Master's thesis

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

Jim Oskar Zachrisson Totland

Geostatistical Modeling under Positional Uncertainty

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

NTNU
Norwegian University of
Science and Technology

Jim Oskar Zachrisson Totland

Geostatistical Modeling under Positional Uncertainty

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

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

NTNU
Norwegian University of
Science and Technology

Preface

This thesis is a product of the course TMA4900 - Industrial Mathematics, Master's Thesis and concludes my M.Sc. degree in Applied Physics and Mathematics at NTNU. My path through university has been anything but linear. I have swung by Industrial Design and Nanotechnology before I ended up specializing in Industrial Mathematics. From then on, my focus has been on the field of statistics, which I have come to deeply appreciate for its practical, mathematical and philosophical dimensions.

I am sincerely grateful to my supervisor, Geir-Arne Fuglstad, whose consistent guidance has been instrumental throughout the entire thesis writing process. I extend my gratitude to friends and fellow students, who have made the process enjoyable despite being hard work. Finally, I would like to acknowledge my partner, Silje, and my family for always being patient and supportive.

Jim Oskar Zachrisson Totland,
Trondheim, June 2023

Abstract

In this thesis, we study positional uncertainty in the setting of geostatistical modeling. A central example of positional uncertainty is the application of geographic masking (geomasking) intended to protect the privacy of data subjects. Geomasking is applied in Demographic and Health Surveys (DHS) (Burgert et al., 2013), which collect data on population, health and nutrition in many low and middle income countries. These data are important for monitoring and facilitating the progress towards the Sustainable Development Goals, which motivates the study of how positional uncertainty from geomasking impacts geostatistical inference.

A contribution of this thesis is to summarize the geostatistics-literature on positional uncertainty. Additionally, we focus on two geomasking strategies, referred to as *disk jittering* and *donut jittering*. To account for the uncertainty these *jittering schemes* introduce, we use the computationally efficient *numerical approach* proposed by Altay et al. (2022a). We extend the numerical approach to handle donut jittering. Furthermore, we investigate the accuracy of the method by constructing, implementing and applying a *sample-based* approach, which demonstrates that the numerical approach may effectively approximate exact inference under jittering. Subsequently, we conduct a simulation study to assess the impact of *moderate* and *substantial* jittering on inference with the numerical approach. The results agree with previous work and expectations from theory, where adjusting for jittering leads to less attenuation in estimated covariate coefficients, substantially less biased estimates of the *nugget* variance and superior predictive measures. For example, when there are 10^3 observations and spatial covariates account for 68% of the variation in observations, we find that adjusting for, relative to ignoring, substantial donut jittering gives a median prediction root-mean-square error that is 18% lower. Another contribution of this work is to explore how privacy protection can be evaluated for a given jittering scheme.

We apply the numerical approach to real data in a case study of vaccination coverage in Nigeria, demonstrating its applicability to arbitrary geographies and generalized linear geostatistical models. Due to its favorable qualities and results from this and previous work, we advocate the use of the numerical approach to account for positional uncertainty in geostatistical analyses of DHS data and in other applications with positional uncertainty of comparable scale.

Contents

Chapter 1

Introduction

Analyses of spatial data almost always assume that the recorded spatial locations are exact. However, there are many situations where this assumption is invalid because of uncertainty in the spatial locations. Such *positional uncertainty* could come from multiple sources, e.g., positional error in maps (Barber et al., 2006), rounding of coordinates to the nearest grid point for convenience (Cressie and Kornak, 2003), or intentional masking of locations meant to preserve the privacy of data subjects (Armstrong et al., 1999; Zandbergen, 2014). In this thesis, we mainly focus on the latter example of positional uncertainty, which is referred to as *geomasking*. The idea underpinning geomasking is that it can be used to strike a balance between making the spatial locations in datasets broadly available, enabling, e.g., more research on geographic patterns in the data, while at the same time maintaining the privacy of data subjects for both ethical and legal reasons (European Parliament, Council of the European Union, 2016).

The most prominent example of geomasking is perhaps provided by the Demographic and Health Surveys (DHS) Program, which is a global initiative that has collected data on health, population and nutrition through more than 400 surveys in over 90 participating countries, where surveys are typically conducted every 5 years (The DHS Program, n.d.). In order to preserve the confidentiality of respondents in DHS datasets, the coordinates of survey clusters are randomly displaced Burgert et al. (2013). Moreover, the surveys are mainly conducted in low- and middle-income countries, and are a valuable source of data for both policymakers and researchers trying to facilitate and monitor the progress towards the Sustainable Development Goals (SDGs) set forth by the United Nations (UN) (UN, 2015). Consequently, accurate statistical inference is of great importance when considering this data, which motivates the study of how positional uncertainty from geomasking impacts inference. Furthermore, studies have demonstrated that positional uncertainty may lead to increased bias in parameter estimates and worse predictive measures when it is ignored (Gabrosek and Cressie, 2002; Cressie and Kornak, 2003; Altay et al., 2022a,b; Totland, 2022).

Some approaches to account for positional uncertainty have been presented in the literature of spatial statistics. Cressie and Kornak (2003) adjust for positional error in a Gaussian observation process by adjusting the mean and covariance function, and assume that the resulting adjusted process is Gaussian. This assumption is, however, not true in general and the approach does not easily generalize to generalized linear geostatistical models. Fanshawe and Diggle (2011) present a

method which accounts for error in prediction and observation locations, where maximum likelihood estimation is used on a hierarchical geostatistical model. However, they encounter prohibitive computation times with positional uncertainty in the observation locations. Later work proposes a remedy for this through a composite likelihood approach for linear geostatistical models with a Gaussian likelihood (Fronterrè et al., 2018). Wilson and Wakefield (2021) consider the positional uncertainty in DHS data and develop a computational procedure which uses integrated nested Laplace approximations (INLA) (Rue et al., 2009) within a Markov chain Monte Carlo (MCMC) algorithm (Gómez-Rubio and Rue, 2018), which for each simulation scenario had a computation time of around 52 hours. By contrast, the novel approach proposed by Altay et al. (2022a) enables fast inference measured in minutes in the presence of positional uncertainty by using the stochastic partial differential equation (SPDE) approach (Lindgren et al., 2011) to approximate the spatial field and the Template Model Builder (TMB) (Kristensen et al., 2016) R-package for fast computations. Furthermore, the approach easily extends to generalized linear geostatistical models, and has recently been made available as an R-package (Altay et al., 2023), where Gaussian, Binomial and Poisson observation likelihoods are supported. This model estimation approach uses numerical integration to approximate the mixture distribution which arises with positional uncertainty, and we therefore refer to it as the *numerical approach*.

Because of the favorable qualities of the numerical approach, a key focus of this thesis is to further study and develop the method. The numerical approach has been applied and studied in the context of DHS data and the geomasking strategy employed there. Altay et al. (2022a) find that using it to account for the geomasking in a smoothly varying spatial field leads to equally or better performance in terms of parameter bias and predictive measures compared to the common approach of ignoring positional uncertainty. Altay et al. (2022b) introduce raster-based covariates into the geostatistical model and find that accounting for geomasking reduces attenuation in the estimated covariate coefficients and improves predictive measures. Totland (2022) also includes spatial covariates in the numerical approach and considers a Gaussian linear geostatistical model, for which accounting for positional uncertainty as opposed to ignoring it is found to give substantially less or equally biased parameter estimates when there are sufficiently many observations and the covariate signal is moderately strong.

