., and Blangiardo, M. (2020). A joint Bayesian space–time model to integrate spatially misaligned air pollution data in R-INLA. *Environmetrics*, 31(8):e2644. 13, 58, 65, 135, 137, 214

Foster, B. (2001). Ipcc third assessment report. *The Scientific Basis: Geneva, Switzerland*. 168

Francisco, M. E., Carvajal, T. M., Ryo, M., Nukazawa, K., Amalin, D. M., and Watanabe, K. (2021). Dengue disease dynamics are modulated by the combined influences of precipitation and landscape: A machine learning approach. *Science of The Total Environment*, 792:148406. 140, 145

Franco-Villoria, M., Ventrucci, M., and Rue, H. (2022). Variance partitioning in spatio-temporal disease mapping models. *Statistical Methods in Medical Research*, 31(8):1566–1578. 166, 232

Freni-Sterrantino, A., Ventrucci, M., and Rue, H. (2018). A note on intrinsic conditional autoregressive models for disconnected graphs. *Spatial and spatio-temporal epidemiology*, 26:25–34. 149, 150

Fuentes, M. and Raftery, A. E. (2001). Model validation and spatial interpolation by combining observations with outputs from numerical models via bayesian melding. Technical report, WASHINGTON UNIV SEATTLE DEPT OF STATISTICS.

Fuentes, M. and Raftery, A. E. (2005). Model evaluation and spatial interpolation by bayesian combination of observations with outputs from numerical models. *Biometrics*, 61(1):36–45. 55, 210

Fuglstad, G.-A., Hem, I. G., Knight, A., Rue, H., and Riebler, A. (2020). Intuitive joint priors for variance parameters. *Bayesian Analysis*, 15(4):1109–1137. 232

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2019). Constructing priors that penalize the complexity of gaussian random fields. *Journal of the American Statistical Association*, 114(525):445–452. 78, 79, 103, 111, 122, 186

Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M. (2010). *Handbook of spatial statistics*. CRC press. 20, 21, 147

Gelfand, A. E., Schmidt, A. M., Banerjee, S., and Sirmans, C. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test*, 13:263–312. 57

Gelfand, A. E. and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410):398–409. 37

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (1995). *Bayesian data analysis*. Chapman and Hall/CRC.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., and Modrák, M. (2020). Bayesian workflow. *arXiv preprint arXiv:2011.01808*. 170, 207, 215

Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, pages 721–741. 32, 37

Gettelman, A., Geer, A. J., Forbes, R. M., Carmichael, G. R., Feingold, G., Posselt, D. J., Stephens, G. L., van den Heever, S. C., Varble, A. C., and Zuidema, P. (2022). The future of earth system prediction: Advances in model-data fusion. *Science Advances*, 8(14):eabn3488. 11, 135

Ghysels, E., Kvedaras, V., and Zemlys, V. (2016). Mixed frequency data sampling regression models: The r package midasr. *Journal of statistical software*, 72:1–35. 225

Ghysels, E., Kvedaras, V., and Zemlys-Balevičius, V. (2020). Mixed data sampling (midas) regression models. In *Handbook of statistics*, volume 42, pages 117–153. Elsevier. 224

Ghysels, E. and Qian, H. (2019). Estimating midas regressions via ols with polynomial parameter profiling. *Econometrics and statistics*, 9:1–16. 225

Ghysels, E., Sinko, A., and Valkanov, R. (2007). Midas regressions: Further results and new directions. *Econometric reviews*, 26(1):53–90. 226

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American statistical Association*, 102(477):359–378. 118

Gómez-Rubio, V. (2020). *Bayesian inference with INLA*. Chapman and Hall/CRC. 112, 176

Gómez-Rubio, V., Bivand, R. S., and Rue, H. (2020). Bayesian model averaging with the integrated nested laplace approximation. *Econometrics*, 8(2):23. 102, 213, 218