In this thesis, we aim to study positional uncertainty from a wider perspective than the application of geomasking in DHS data, and therefore give a detailed account on relevant previous work on positional uncertainty. We focus on two types of geomasking strategies, which we collectively refer to as *jittering schemes*. Specifically, we consider *disk jittering* and *donut jittering*, whose names refer to the shape of the positional uncertainty distribution. The numerical approach is previously implemented for disk jittering, so we extend it to handle donut jittering. Additionally, we develop a *sample-based* approach for inference under positional uncertainty, and compare it to the numerical approach, both as an alternative and as a verification tool. To study the impact of positional uncertainty in geostatistical modeling, we apply jittering schemes of varying scale to simulated data in a simulation study and consider how adjusting for it compares to ignored it with the numerical approach. We also explore how the privacy protection of a jittering scheme can be evaluated, to hopefully give data managers a better way to justify

the choice of a specific jittering scheme. Finally, we demonstrate an application of the numerical approach to real data, in a study of vaccination coverage in Nigeria.

To evaluate and compare models, we consider both parameter estimation and predictive power. Parameter estimation is evaluated through bias, and predictive power is evaluated with root-mean-square error (RMSE) and continuous ranked probability score (CRPS) (Gneiting and Raftery, 2007). We do not consider cross-validation, because the datasets we use for estimation contain observations with positional uncertainty, which are not the target of prediction. Hence, we in stead simulate datasets and use the underlying latent process as the target of prediction.

The thesis is structured as follows. Chapter 2 summarizes the traditional geostatistical modeling approach without positional uncertainty, in addition to the tools employed to achieve computationally efficient inference. Then, Chapter 3 gives an overview of the geostatistics-literature on positional uncertainty and describes the jittering schemes we focus on. The model which incorporates jittering and our two methods of model estimation and inference are presented in Chapter 4. In Chapter 5, the sample-based approach is used to investigate the accuracy of the numerical approach. Chapter 6 contains a simulation study to compare how different jittering schemes impact inference. The numerical approach is then applied to DHS data in a case study of prevalences in Chapter 7. Finally, Chapter 8 ends the thesis with a discussion of the presented material in relation to the objectives set out here, weaknesses and future work.

Chapter 2

Background

We present the traditional approach to geostatistical modeling in Section 2.1. Then, Section 2.2 presents some important MCMC methods that can be used for sample-based inference. Section 2.3 describes some tools that can be used to achieve computationally efficient inference. Lastly, Section 2.4 presents the scoring rules we use to evaluate our estimated models. We remark that Sections 2.1, 2.3 and 2.4 are largely based on the work of Totland (2022).

2.1 Geostatistical modeling

Geostatistics is a sub-discipline of spatial statistics, where a finite sample of georeferenced data is used to study some related, underlying spatially continuous phenomenon (Diggle and Ribeiro, 2007). We can model this phenomenon as a spatial process, or field, which is a family of random variables $u = \{u(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ where $\mathcal{D}$ is the domain under consideration and $\boldsymbol{s}$ is a location in that domain. In the context of geostatistical modeling, $\mathcal{D} \subset \mathbb{R}^d$, where $d \in \mathbb{N}^+$, and u takes on values in a real-valued state space. In terms of this general categorisation, geostatistical modeling has seen applications in a diversity of fields, such as ecology, hydrology, meteorology and environmental sciences (Kyriakidis and Journel, 1999). The first step in geostatistical modeling is typically to model the distribution of the spatial field. Then, Gaussian random fields are often particularly appealing.

2.1.1 Gaussian Random Fields

Gaussian random fields (GRFs) play a central role in geostatistical modeling by providing a powerful yet simple tool for modeling spatial dependence. We define GRFs in terms of their finite dimensional distribution.

Definition 1 (Gaussian Random Field)**.** For a domain $\mathcal{D} \subset \mathbb{R}^d$, the random field $u = \{u(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ is a Gaussian random field if $\forall m \geq 1$ and $\forall \boldsymbol{s}_1, \dots \boldsymbol{s}_m \in \mathcal{D}$, the vector $(u(\boldsymbol{s}_1), \dots, u(\boldsymbol{s}_m))^{\mathrm{T}}$ has a multivariate Gaussian distribution. ‖

GRFs have the appealing property that they are uniquely determined by a mean function $m : \mathcal{D} \to \mathbb{R}$ and a positive semi-definite covariance function $c : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$. A *centered* GRF has $m(\boldsymbol{s}) = 0 \; \forall \boldsymbol{s} \in \mathcal{D}$. We will exclusively consider centered GRFs with a Matérn covariance function, which for $\boldsymbol{s}_i, \boldsymbol{s}_j \in \mathcal{D}$ is given by

$$c(\boldsymbol{s}_i, \boldsymbol{s}_j) = \sigma_{\mathrm{S}}^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{8\nu}\|\boldsymbol{s}_i - \boldsymbol{s}_j\|}{\rho_{\mathrm{S}}} \right)^{\nu} K_\nu \left(\frac{\sqrt{8\nu}\|\boldsymbol{s}_i - \boldsymbol{s}_j\|}{\rho_{\mathrm{S}}} \right), \tag{2.1}$$

where σ_{S}^2 is the marginal variance, ρ_{S} is the range, ν is the smoothness parameter, Γ is the gamma function and K_ν is the modified Bessel function of second kind of order ν. This parametrization follows Lindgren et al. (2011) and makes ρ_{S} the distance at which the correlation is approximately 0.1. We note that the Matérn covariance function only depends on the distance $h = \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$ between locations, so we can equivalently let $c(\boldsymbol{s}_i, \boldsymbol{s}_j) = C(h)$. This property characterizes an *isotropic* covariance function.

2.1.2 Bayesian Hierarchical Modeling

In this thesis, we adhere to the theory of Bayesian statistics. This means that we model both the observed data and any unknown model parameters as random variables. Now, to construct a Bayesian model, we let $\boldsymbol{\theta}$ be the vector of model parameters, $\boldsymbol{\eta} \in \mathbb{R}^n$ be a discretized unobservable latent process, and $\boldsymbol{y} \in \mathbb{R}^n$ be the vector of observations which are connected to the discretized latent process, where $n \in \mathbb{N}^+$. We model the distributional relationships between these quantities hierarchically. Specifically, we model

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta} &\sim \pi(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta}), \\ \boldsymbol{\eta}|\boldsymbol{\theta} &\sim \pi(\boldsymbol{\eta}|\boldsymbol{\theta}), \\ \boldsymbol{\theta} &\sim \pi(\boldsymbol{\theta}), \end{aligned} \tag{2.2}$$

where each line in 2.2 represents a layer in the hierarchical model. The first line represents the observation layer, the second line represents the underlying latent layer, and the third line represents the parameter prior layer, which encodes any *a priori* beliefs about the model. This hierarchical structure characterizes *Bayesian hierarchical models* (Cressie and Wikle, 2011, Chapter 2.1), and serves to make the model easier to understand and modify. Inference with the Bayesian hierarchical model is based on the *posterior*, $\pi(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y})$, which is the *a posteriori* distribution of the unobserved quantities in the model, conditional on what we have observed. Importantly, *Bayes' rule* (Bayes, 1763) gives that

$$\pi(\boldsymbol{\eta}, \boldsymbol{\theta}|\boldsymbol{y}) = \frac{\pi(\boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{\theta})}{\pi(\boldsymbol{y})} = \frac{\pi(\boldsymbol{y}|\boldsymbol{\eta}, \boldsymbol{\theta})\pi(\boldsymbol{\eta}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\boldsymbol{y})}, \tag{2.3}$$

enabling us to express the posterior (up to a normalizing constant) in terms of the distributions modeled in (2.2).

We follow the hierarchical approach to construct the *standard* linear geostatistical model. First, let $y_1, \ldots y_n$ denote the observations at the corresponding locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ in the domain $\mathcal{D} \subset \mathbb{R}^2$. We model our observations to be composed of a stochastic latent process, $\eta = \{\eta(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$, and a noise process, $\varepsilon = \{\varepsilon(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$, such that

$$y_i = \eta(\boldsymbol{s}_i) + \varepsilon(\boldsymbol{s}_i), \quad i = 1, \ldots, n, \tag{2.4}$$

where we let $\varepsilon(\boldsymbol{s}_i) = \varepsilon_i$ be the measurement error in observation i (also referred to as the *nugget effect* or simply the *nugget*) which is considered to be independent and identically distributed (iid) to the measurement error in other observations,

$$\varepsilon_1, \dots \varepsilon_n \mid \sigma^2_{\mathrm{N}} \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2_{\mathrm{N}}).$$

The latent process is is composed of spatial covariates and a GRF,

$$\eta(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + u(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D}, \tag{2.5}$$

where $\boldsymbol{x}(\boldsymbol{s}) = (1, x_1(\boldsymbol{s}), \dots, x_{p-1}(\boldsymbol{s}))^{\mathrm{T}}$ is a p-dimensional vector of deterministic covariate values, $\boldsymbol{\beta} = (\mu, \beta_1, \dots, \beta_{p-1})^{\mathrm{T}}$ is a p-dimensional vector of covariate coefficients, and u is a centered GRF with a Matérn covariance function with marginal marginal variance σ^2_{S}, range ρ_{S} and fixed smoothness $\nu = 1$.

It remains to specify a prior on $\boldsymbol{\beta}$ and a prior $\pi(\boldsymbol{\theta})$ on $\boldsymbol{\theta} = (\sigma^2_{\mathrm{N}}, \sigma^2_{\mathrm{S}}, \rho_{\mathrm{S}})^{\mathrm{T}}$. We let $\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, V\mathbf{I}_p)$, where V is a fixed positive number which can be set to give a sufficiently vague or informative prior. For the parameters associated with the GRF and the nugget variance, we use penalized complexity priors, which are described in Section 2.1.3. Finally, let $\boldsymbol{y} = (y_1, \dots, y_n)^{\mathrm{T}}$ and $\boldsymbol{\eta} = (\eta(\boldsymbol{s}_1), \dots, \eta(\boldsymbol{s}_n))^{\mathrm{T}}$. The standard linear geostatistical model can then be presented as a Bayesian hierarchical model,

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{\eta}, \sigma^2_{\mathrm{N}} &\sim \mathcal{N}_n(\boldsymbol{\eta}, \sigma^2_{\mathrm{N}}\mathbf{I}_n), \\ \boldsymbol{\eta}|\boldsymbol{\beta}, \sigma^2_{\mathrm{S}}, \rho_{\mathrm{S}} &\sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathrm{S}}(\sigma^2_{\mathrm{S}}, \rho_{\mathrm{S}})), \\ \boldsymbol{\beta} &\sim \mathcal{N}_p(\mathbf{0}, V\mathbf{I}_p), \ \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta}), \end{aligned} \tag{2.6}$$

where $\mathbf{X}$ is an $n \times p$ design matrix with i'th row given by $\boldsymbol{x}(\boldsymbol{s}_i)$ and $\boldsymbol{\Sigma}_{\mathrm{S}}(\sigma^2_{\mathrm{S}}, \rho_{\mathrm{S}})$ is the $n \times n$ covariance matrix arising from the covariance function of the GRF, with element (i, j) equal to $c(\boldsymbol{s}_i, \boldsymbol{s}_j)$. We note that this model can easily extended to generalized linear geostatistical models by introducing a link function, g, and letting $\mathrm{E}[\boldsymbol{y}|\boldsymbol{\eta}] = g(\boldsymbol{\eta})$ where the distribution of $\boldsymbol{y}|\boldsymbol{\eta}$ belongs to an exponential family. Furthermore, we assume that the observations are conditionally independent given the discretized latent process and the parameter(s) of the observation layer. That is,

$$\pi(\boldsymbol{y}|\boldsymbol{\eta}, \sigma^2_{\mathrm{N}}) = \prod_{i=1}^{n} \pi(y_i|\eta(\boldsymbol{s}_i), \sigma^2_{\mathrm{N}}).$$

2.1.3 Penalised Complexity Priors

To complete the Bayesian hierarchical model presented in Section 2.1.2, we need to assign priors to the parameters, $\boldsymbol{\theta} = (\rho_{\mathrm{S}}, \sigma^2_{\mathrm{S}}, \sigma^2_{\mathrm{N}})^{\mathrm{T}}$. The choice of priors is often dictated by convenience, e.g., by using *conjugate* priors for computational convenience or priors that have previously been used on similar problems in the literature. Simpson et al. (2017) argue and demonstrate that such "ad hoc" approaches to prior selection can have an unfortunate impact on inference, and, in an attempt to alleviate the difficulty of prior selection, they propose a principled approach to constructing *penalized complexity* (PC) priors.

In the PC prior framework, model components are considered to be extensions of less flexible base models. For example, the measurement error component in the hierarchical model,

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^{\mathrm{T}} \mid \sigma_{\mathrm{N}}^2 \sim \mathcal{N}_n(\mathbf{0}, \sigma_{\mathrm{N}}^2 \mathbf{I}_n),$$

consists of random effects, and a natural base model is the absence of these effects, i.e., $\boldsymbol{\varepsilon}|\sigma_{\mathrm{N}}^2 = 0$. Furthermore, PC priors are constructed from a set of principles. The principles state that the concept of *Occam's razor* should be invoked, namely that a less complex model should be preferred until there is sufficient support for a more complex model. The measure of complexity is chosen to be the Kullback-Leibler divergence (KLD) from the base model. If the probability density function (pdf) of the base model is given by g and the pdf of the model component is given by f, the KLD of f from g is given by (Kullback and Leibler, 1951)

$$D_{\mathrm{KL}}(f \parallel g) = \int_{\mathbb{R}^m} f(\boldsymbol{z}) \log\left(\frac{f(\boldsymbol{z})}{g(\boldsymbol{z})}\right) \mathrm{d}\boldsymbol{z}, \tag{2.7}$$

where $\boldsymbol{z} \in \mathbb{R}^m$, $m \in \mathbb{N}^+$. A measure of distance is constructed from (2.7), $d(f\|g) = \sqrt{2D_{\mathrm{KL}}(f \parallel g)}$. Then, a prior is placed on the distance, d, such that the rate of penalization of deviation from the base model is constant. This amounts to placing an exponential prior on d,

$$d \sim \mathrm{Exp}\left(\tilde{\lambda}\right).$$

For the measurement error component, $d = d(\sigma_{\mathrm{N}}^2)$, so a prior on, e.g., the nugget variance, $\pi(\sigma_{\mathrm{N}}^2)$, can be derived by an ordinary change of variables transformation.

Finally, to completely specify a PC prior, the principle of user-defined scaling must be applied. That is, the practitioner should be able to control how the prior mass is distributed through an interpretable scaling parameter, yielding a weakly informative prior. In terms of the measurement error component one can construct a PC prior which allows the user to specify (U, α) such that $\mathrm{P}(\sigma_{\mathrm{N}} > U) = \alpha$. With this choice, the prior of $\xi = 1/\sigma_{\mathrm{N}}^2$ can be derived as

$$\pi(\xi) = \frac{\lambda}{2} \xi^{-3/2} \exp(-\lambda \xi^{-1/2}), \quad \xi > 0, \lambda < 0,$$

where $\lambda = -\log(\alpha)/U$. More details on this derivation can be found in Appendix A.1 of Simpson et al. (2017).