Gómez-Rubio, V. and Rue, H. (2018). Markov chain monte carlo with the integrated nested laplace approximation. *Statistics and Computing*, 28:1033–1051. 176, 218, 224

Goody, R. (1995). *Principles of atmospheric physics and chemistry*. Oxford University Press. 126

Gotway, C. A. and Stroup, W. W. (1997). A generalized linear model approach to spatial data analysis and prediction. *Journal of Agricultural, Biological, and Environmental Statistics*, pages 157–178. 35, 36

Gotway, C. A. and Young, L. J. (2002). Combining incompatible spatial data. *Journal of the American Statistical Association*, 97(458):632–648. 20

Green, P. J., Łatuszyński, K., Pereyra, M., and Robert, C. P. (2015). Bayesian computation: a summary of the current state, and samples backwards and forwards. *Statistics and Computing*, 25:835–862. 37, 38, 39

Greven, S., Dominici, F., and Zeger, S. (2011). An approach to the estimation of chronic air pollution effects using spatio-temporal information. *Journal of the American Statistical Association*, 106(494):396–406. 5, 11

Gryparis, A., Paciorek, C. J., Zeka, A., Schwartz, J., and Coull, B. A. (2009). Measurement error caused by spatial misalignment in environmental epidemiology. *Biostatistics*, 10(2):258–274. 7, 9, 138, 141, 174, 204, 209

Gubler, D. J., Ooi, E. E., Vasudevan, S., and Farrar, J. (2014). *Dengue and dengue hemorrhagic fever*. CABI. 146

Gubler, D. J., Reiter, P., Ebi, K. L., Yap, W., Nasci, R., and Patz, J. A. (2001). Climate variability and change in the united states: potential impacts on vector- and rodent-borne diseases. *Environmental health perspectives*, 109(suppl 2):223–233. 145

Hales, S., De Wet, N., Maindonald, J., and Woodward, A. (2002). Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. *The Lancet*, 360(9336):830–834. 2, 145, 165

Hales, S., Weinstein, P., Souares, Y., and Woodward, A. (1999). El niño and the dynamics of vectorborne disease transmission. *Environmental Health Perspectives*, 107(2):99–102. 168

Hales, S., Weinstein, P., and Woodward, A. (1996). Dengue fever epidemics in the south pacific: driven by el nino southern oscillation? *The Lancet*, 348(9042):1664–1665. 168

Harville, D. A. and Jeske, D. R. (1992). Mean squared error of estimation or prediction under a general linear model. *Journal of the American Statistical Association*, 87(419):724–731. 26

Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*. 37

Henderson, C. R. (1950). Estimation of genetic parameters (abstract). *Biometrics*, 9:226–252. 28

Hoffman, M. D., Gelman, A., et al. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623. 189

Iguchi, J. A., Seposo, X. T., and Honda, Y. (2018). Meteorological factors affecting dengue incidence in davao, philippines. *BMC public health*, 18:1–10. 140

Jaakkola, T. S. and Jordan, M. I. (2000). Bayesian parameter estimation via variational methods. *Statistics and Computing*, 10:25–37. 38

Jaya, I. and Folmer, H. (2022). Spatiotemporal high-resolution prediction and mapping: methodology and application to dengue disease. *Journal of geographical systems*, pages 1–55. 11, 94

Journel, A. G. and Huijbregts, C. J. (1976). Mining geostatistics. 29

Kackar, R. N. and Harville, D. A. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. *Journal of the American Statistical Association*, 79(388):853–862. 26

Kintanar, R. L. (1984a). *Climate of the Philippines*. Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA). 96, 128, 130, 155

Kintanar, R. L. (1984b). *Climate of the Philippines*. Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA).

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in medicine*, 19(17-18):2555–2567. 71, 151, 232

Koh, J. and Opitz, T. (2023). Extreme-value modelling of migratory bird arrival dates: Insights from citizen science data. *arXiv preprint arXiv:2312.01870*. 137

Krall, J. R., Chang, H. H., Sarnat, S. E., Peng, R. D., and Waller, L. A. (2015). Current methods and challenges for epidemiological studies of the associations between chemical constituents of particulate matter and health. *Current environmental health reports*, 2:388–398. 62

Lawson, A. B., Banerjee, S., Haining, R. P., and Ugarte, M. D. (2016). *Handbook of spatial epidemiology*. CRC Press. 11, 55, 56, 57, 58, 59, 61, 65, 108, 135, 141, 144

Lee, A., Szpiro, A., Kim, S., and Sheppard, L. (2015). Impact of preferential sampling on exposure prediction and health effect inference in the context of air pollution epidemiology. *Environmetrics*, 26(4):255–267. 11, 62

Lee, D. (2011). A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and spatio-temporal epidemiology*, 2(2):79–89. 192

Lee, D., Ferguson, C., and Mitchell, R. (2009). Air pollution and health in scotland: a multicity study. *Biostatistics*, 10(3):409–423. 5

Lee, D., Mukhopadhyay, S., Rushworth, A., and Sahu, S. K. (2017). A rigorous statistical framework for spatio-temporal pollution prediction and estimation of its long-term impact on health. *Biostatistics*, 18(2):370–385. 5, 6, 7, 10, 11, 56, 65, 73, 137, 138, 141, 157, 173, 191, 192, 202, 209

Lee, S. A., Economou, T., de Castro Catão, R., Barcellos, C., and Lowe, R. (2021). The impact of climate suitability, urbanisation, and connectivity on the

expansion of dengue in 21st century brazil. *PLoS Neglected Tropical Diseases*, 15(12):e0009773. 191

Lindgren, F., Bachl, F., Illian, J., Suen, M. H., Rue, H., and Seaton, A. E. (2024). inlabru: software for fitting latent Gaussian models with non-linear predictors. *arXiv preprint arXiv:2407.00791*. 49, 108, 117, 138, 167, 175, 191, 192, 216, 218, 224

Lindgren, F., Bakka, H., Bolin, D., Krainski, E., and Rue, H. (2020). A diffusion-based spatio-temporal extension of gaussian matérn fields. *arXiv e-prints*, pages arXiv–2006. 94

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with r-inla. *Journal of statistical software*, 63:1–25. 78, 79, 141, 186

Lindgren, F., Rue, H., and Lindström, J. (2011). 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. 51, 53, 54, 61, 98, 110, 137, 141, 152, 176, 185

Liu, Y., Shaddick, G., and Zidek, J. V. (2017). Incorporating high-dimensional exposure modelling into studies of air pollution and health. *Statistics in Biosciences*, 9(2):559–581. 7, 9, 10, 73, 141, 157, 173, 191, 202, 209

Liu, Z., Le, N. D., and Zidek, J. V. (2011). An empirical assessment of bayesian melding for mapping ozone pollution. *Environmetrics*, 22(3):340–353.

Liu, Z. and Rue, H. (2022). Leave-group-out cross-validation for latent gaussian models. *arXiv preprint arXiv:2210.04482*. 96, 132, 133, 136

Liu, Z., Zhang, Q., Li, L., He, J., Guo, J., Wang, Z., Huang, Y., Xi, Z., Yuan, F., Li, Y., et al. (2023). The effect of temperature on dengue virus transmission by aedes mosquitoes. *Frontiers in cellular and infection microbiology*, 13:1242173. 165

Macdonald. G, M. G. (1957). *The epidemiology and control of malaria.* Oxford University Press. 144

Madsen, L., Ruppert, D., and Altman, N. S. (2008). Regression with spatially misaligned data. *Environmetrics: The official journal of the International Environmetrics Society*, 19(5):453–467. 185

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The review of economic studies*, 60(3):531–542. 218

Marigmen, J. L. D. and Addawe, R. C. (2022a). Climatic influences on dengue incidence in baguio city, philippines: A multiple linear regression approach. In *AIP Conference Proceedings*, volume 2472. AIP Publishing. 140