The spatial field in the hierarchical model is another model component which needs parameter priors. Fuglstad et al. (2019) develop a joint PC prior for the range and marginal variance of a GRF with Matérn covariance function with fixed smoothness parameter. They select a base model with infinite range and zero marginal variance, and develop a PC prior where the user specifies $(\rho_0, \alpha_1, \sigma_0, \alpha_2)$ such that

$$\mathrm{P}(\rho_{\mathrm{S}} < \rho_0) = \alpha_1 \quad \text{and} \quad \mathrm{P}(\sigma_{\mathrm{S}} > \sigma_0) = \alpha_2.$$

In the following, we will place these prior on the median, i.e., we let $\alpha_1 = \alpha_2 = 0.5$, such that

$$\mathrm{P}(\rho_{\mathrm{S}} < \rho_0) = 0.5 \quad \text{and} \quad \mathrm{P}(\sigma_{\mathrm{S}} > \sigma_0) = 0.5,$$

because we want to facilitate interpretability through the general familiarity with the median.

2.2 MCMC Methods

When priors are defined and data is collected, posteriors are completely determined. However, geostatistical models such as the one presented in (2.6) are typically analytically intractable, making approximate inference necessary. If that is the case, one way to proceed is with sample-based inference by using MCMC methods. MCMC methods comprise a collection of methods in which Markov chains are used to sample from a target probability distribution. Without loss of generality, assume that we want to sample from the marginal distribution of $\boldsymbol{z} = (z_1, \ldots, z_m)^{\mathrm{T}}, m \in \mathbb{N}^+$, i.e., the target distribution is $\pi(\boldsymbol{z})$. Then, in the MCMC framework, one constructs an *ergodic* Markov Chain that has $\pi(\boldsymbol{z})$ as its stationary distribution. The theory of Markov chains and MCMC methods is vast, and the reader is referred to, e.g., Robert and Casella (2004) for a comprehensive overview. Here, we shortly present the most important building blocks for constructing MCMC algorithms.

2.2.1 Gibbs Sampling

MCMC methods are often used because the dimension of $\boldsymbol{z}$ is large, making it difficult to sample with, e.g., rejection sampling. In MCMC-algorithms, this problem is solved by dividing the sampling into smaller parts. The Gibbs sampler is the foremost example of this, where the full conditional distribution of each element of $\boldsymbol{z}$ is sampled in each iteration of the chain. Specifically, in iteration b, we loop through the elements of $\boldsymbol{z}$ and sample from

$$z_i^{(b)} | z_1^{(b)}, z_2^{(b)}, \ldots, z_{i-1}^{(b)}, z_{i+1}^{(b-1)}, \ldots, z_p^{(b-1)}, \quad i = 1, \ldots m,$$

where $z_i^{(b)}$ is element i in $\boldsymbol{z}$ in iteration b of the chain. This of course presupposes that the full conditionals are available and easy to sample from. It can then be shown, under mild regularity conditions, that for large values of b,

$$(z_1^{(b)}, z_2^{(b)}, \ldots, z_n^{(b)}),\ (z_1^{(b+1)}, z_2^{(b+1)}, \ldots, z_p^{(b+1)}),\ \text{etc.},$$

can be regarded as (correlated) samples drawn from $\pi(\boldsymbol{z})$ (Roberts and Smith, 1994).

2.2.2 Metropolis-Hastings Algorithm

The Gibbs sampler can in fact be framed as a special case of a more general algorithm, namely the Metropolis-Hastings (MH) algorithm, which was first introduced by Metropolis et al. (2004) in a statistical physics context and later generalized by Hastings (1970). The MH algorithm relies on a proposal distribution $q(\boldsymbol{z}'|\boldsymbol{z})$ for proposing new samples of $\boldsymbol{z}$, and an acceptance probability,

$$\alpha(\boldsymbol{z}', \boldsymbol{z}) = \min \left\{ 1, \frac{\pi(\boldsymbol{z}')q(\boldsymbol{z}|\boldsymbol{z}')}{\pi(\boldsymbol{z})q(\boldsymbol{z}'|\boldsymbol{z})} \right\}, \tag{2.8}$$

for deciding whether to accept or reject the newly proposed value $\boldsymbol{z}'$ at each iteration of the Markov chain. The MH algorithm is given in Algorithm 1, and it can be shown,

under simple conditions, that the distribution of the samples it produces converges to the target, $\pi(\boldsymbol{z})$ (Roberts and Smith, 1994).

Algorithm 1 The MH algorithm.

Let $B \in \mathbb{N}^+$ be the number of iterations
Assign initial values to $\boldsymbol{z}^{(0)}$
for $b = 1, \ldots, B$ **do**
  Sample $\boldsymbol{z}' \sim q(\boldsymbol{z}'|\boldsymbol{z}^{(b-1)})$
  Sample $u \sim \mathcal{U}_{[0,1]}$
  if $u \leq \alpha(\boldsymbol{z}', \boldsymbol{z}^{(b-1)})$ **then**
    $\boldsymbol{z}^{(b)} \leftarrow \boldsymbol{z}'$
  else
    $\boldsymbol{z}^{(b)} \leftarrow \boldsymbol{z}^{(b-1)}$
  end if
end for

2.3 Computationally Efficient Inference

Though MCMC methods are flexible and can be applied to nearly any geostatistical model, they are often computationally intensive. Furthermore, doing likelihood-based inference with the geostatistical model presented in (2.6) requires the evaluation of a multivariate Gaussian density with a dense $n \times n$ covariance matrix, $\boldsymbol{\Sigma}$. This entails calculating $|\boldsymbol{\Sigma}|$ and $\boldsymbol{\Sigma}^{-1}\boldsymbol{v}$ where $\boldsymbol{v}$ is an $n \times 1$ vector. Typically, a Cholesky factorization is used to speed up the calculations, but $\mathcal{O}(n^3)$ operations are still needed in general. This makes fast inference practically infeasible for standard approaches when $n \sim 10^3$ and higher. Therefore, additional tools are needed to ensure computational efficiency. One such tool is the class of Gaussian Markov random fields, whose spatial sparsity structure can be utilized to develop a computationally efficient inference scheme.

2.3.1 Gaussian Markov Random Fields

Gaussian Markov random fields (GMRFs) are defined in terms of a graph $\mathcal{G}$ and a precision matrix, which for an $n \times n$ covariance matrix $\boldsymbol{\Sigma}$, $n \geq 1$, is given by $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$. The graph consists of a vertex set, $\mathcal{V} = \{1, \ldots, n\}$, and a set of edges, $\mathcal{E}$. The graph of a GMRF with $\mathcal{V} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{2, 4\}\}$ is illustrated in Figure 2.1. Note that the edges are undirected, so an edge connecting vertex i and vertex j can be written as both $\{i, j\}$ and $\{j, i\}$. We define GMRFs in line with Rue and Held (2005).

Definition 2 (GMRF)**.** A random vector $\boldsymbol{x} = (x_1, \ldots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$, $n \geq 1$, is called a GMRF with respect to the graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with mean $\boldsymbol{\mu} \in \mathbb{R}^m$ and $n \times n$ positive definite precision matrix $\mathbf{Q}$ if

$$\pi(\boldsymbol{x}) = (2\pi)^{-n/2}|\mathbf{Q}|^{1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}}\mathbf{Q}(\boldsymbol{x} - \boldsymbol{\mu})\right), \quad \boldsymbol{x} \in \mathbb{R}^n, \tag{2.9}$$

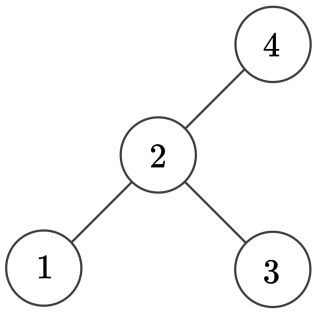


Figure 2.1: A graph with four vertices and three edges.

and

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

‖

By definition, a GMRF has a sparse precision matrix when the graph is sparse. To illustrate this more clearly, we note that GMRFs exhibit the pairwise Markov property, namely that

$$x_i \perp x_j \mid \boldsymbol{x}_{-ij}, \quad \text{if } \{i,j\} \notin \mathcal{E} \text{ and } i \neq j,$$

where $\boldsymbol{x}_{-ij}$ denotes the vector $\boldsymbol{x}$ stripped of element i and j. Furthermore, the diagonal elements of the precision matrix are given by the relationship $\mathrm{Q}_{ii}^{-1} = \mathrm{Var}(x_i|\boldsymbol{x}_{-i})$ and the off-diagonal elements are given by

$$\frac{\mathrm{Q}_{ij}}{\sqrt{\mathrm{Q}_{ii}\mathrm{Q}_{jj}}} = -\mathrm{Corr}(x_i, x_j|\boldsymbol{x}_{-ij}), \quad i \neq j.$$

Thus, we know that for a GMRF equipped with the graph in Figure 2.1, x_1 and x_4 are conditionally independent given all the other elements of the GMRF, since there is no direct edge between vertex 1 and vertex 4. Consequently, $\mathrm{Q}_{14} = \mathrm{Q}_{41} = 0$. On the other hand, since there is a direct edge between vertex 1 and vertex 2, we know that $\mathrm{Q}_{12} = \mathrm{Q}_{21} \neq 0$. Thus, x_1 and x_2 are conditionally *dependent* given all the other elements of the GMRF. We note the contrast between $\mathbf{Q}$ and $\boldsymbol{\Sigma}$, namely that the precision matrix describes the conditional variance structure of $\boldsymbol{x}$, whereas the covariance matrix describes the marginal variance structure of $\boldsymbol{x}$.

It is usually of interest to evaluate the log-density of a GMRF, i.e.,

$$\log \pi(\boldsymbol{x}) = -\frac{n}{2}\log(2\pi) + \log(|\mathbf{Q}^{1/2}|) - \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathbf{Q}\boldsymbol{x}.$$

In practice, this is calculated by computing the Cholesky factorization $\mathbf{L}\mathbf{L}^{\mathrm{T}} = \mathbf{Q}$, where $\mathbf{L}$ is a lower-triangular matrix. The log-density then simplifies to

$$\log \pi(\boldsymbol{x}) = -\frac{n}{2}\log(2\pi) + \sum_{i=1}^{n} \log|\mathrm{L}_{ii}| - \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\mathbf{L}\mathbf{L}^{\mathrm{T}}\boldsymbol{x}.$$

There exists numerical techniques such as bandwidth reduction and nested segmentation which ensure that $\mathbf{L}$ inherits the sparsity of $\mathbf{Q}$, which makes both the computation of $\mathbf{L}$ and the calculation of $\mathbf{L}\boldsymbol{x}$ and $\mathbf{L}^{\mathrm{T}}\boldsymbol{x}$ quicker. Typically, evaluation of the log-density requires $\mathcal{O}(n^{3/2})$ operations for $\mathbf{Q}$ with two-dimensional spatial sparsity (Rue and Held, 2005).

2.3.2 The SPDE Approach

As previously mentioned, inference with GRFs is computationally constrained by the need to factorize a dense covariance matrix. However, GMRFs (with sufficient spatial sparsity) do not suffer from this computational limitation. In their article, Lindgren et al. (2011) aim to take advantage of this, and establish a connection between GRFs and GMRFs. They propose an approach where a GRF is used in the modeling process, while computations are done on a corresponding GMRF that represents the GRF. This approach emerges from an SPDE. Specifically, if we let u be a centered GRF with a Matérn covariance function as defined in (2.1) with $\nu = 1$, the approach is based on the fact that u is a stationary solution to the SPDE given by

$$(\kappa^2 - \Delta)(\tau u(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathbb{R}^2, \tag{2.10}$$

where

$$\kappa = \frac{\sqrt{8}}{\rho_{\mathrm{S}}}, \quad \tau = \frac{1}{\sqrt{4\pi}\kappa\sigma_{\mathrm{S}}}, \tag{2.11}$$

Δ is the Laplacian, and $\mathcal{W}$ is standard Gaussian white noise. The operator $(\kappa^2 - \Delta)$ in Equation (2.10) is a local differential operator in the sense that it only depends on the value of u in an infinitesimal neighbourhood of $\boldsymbol{s}$. This implies that the stationary solution of the SPDE is a *Markovian* GRF (Simpson et al., 2012). A Markovian GRF is characterized informally as follows. Let A, B and C be a partition of $\mathcal{D}$ such that B separates A from C, as illustrated in Figure 2.2. Then, if u is a Markovian GRF,

$$\{u(\boldsymbol{s}) : \boldsymbol{s} \in A\} \perp \{u(\boldsymbol{s}) : \boldsymbol{s} \in C\} \mid \{u(\boldsymbol{s}) : \boldsymbol{s} \in B\}. \tag{2.12}$$

That is, the field evaluated on A is independent of the field evaluated on C, conditional on the field evaluated on B, which separates A from C. It is the "local" character of (2.10) which leads to the spatial Markov structure in the stationary solutions. In turn, this leads to sparse precision matrices and GMRFs when we (approximately) solve the SPDE. The reader is referred to Simpson et al. (2012) for more details on this topic.

Now, we find an approximation to the stochastic weak solution of (2.10) on the domain $\widetilde{\mathcal{D}} \supset \mathcal{D}$, which necessitates some boundary conditions. We follow Lindgren et al. (2011) and use a Neumann boundary condition. That is, we solve

$$(\kappa^2 - \Delta)(\tau u(\boldsymbol{s})) = \mathcal{W}(\boldsymbol{s}), \quad \boldsymbol{s} \in \widetilde{\mathcal{D}}, \quad \partial_{\boldsymbol{n}} u(\boldsymbol{s}) = 0, \quad \boldsymbol{s} \in \partial\widetilde{\mathcal{D}}, \tag{2.13}$$

where $\partial\widetilde{\mathcal{D}}$ is the boundary of $\widetilde{\mathcal{D}}$ and $\partial_{\boldsymbol{n}}$ is the gradient in the direction of the normal to the boundary, $\boldsymbol{n}$. The reason why we introduce the extension $\widetilde{\mathcal{D}}$ of $\mathcal{D}$ is to reduce boundary effects.

A triangulated mesh with $K \in \mathbb{N}^+$ vertices is constructed on the domain $\widetilde{\mathcal{D}}$, such as illustrated in Figure 2.3. Defining the inner product $\langle f, g \rangle = \int_{\widetilde{\mathcal{D}}} f(\boldsymbol{s}) g(\boldsymbol{s}) \mathrm{d}\boldsymbol{s}$, the approximation to the stochastic weak solution is found by requiring that

$$\{\langle \phi_j, (\kappa^2 - \Delta)u \rangle, j = 1, \ldots, K\} \overset{\mathrm{d}}{=} \{\tau^{-1} \langle \phi_j, \mathcal{W} \rangle, j = 1, \ldots K\}, \tag{2.14}$$

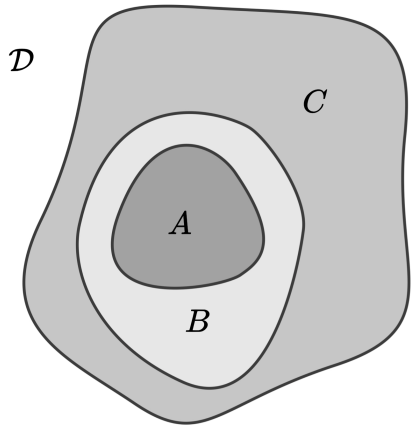


Figure 2.2: A partition of the domain, $\mathcal{D} = A \cup B \cup C$ where B separates A and C.