Marigmen, J. L. D. C. and Addawe, R. C. (2022b). Forecasting and on the influence of climatic factors on rising dengue incidence in baguio city, philippines. *Journal of Computational Innovation and Analytics (JCIA)*, 1(01):43–68. 140

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013a). Bayesian computing with inla: new features. *Computational Statistics & Data Analysis*, 67:68–83. 45, 66

Martins, T. G., Simpson, D., Lindgren, F., and Rue, H. (2013b). Bayesian computing with INLA: new features. *Computational Statistics & Data Analysis*, 67:68–83.

McCullagh, P. (2019). *Generalized linear models*. Routledge. 33, 35

McMichael, A. J. (2003). *Climate change and human health: risks and responses*. World Health Organization. 2, 145, 146, 165, 167, 168

McMillan, N. J., Holland, D. M., Morara, M., and Feng, J. (2010). Combining numerical model output and particulate data using bayesian space–time modeling. *Environmetrics: The official journal of the International Environmetrics Society*, 21(1):48–65. 56, 59

McShane, L. M., Albert, P. S., and Palmatier, M. A. (1997). A latent process regression model for spatially correlated count data. *Biometrics*, pages 698–706. 36

MetOffice (2024). Why is humidity important. 145

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6):1087–1092. 37

Minka, T. P. (2013). Expectation propagation for approximate bayesian inference. *arXiv preprint arXiv:1301.2294*. 38

Modrák, M., Moon, A. H., Kim, S., Bürkner, P., Huurre, N., Faltejsková, K., Gelman, A., and Vehtari, A. (2023). Simulation-based calibration checking for bayesian computation: The choice of test quantities shapes sensitivity. *Bayesian Analysis*, 1(1):1–28. 179, 180

Molitor, J., Molitor, N.-T., Jerrett, M., McConnell, R., Gauderman, J., Berhane, K., and Thomas, D. (2006). Bayesian modeling of air pollution health effects with missing exposure data. *American journal of epidemiology*, 164(1):69–76.

Montero, J.-M., Fernández-Avilés, G., and Mateu, J. (2015). *Spatial and spatio-temporal geostatistical modeling and kriging*. John Wiley & Sons. 20

Moraga, P., Cramb, S. M., Mengersen, K. L., and Pagano, M. (2017). A geostatistical model for combined analysis of point-level and area-level data using INLA and SPDE. *Spatial Statistics*, 21:27–41. 13, 57, 102, 135

Murphy, A. K., Salazar, F. V., Bonsato, R., Uy, G., Ebol, A. P., Boholst, R. P., Davis, C., Frentiu, F. D., Bambrick, H., Devine, G. J., et al. (2022). Climate variability and aede s vector indices in the southern philippines: An empirical analysis. *PLOS Neglected Tropical Diseases*, 16(6):e0010478. 140, 168

Murphy, F. A. and Nathanson, N. (1994). The emergence of new virus diseases: an overview. In *Seminars in Virology*, volume 5, pages 87–102. Elsevier. 146

Murray, N. E. A., Quam, M. B., and Wilder-Smith, A. (2013). Epidemiology of dengue: past, present and future prospects. *Clinical epidemiology*, pages 299–309. 2, 3, 198, 202

Myer, M. H., Fizer, C. M., Mcpherson, K. R., Neale, A. C., Pilant, A. N., Rodriguez, A., Whung, P.-Y., and Johnston, J. M. (2020). Mapping aedes aegypti (diptera:

Culicidae) and aedes albopictus vector mosquito distribution in brownsville, tx. *Journal of medical entomology*, 57(1):231–240. 165, 231

Naish, S., Dale, P., Mackenzie, J. S., McBride, J., Mengersen, K., and Tong, S. (2014). Climate change and dengue: a critical and systematic review of quantitative modelling approaches. *BMC infectious diseases*, 14(1):1–14. 3, 98, 165, 198, 202