where "$\stackrel{\mathrm{d}}{=}$" denotes equality in distribution and $\{\phi_k\}_{k=1}^K$ is a specific set of test functions which we choose to be piecewise linear basis functions, such that $\phi_k = 1$ at vertex k and $\phi_k = 0$ at any other vertex in the mesh. Then, we construct a finite element representation of u by using the test functions as basis functions,

$$\tilde{u}(\boldsymbol{s}) = \sum_{k=1}^{K} \phi_k(\boldsymbol{s}) w_k, \quad \boldsymbol{s} \in \widetilde{\mathcal{D}}, \tag{2.15}$$

where the weights $\boldsymbol{w} = (w_1, \ldots, w_K)^{\mathrm{T}}$ are multivariate Gaussian-distributed. The left hand side of Equation (2.14) then becomes $\{\sum_{k=1}^K w_k \langle \phi_j, (\kappa^2 - \Delta)\phi_k \rangle, j = 1, \ldots, K\}$, which can be written as the matrix-vector product $\mathbf{K}\boldsymbol{w}$, where $\mathbf{K}$ is a $K \times K$ matrix with elements

$$\mathrm{K}_{ij} = \langle \phi_i, (\kappa^2 - \Delta)\phi_j \rangle = \kappa^2 \langle \phi_i, \phi_j \rangle + \langle \nabla \phi_i, \nabla \phi_j \rangle,$$

where the second transition follows from an application of Green's formula and the Neumann boundary condition.

Next, we define the two $K \times K$ matrices $\mathbf{C}$ and $\mathbf{G}$ with elements $\mathrm{C}_{ij} = \langle \phi_i, \phi_j \rangle$ and $\mathrm{G}_{ij} = \langle \nabla \phi_i, \nabla \phi_j \rangle$, respectively. Then, by the properties of standard Gaussian white noise (Lindgren, 2012, Chapter 2), the right hand side of Equation (2.14) is distributed as $\mathcal{N}(\mathbf{0}, \tau^{-2}\mathbf{C})$. Thus, the distribution of the weights is given by

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

so $\boldsymbol{w}$ has a precision matrix given by $\mathbf{Q}_w = \tau^2 \mathbf{K}\mathbf{C}^{-1}\mathbf{K}$. The matrices $\mathbf{C}$ and $\mathbf{G}$ are easily obtained, since their elements are nonzero only for the pairs of basis functions which share the same triangle in the mesh. Additionally, they do not depend on the parameters, κ and τ. The matrix $\mathbf{C}^{-1}$ is, however, dense, so the precision matrix of $\boldsymbol{w}$ is not yet sparse. But, as shown in Appendix C.5 of Lindgren et al. (2011), we can replace $\mathbf{C}$ by the diagonal matrix $\widetilde{\mathbf{C}}$ (in the construction of $\mathbf{K}$ also) with elements $\widetilde{\mathrm{C}}_{ii} = \langle \phi_i, 1 \rangle$. Then $\mathbf{Q}_w$ will be sparse, and we have a mapping from the parameters of the GRF u to the sparse precision matrix of its GMRF representation which takes $\mathcal{O}(K)$ operations to compute for any triangulation (Lindgren et al., 2011).

The GMRF representation of u can be used to approximate the latent process given by Equation (2.5),

$$\eta(\boldsymbol{s}) \approx \tilde{\eta}(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(\boldsymbol{s}) w_k, \quad \boldsymbol{s} \in \mathcal{D},$$

where the domain changes from $\widetilde{\mathcal{D}}$ to $\mathcal{D}$, because $\boldsymbol{x} : \mathcal{D} \to \mathbb{R}^p$ is only defined on $\mathcal{D}$. This gives the discretized latent process

$$\tilde{\boldsymbol{\eta}} = (\tilde{\eta}(\boldsymbol{s}_1), \ldots, \tilde{\eta}(\boldsymbol{s}_n))^{\mathrm{T}} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{w}, \tag{2.16}$$

where $\mathbf{A}$ is an $n \times K$ projection matrix with elements $\mathrm{A}_{ij} = \phi_j(\boldsymbol{s}_i)$, $i = 1, \ldots n$, $j = 1, \ldots K$. Note that $\mathbf{A}$ is sparse, since it has at most 3 non-zero elements per row by the choice of pyramidal basis functions. Thus, we can use numerical methods for sparse matrices to avoid the $\mathcal{O}(nK)$ cost of calculating $\mathbf{A}\boldsymbol{w}$ if $\mathbf{A}$ were dense.

To illustrate the computational gain from applying the SPDE approach, we consider the cost of evaluating the joint probability density of the geostatistical model in (2.6). That is, we want to evaluate the joint density

$$\pi(\boldsymbol{y}, \boldsymbol{\eta}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \pi(\boldsymbol{y}|\boldsymbol{\eta}, \sigma_{\mathrm{N}}^2) \cdot \pi(\boldsymbol{\eta}|\boldsymbol{\beta}, \sigma_{\mathrm{S}}^2, \rho_{\mathrm{S}}) \cdot \pi(\boldsymbol{\beta}) \cdot \pi(\boldsymbol{\theta}). \tag{2.17}$$

The first factor on the right hand side of Equation (2.17) requires $\mathcal{O}(n)$ operations, since we need to evaluate a univariate Gaussian density for each of the n observation. The second factor requires $\mathcal{O}(n^3)$ operations, since we in general need to factorize a dense $n \times n$ covariance matrix. The last two factors have a constant contribution to the computational cost (they do not scale with n), so evaluation of the density in (2.17) requires $\mathcal{O}(n + n^3) = \mathcal{O}(n^3)$ operations in total. However, we can use the approximation provided by the SPDE-approach, and in stead consider

$$\pi(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \pi(\boldsymbol{y}|\tilde{\boldsymbol{\eta}}, \sigma_{\mathrm{N}}^2) \cdot \pi(\boldsymbol{w}|\sigma_{\mathrm{S}}^2, \rho_{\mathrm{S}}) \cdot \pi(\boldsymbol{\beta}) \cdot \pi(\boldsymbol{\theta}). \tag{2.18}$$

The first factor on the right hand side of Equation (2.18) still requires $\mathcal{O}(n)$ operations, while the second factor by Section 2.3.1 now only requires $\mathcal{O}(K + K^{3/2}) = \mathcal{O}(K^{3/2})$ operations. Thus, evaluation of the density in Equation (2.18) requires $\mathcal{O}(n + K^{3/2})$ operations in total, which is linear as opposed to cubic in the number of observations, which represents a major reduction in computational cost.

2.3.3 TMB

Applying the SPDE approach to the geostatistical model in (2.6) gives a model which is parameterized by $\boldsymbol{\theta} = (\sigma_{\mathrm{N}}^2, \sigma_{\mathrm{S}}^2, \rho_{\mathrm{S}})^{\mathrm{T}}$ and the *random effects* $\boldsymbol{z} = (\boldsymbol{\beta}, \boldsymbol{w})^{\mathrm{T}}$. To estimate this model in a computationally efficient manner, we use the R-package TMB (Kristensen et al., 2016). TMB provides a fast and general framework for estimation of statistical models, by allowing the user to specify the joint distribution of the data and the latent variables (random effects) as a C++ template function. TMB also utilizes C++ libraries such as `CppAD` to perform automatic differentiation, making calculation of gradients and the Hessians quick (Kristensen et al., 2016).

We outline the estimation procedure in TMB. Since TMB is rooted in frequentist inference, we must distinguish between fixed effects and random effects. We treat $\boldsymbol{\theta} = (\rho_{\mathrm{S}}, \sigma_{\mathrm{N}}^2, \sigma_{\mathrm{S}}^2)^{\mathrm{T}}$ as fixed effects, but we will work with $\boldsymbol{\phi} = (\log\kappa, \log\tau, \log\sigma_{\mathrm{N}})^{\mathrm{T}}$ instead, due to how priors are defined and because we want to avoid constraints

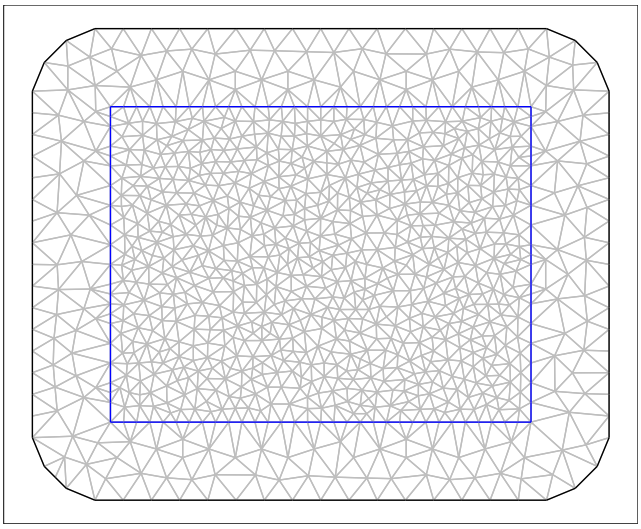


Figure 2.3: A triangulated mesh over a rectangular domain. The area outside the blue boundary is needed to reduce boundary effects.

such as $\sigma_{\mathrm{N}}^2 > 0$ in the optimization routine. Recall that Equation (2.11) gives the relationship between $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$. The vector $\boldsymbol{z} = (\boldsymbol{\beta}, \boldsymbol{w})^{\mathrm{T}}$ denotes the random effects, which consists of $p + K$ elements, p covariate coefficients and K weights. TMB produces marginal maximum *a posteriori* (MMAP) estimates of the fixed effects, $\boldsymbol{\phi}$, by maximizing the marginal posterior of $\boldsymbol{\phi}$, $\pi(\boldsymbol{\phi}|\boldsymbol{y})$. We can express the marginal posterior of the fixed effects as

$$\begin{aligned}
\pi(\boldsymbol{\phi}|\boldsymbol{y}) &= \int \pi(\boldsymbol{\phi}, \boldsymbol{z}|\boldsymbol{y}) \mathrm{d}\boldsymbol{z} \propto \int \pi(\boldsymbol{\phi}, \boldsymbol{z}, \boldsymbol{y}) \mathrm{d}\boldsymbol{z} \\
&= \int \pi(\boldsymbol{y}|\boldsymbol{\phi}, \boldsymbol{z}) \pi(\boldsymbol{z}|\boldsymbol{\phi}) \pi(\boldsymbol{\phi}) \mathrm{d}\boldsymbol{z} \\
&= \int \exp\left\{\log \pi(\boldsymbol{y}|\boldsymbol{\phi}, \boldsymbol{z}) + \log \pi(\boldsymbol{z}|\boldsymbol{\phi}) + \log \pi(\boldsymbol{\phi})\right\} \mathrm{d}\boldsymbol{z} \\
&= \int \exp\left\{-f(\boldsymbol{\phi}, \boldsymbol{z})\right\} \mathrm{d}\boldsymbol{z} =: \mathcal{L}^*(\boldsymbol{\phi}). \qquad (2.19)
\end{aligned}$$

where $f(\boldsymbol{\phi}, \boldsymbol{z}) = -(\log \pi(\boldsymbol{y}|\boldsymbol{\phi}, \boldsymbol{z}) + \log \pi(\boldsymbol{z}|\boldsymbol{\phi}) + \log \pi(\boldsymbol{\phi}))$.

The estimation procedure in TMB consists of a two-step nested optimization routine. First, given a set of parameter values for $\boldsymbol{\phi}$, new modal values are found for the random effects by nonlinear optimization of

$$\widehat{\boldsymbol{z}}(\boldsymbol{\phi}) := \arg\min_{\boldsymbol{z}} \ f(\boldsymbol{\phi}, \boldsymbol{z}). \qquad (2.20)$$

Then, $\mathcal{L}^*(\boldsymbol{\phi})$ as defined in (2.19) is approximated with a Laplace approximation,

$$\mathcal{L}^*(\boldsymbol{\phi}) \approx \widetilde{\mathcal{L}}^*(\boldsymbol{\phi}) = (2\pi)^{(p+K)/2} |\mathbf{H}(\boldsymbol{\phi})|^{1/2} \exp\{-f(\boldsymbol{\phi}, \widehat{\boldsymbol{z}}(\boldsymbol{\phi}))\}, \qquad (2.21)$$

where $\mathbf{H}(\boldsymbol{\phi})$ is the Hessian of $f(\boldsymbol{\phi}, \boldsymbol{z})$ with respect to $\boldsymbol{z}$, denoted by $\mathbf{H}$, evaluated at $\boldsymbol{z} = \widehat{\boldsymbol{z}}(\boldsymbol{\phi})$. The current values of the random effects are inserted into (2.21), which is optimized to find new modal values $\widehat{\boldsymbol{\phi}}$ of $\boldsymbol{\phi}$. If the maximum gradient

component of $\widetilde{\mathcal{L}}^*(\boldsymbol{\phi})$ at $\boldsymbol{\phi} = \widehat{\boldsymbol{\phi}}$ is below some threshold, the estimation procedure stops. Otherwise, we return to (2.20) and insert the newly found estimates of $\boldsymbol{\phi}$.

To ensure computational efficiency, it is important that $\mathbf{H}$ is sparse. In that regard, note that for $i, j > 3 + p$,

$$\begin{aligned} \mathrm{H}_{ij} &= \frac{\partial^2 f(\boldsymbol{\phi}, \boldsymbol{z})}{\partial w_i \partial w_j} \\ &= -\frac{\partial^2}{\partial w_i \partial w_j}(\log \pi(\boldsymbol{y}|\boldsymbol{\eta}, \sigma_{\mathrm{N}}^2) + \log \pi(\boldsymbol{w}|\sigma_{\mathrm{S}}^2, \rho_{\mathrm{S}})) \\ &= -\left[\frac{\partial^2}{\partial w_i \partial w_j}(\log \pi(\boldsymbol{y}|\boldsymbol{\eta}, \sigma_{\mathrm{N}}^2) + (\mathbf{Q}_w)_{ij}\right]. \end{aligned} \tag{2.22}$$

Thus, $\mathrm{H}_{ij} = (\mathbf{Q}_w)_{ij}$ when $\pi(\boldsymbol{y}|\boldsymbol{\eta}, \sigma_{\mathrm{N}}^2)$ does not depend on w_i or w_j. In that way, $\mathbf{H}$ inherits the sparsity of $\mathbf{Q}_w$. TMB has automatic sparsity detection, which utilizes this effect (Kristensen et al., 2016).

The R-package INLA Rue et al. (2009) can also be used to estimate the model considered here, and we do precisely this in order to verify the implementation of a sample-based approach in Appendix C. INLA is amenable to the class of latent Gaussian models, and estimates the marginal posteriors of model parameters. However, when we introduce positional uncertainty, the observation likelihood becomes a mixture distribution which INLA does not have the flexibility to handle. Thus, we use TMB, which enables the user to manually specify the distribution of the data. For a comprehensive overview of TMB and a comparison to INLA in the context of spatial statistics, the reader is referred to Osgood-Zimmerman and Wakefield (2021).