Nascimento, L. F. C., Vieira, L. C. P. F., Mantovani, K. C. C., and Moreira, D. S. (2016). Air pollution and respiratory diseases: ecological time series. *Sao Paulo Medical Journal*, 134:315–321. 5

Neal, R. M. (2012). Mcmc using hamiltonian dynamics. *arXiv preprint arXiv:1206.1901*. 38

Ong, E. P., Obeles, A. J. T., Ong, B. A. G., and Tantengco, O. A. G. (2022). Perspectives and lessons from the philippines' decades-long battle with dengue. *The Lancet Regional Health–Western Pacific*, 24. 139

Owen, A. B. (2001). *Empirical likelihood*. Chapman and Hall/CRC. 38

PAGASA (2023). Climate of the Philippines. https://www.pagasa.dost.gov.ph/information/climate-philippines. Accessed: 2023-06-12. 3, 95, 96, 97, 126, 127, 128, 129, 130, 250

Patz, J. A., McGeehin, M. A., Bernard, S. M., Ebi, K. L., Epstein, P. R., Grambsch, A., Gubler, D. J., Reither, P., Romieu, I., Rose, J. B., et al. (2000). The potential health impacts of climate variability and change for the united states: executive summary of the report of the health sector of the us national assessment. *Environmental health perspectives*, 108(4):367–376. 146

Pebesma, E. and Bivand, R. (2023). *Spatial data science: With applications in R*. Chapman and Hall/CRC. 20, 21

Peng, R. D. and Bell, M. L. (2010). Spatial misalignment in time series studies of air pollution and health data. *Biostatistics*, 11(4):720–740. 170, 174

Pettit, L. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(1):175–184. 134

Pineda-Cortel, M. R. B., Clemente, B. M., and Nga, P. T. T. (2019). Modeling and predicting dengue fever cases in key regions of the philippines using remote sensing data. *Asian Pacific Journal of Tropical Medicine*, 12(2):60–66. 140

Plummer, M. (2015). Cuts in bayesian graphical models. *Statistics and Computing*, 25:37–43. 9

Prasad, N. N. and Rao, J. N. (1990). The estimation of the mean squared error of small-area estimators. *Journal of the American statistical association*, 85(409):163–171. 27

Promprou, S., Jaroensutasinee, M., and Jaroensutasinee, K. (2005). Climatic factors affecting dengue haemorrhagic fever incidence in southern thailand. *Bulletin of the World Health Organization*. 144, 145

Riebler, A., Sørbye, S. H., Simpson, D., and Rue, H. (2016). An intuitive bayesian spatial model for disease mapping that accounts for scaling. *Statistical methods in medical research*, 25(4):1145–1165. 148, 152, 233

Rigau-Pérez, J. G., Clark, G. G., Gubler, D. J., Reiter, P., Sanders, E. J., and Vorndam, A. V. (1998). Dengue and dengue haemorrhagic fever. *The lancet*, 352(9132):971–977. 146

Roberts, G. O. and Tweedie, R. L. (1996). Geometric convergence and central limit theorems for multidimensional hastings and metropolis algorithms. *Biometrika*, 83(1):95–110. 38

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC press. 41, 69

Rue, H. and Martino, S. (2007). Approximate bayesian inference for hierarchical gaussian markov random field models. *Journal of statistical planning and inference*, 137(10):3177–3192. 43, 45, 165

Rue, H., Martino, S., and Chopin, N. (2009). Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392. 38, 39, 43, 44, 61, 99, 108, 137, 141, 152, 174

Ruiz-Cárdenas, R., Krainski, E. T., and Rue, H. (2012). Direct fitting of dynamic models using integrated nested laplace approximations - inla. *Computational Statistics & Data Analysis*, 56(6):1808–1828. 66

Ruppert, D., Wand, M. P., and Carroll, R. J. (2003). *Semiparametric regression*. Cambridge university press. 28

Rustand, D., Van Niekerk, J., Krainski, E. T., Rue, H., and Proust-Lima, C. (2024). Fast and flexible inference for joint models of multivariate longitudinal and survival data using integrated nested laplace approximations. *Biostatistics*, 25(2):429–448. 7

Ryan, S. J., Carlson, C. J., Mordecai, E. A., and Johnson, L. R. (2019). Global expansion and redistribution of aedes-borne virus transmission risk with climate change. *PLoS neglected tropical diseases*, 13(3):e0007213. 144

Sahu, S. K., Gelfand, A. E., and Holland, D. M. (2010). Fusing point and areal level space–time data with application to wet deposition. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(1):77–103. 56, 58

Säilynoja, T., Bürkner, P.-C., and Vehtari, A. (2022). Graphical test for discrete uniformity and its applications in goodness-of-fit evaluation and multiple sample comparison. *Statistics and Computing*, 32(2):32. 180

Santos, A. P. (2019). Philippines: Worst dengue outbreak in years kills over a thousand. https://www.aljazeera.com/news/2019/9/17/philippines-worst-dengue-outbreak-in-years-kills-over-a-thousand. Accessed: 2025-17-02. 162

Schabenberger, O. and Gotway, C. A. (2017). *Statistical methods for spatial data analysis*. Chapman and Hall/CRC. 20, 22, 23, 24, 34

Schmidt, A. M. and Gelfand, A. E. (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108(D24). 58

Schoenbach, V. J. and Rosamond, W. D. (2000). Understanding the fundamentals of epidemiology: an evolving text. *Chapel Hill: University of North Carolina*, pages 129–151. 199

Schrödle, B. and Held, L. (2011). Spatio-temporal disease mapping using inla. *Environmetrics*, 22(6):725–734. 141

Seposo, X., Valenzuela, S., and Apostol, G. L. (2023). Socio-economic factors and its influence on the association between temperature and dengue incidence in 61 provinces of the philippines, 2010–2019. *PLOS Neglected Tropical Diseases*, 17(10):e0011700. 140

Seposo, X., Valenzuela, S., Apostol, G. L. C., Wangkay, K. A., Lao, P. E., and Enriquez, A. B. (2024). Projecting temperature-related dengue burden in the philippines under various socioeconomic pathway scenarios. *Frontiers in Public Health*, 12:1420457. 145

Seposo, X. T. (2021). Dengue at the time of covid-19 in the philippines. *Western Pacific Surveillance and Response Journal: WPSAR*, 12(2):38. 3, 142, 143

Serafini, F., Lindgren, F., and Naylor, M. (2023). Approximation of Bayesian hawkes process with inlabru. *Environmetrics*, 34(5):e2798. 138

Shaddick, G. and Wakefield, J. (2002). Modelling daily multivariate pollutant data at multiple sites. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 51(3):351–372. 9, 204, 209

Shaddick, G., Zidek, J. V., and Schmidt, A. M. (2023). *Spatio–Temporal Methods in Environmental Epidemiology with R*. Chapman and Hall/CRC. 70

Simkovich, S. M., Goodman, D., Roa, C., Crocker, M. E., Gianella, G. E., Kirenga, B. J., Wise, R. A., and Checkley, W. (2019). The health and social implications

of household air pollution and respiratory diseases. *NPJ primary care respiratory medicine*, 29(1):12. 5

Simpson, D., Rue, H., Riebler, A., Martins, T. G., and Sørbye, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*. 78, 79, 103, 111, 122, 186

Sørbye, S. H. and Rue, H. (2014). Scaling intrinsic gaussian markov random field priors in spatial modelling. *Spatial Statistics*, 8:39–51. 149

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). Winbugs user manual. 9

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, 64(4):583–639.

Stein, M. L. (2012). *Interpolation of spatial data: some theory for kriging*. Springer Science & Business Media. 27

Stoddard, S. T., Forshey, B. M., Morrison, A. C., Paz-Soldan, V. A., Vazquez-Prokopec, G. M., Astete, H., Reiner Jr, R. C., Vilcarromero, S., Elder, J. P., Halsey, E. S., et al. (2013). House-to-house human movement drives dengue virus transmission. *Proceedings of the National Academy of Sciences*, 110(3):994–999. 3, 143

Stolerman, L. M., Maia, P. D., and Kutz, J. N. (2019). Forecasting dengue fever in brazil: An assessment of climate conditions. *PloS one*, 14(8):e0220106.

Su, G. L. S. (2008). Correlation of climatic factors and dengue incidence in metro manila, philippines. *AMBIO: A Journal of the Human Environment*, 37(4):292–294. 140

Subido, M. E. and Aniversario, I. S. (2022). A correlation study between dengue incidence and climatological factors in the philippines. *Asian Res J Math*, 18:110–119. 140, 145

Sumi, A., Telan, E., Chagan-Yasutan, H., Piolo, M., Hattori, T., and Kobayashi, N. (2017). Effect of temperature, relative humidity and rainfall on dengue fever and leptospirosis infections in manila, the philippines. *Epidemiology & Infection*, 145(1):78–86. 140

Szpiro, A. A., Sheppard, L., and Lumley, T. (2011). Efficient measurement error correction with spatially misaligned data. *Biostatistics*, 12(4):610–623. 7, 141

Talts, S., Betancourt, M., Simpson, D., Vehtari, A., and Gelman, A. (2018). Validating bayesian inference algorithms with simulation-based calibration. *arXiv preprint arXiv:1804.06788*. 169, 170, 178, 179, 204, 205, 207, 210, 214, 215

Thu, H. M., Aye, K. M., and Thein, S. (1998). The effect of temperature and humidity on dengue virus propagation in aedes aegypti mosquitos. *Southeast Asian J Trop Med Public Health*, 29(2):280–284. 198, 202

Tobler, W. R. (1979). Smooth pycnophylactic interpolation for geographical regions. *Journal of the American Statistical Association*, 74(367):519–530.

Tran, H. M., Tsai, F.-J., Lee, Y.-L., Chang, J.-H., Chang, L.-T., Chang, T.-Y., Chung, K. F., Kuo, H.-P., Lee, K.-Y., Chuang, K.-J., et al. (2023). The impact of air pollution on respiratory diseases in an era of climate change: A review of the current evidence. *Science of the Total Environment*, 898:166340. 5

Undurraga, E. A., Edillo, F. E., Erasmo, J. N. V., Alera, M. T. P., Yoon, I.-K., Largo, F. M., and Shepard, D. S. (2017). Disease burden of dengue in the philippines: adjusting for underreporting by comparing active and passive dengue surveillance in punta princesa, cebu city. *The American journal of tropical medicine and hygiene*, 96(4):887. 139

Undurraga, E. A., Halasa, Y. A., and Shepard, D. S. (2013). Use of expansion factors to estimate the burden of dengue in southeast asia: a systematic analysis. *PLoS neglected tropical diseases*, 7(2):e2056. 139

Van Niekerk, J., Krainski, E., Rustand, D., and Rue, H. (2023). A new avenue for

bayesian inference with inla. *Computational Statistics & Data Analysis*, 181:107692. 47, 48, 49, 165, 174, 175

Van Strien, A. J., Van Swaay, C. A., and Termaat, T. (2013). Opportunistic citizen science data of animal species produce reliable estimates of distribution trends if analysed with occupancy models. *Journal of Applied Ecology*, 50(6):1450–1458. 137

Villejo, S. J., Illian, J. B., and Swallow, B. (2023). Data fusion in a two-stage spatio-temporal model using the inla-spde approach. *Spatial Statistics*, 54:100744. 135, 141, 191

Villejo, S. J., Martino, S., Lindgren, F., and Illian, J. (2024). A data fusion model for meteorological data using the inla-spde method. *arXiv preprint arXiv:2404.08533*.

Wackernagel, H. (2003). *Multivariate geostatistics: an introduction with applications*. Springer Science & Business Media. 27

Wakefield, J. and Shaddick, G. (2006). Health-exposure modeling and the ecological fallacy. *Biostatistics*, 7(3):438–455. 9, 204, 209

Waller, L. A. and Carlin, B. P. (2010). Disease mapping. *Chapman & Hall/CRC handbooks of modern statistical methods*, 2010:217. 70, 143, 147, 199

Waller, L. A. and Gotway, C. A. (2004). *Applied spatial statistics for public health data*. John Wiley & Sons. 27, 70, 143

Wang, B. and Titterington, D. M. (2005). Inadequacy of interval estimates corresponding to variational bayesian approximations. In *International workshop on artificial intelligence and statistics*, pages 373–380. PMLR. 38

Wedderburn, R. W. (1974). Quasi-likelihood functions, generalized linear models, and the gauss—newton method. *Biometrika*, 61(3):439–447. 35

Whittle, P. (1954). On stationary processes in the plane. *Biometrika*, pages 434–449. 31, 51

WHO (2023a). Dengue - global situation. https://www.who.int/emergencies/disease-outbreak-news/item/2023-DON498. Accessed: 2025-20-02. 3, 4, 143

WHO (2023b). Neglected tropical diseases. https://www.who.int/health-topics/neglected-tropical-diseases. Accessed: 2025-03-01. 2

Wikle, C. K. and Berliner, L. M. (2005). Combining information across spatial scales. *Technometrics*, 47(1):80–91. 56

Wilkinson, R. D. (2013). Approximate bayesian computation (abc) gives exact results under the assumption of model error. *Statistical applications in genetics and molecular biology*, 12(2):129–141. 38

Wolfinger, R. and O'connell, M. (1993). Generalized linear mixed models a pseudo-likelihood approach. *Journal of statistical Computation and Simulation*, 48(3-4):233–243. 36

Xu, L., Stige, L. C., Chan, K.-S., Zhou, J., Yang, J., Sang, S., Wang, M., Yang, Z., Yan, Z., Jiang, T., et al. (2017). Climate variation drives dengue dynamics. *Proceedings of the National Academy of Sciences*, 114(1):113–118. 3

Xu, Z., Bambrick, H., Yakob, L., Devine, G., Frentiu, F. D., Salazar, F. V., Bonsato, R., and Hu, W. (2020). High relative humidity might trigger the occurrence of the second seasonal peak of dengue in the philippines. *Science of The Total Environment*, 708:134849. 140, 145, 155

Ye, W., Lin, X., and Taylor, J. M. (2008). Semiparametric modeling of longitudinal measurements and time-to-event data–a two-stage regression calibration approach. *Biometrics*, 64(4):1238–1246. 7

Yeung, J. and Faidell, S. (2019). Philippines declares a national dengue epidemic after 622 deaths. https://edition.cnn.com/2019/08/07/health/philippines-dengue-epidemic-intl-hnk/index.html. Accessed: 2025-17-02. 162, 164

Yucel, R. M. and Zaslavsky, A. M. (2005). Imputation of binary treatment variables with measurement error in administrative data. *Journal of the American Statistical Association*, 100(472):1123–1132. 9

Zhong, R. and Moraga, P. (2023). Bayesian Hierarchical Models for the Combination of Spatially Misaligned Data: A Comparison of Melding and Downscaler Approaches using INLA and SPDE. *Journal of Agricultural, Biological and Environmental Statistics*, pages 1–20. 13, 57, 102, 135

Zhu, L., Carlin, B. P., and Gelfand, A. E. (2003). Hierarchical regression with misaligned spatial data: relating ambient ozone and pediatric asthma er visits in atlanta. *Environmetrics: The official journal of the International Environmetrics Society*, 14(5):537–557. 141, 173, 191