2.4 Scoring Rules

To evaluate the predictive power of estimated models, we use RMSE for point predictions and CRPS for the predictive distributions. Let $\hat{\boldsymbol{y}} = (\hat{y}_1, \ldots, \hat{y}_{n_{\mathrm{P}}})^{\mathrm{T}}$ denote a set of point predictions with corresponding true values $\boldsymbol{y} = (y_1, \ldots, y_{n_{\mathrm{P}}})^{\mathrm{T}}$. The RMSE is then given by

$$\mathrm{RMSE} = \sqrt{\frac{1}{n_{\mathrm{P}}} \sum_{i=1}^{n_{\mathrm{P}}} (\hat{y}_i - y_i)^2}.$$

We prefer the model with lowest RMSE.

For each observation, y_i, $i = 1 \ldots n$, let F_i denote the cumulative predictive distribution obtained from the estimation procedure. The CRPS compares F_i and y_i, and is defined as follows,

$$\mathrm{CRPS}(F_i, y_i) = \int_{\mathbb{R}} (F_i(x) - \mathbb{I}(x \geq y_i))^2 \mathrm{d}x, \tag{2.23}$$

where $\mathbb{I}$ is the indicator function and lower values of (2.23) are preferred. CRPS generalizes the absolute error. Specifically, if F_i were deterministic, $\mathrm{CRPS}(F_i, y_i) = |\hat{y}_i - y_i|$. Additionally, CRPS is a proper scoring rule, which means that if P is the true cumulative predictive distribution and F is the proposed cumulative predictive distribution, then $\mathrm{E}_P[\mathrm{CRPS}(P, y)] \leq \mathrm{E}_P[\mathrm{CRPS}(F, y)]$ where $y \sim P$ for all P and F (Gneiting and Raftery, 2007).

For a given model, we consider the average CRPS over all predictions,

$$\overline{\text{CRPS}} = \frac{1}{n_{\text{P}}} \sum_{i=1}^{n_{\text{P}}} \text{CRPS}(F_i, y_i),$$

where we prefer the model with lowest average CRPS. In the following, we drop "average", when referring to the average CRPS of a model. To calculate CRPS numerically, samples are drawn from the predictive distributions, and we use the approximation of (2.23) as it is implemented in the R-package `scoringRules` (Jordan et al., 2019).

Chapter 3

Observations with Positional Uncertainty

We opened this thesis by noting that the assumption of exact observation locations may not always be reasonable for a variety of reasons: coordinates could be intentionally rounded for convenience, uncertainty in the measured locations could be large relative to the size of the domain, and masking of the true locations due to privacy concerns could have been applied. In these scenarios, we do not know the true locations, i.e., the locations at which the observations are taken. Instead, we only have access to some related but incorrect counterparts. We say that every unknown true location, $\boldsymbol{s}^* \in \mathcal{D}$, has an observed counterpart, $\boldsymbol{s} \in \mathcal{D}$. Section 3.1 begins by reviewing proposed approaches to account for positional uncertainty. Then, Section 3.2 gives an overview of geomasking and describes the jittering schemes we focus on in detail. Finally, Section 3.3 gives some theoretical results.

3.1 Accounting for Positional Uncertainty

3.1.1 Geostatistical Approaches

Positional uncertainty has received some attention in the literature of geostatistics, starting with the paper of Gabrosek and Cressie (2002), who derive *kriging* equations adjusted for location error, and find that, when the location error is large, this adjustment gives markedly better performance with respect to mean squared prediction error. Cressie and Kornak (2003) continue on this work, and develop more general kriging equations for geostatistical models with non-constant means. They discuss two models of positional error, the *coordinate positioning* (CP) model and the *feature-positioning* (FP) model. The CP model is appropriate when we know the intended sampling location, $\boldsymbol{s}$, but, in the "attempt" to sample from $\boldsymbol{s}$, we "mistakenly" sample from $\boldsymbol{s}^*$ instead. In this situation, the positional error is modeled as

$$\boldsymbol{s}^* = \boldsymbol{s} + \boldsymbol{\epsilon},$$

and we specify the distribution of $\boldsymbol{s}^*|\boldsymbol{s}$. On the other hand, the FP model is appropriate when we know the feature from which to sample from, but the location of the feature is determined with some error. Then, the positional error is modeled as

$$\boldsymbol{s} = \boldsymbol{s}^* + \boldsymbol{\epsilon},$$

and we specify the distribution of $\boldsymbol{s}|\boldsymbol{s}^*$. These two models of positional error are connected to the theory of measurement error models, in which the CP model corresponds to the Berkson measurement error model, while the FP model correspond to the classical measurement error model (Carroll et al., 2006).

Cressie and Kornak (2003) continue with the CP model for positional error, and calculate the moments of an observation process which is adjusted for positional error. Using the latent model described in Section 2.1.2, we can analogously define the adjusted latent process

$$\eta^*(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s}+\boldsymbol{\epsilon})^{\mathrm{T}}\boldsymbol{\beta} + u(\boldsymbol{s}+\boldsymbol{\epsilon}), \quad \boldsymbol{s} \in \mathcal{D}.$$

The mean function of this process can be calculated as

$$m^*(\boldsymbol{s}) = \mathrm{E}[\eta^*(\boldsymbol{s})] = \left(\int \boldsymbol{x}(\boldsymbol{s}+\boldsymbol{\epsilon})\pi(\boldsymbol{\epsilon})\mathrm{d}\boldsymbol{\epsilon}\right)^{\mathrm{T}}\boldsymbol{\beta},$$

while the stationary covariance function is given by

$$C^*(\boldsymbol{h}) = \mathrm{Cov}[\eta^*(\boldsymbol{s}), \eta^*(\boldsymbol{s}+\boldsymbol{h})] = \int\int C(\boldsymbol{h}+\boldsymbol{\epsilon}_1-\boldsymbol{\epsilon}_2)\pi(\boldsymbol{\epsilon}_1)\pi(\boldsymbol{\epsilon}_2)\mathrm{d}\boldsymbol{\epsilon}_1\mathrm{d}\boldsymbol{\epsilon}_2.$$

They then proceed by assuming that the adjusted spatial process with this mean and covariance function is a new Gaussian process, and use this assumption to estimate the model parameters. As they point out, this is a pseudo-likelihood approach, since the assumption of Gaussianity is not true in general. To make predictions of the observation process at a desired location $\boldsymbol{s}_0^* \in \mathcal{D}$, they consider the class of linear predictors using the observations which arise from the latent process η^*. This, combined with the parameter estimates from the pseudo-likelihood approach, gives kriging adjusted for positional error. In a simulation experiment, they find that this adjustment leads to more efficient predictions with less mean squared prediction error. However, the approach has some clear drawbacks, namely the erroneous assumption that the adjusted observation process is Gaussian and the seemingly difficult task of extending the approach to generalized geostatistical models.

Fanshawe and Diggle (2011) consider Gaussian geostatistical models such as (2.6), and specify the positional error model through the conditional distribution $\pi(\boldsymbol{s}^*|\boldsymbol{s})$ and a prior, $\pi(\boldsymbol{s})$, which is taken to be uniform over $\mathcal{D}$. We note that this corresponds to the CP model described above. Specifically, they use a bivariate Gaussian distribution to model the positional error,

$$\boldsymbol{s}^*|\boldsymbol{s} \sim \mathcal{N}_2(\boldsymbol{s}, \delta^2\mathbf{I}_2), \tag{3.1}$$

where $\delta > 0$. The resulting likelihood of the observations is evaluated with Monte Carlo integration. However, computation times are found to be prohibitively long when there is positional error in the observation locations. Fronterrè et al. (2018) continue on this work and try to overcome the computational limits by using a composite likelihood (CL) approach, where only pairs of observations are considered to depend on each other. Additionally, they apply this approach to DHS data, and consequently model the positional error in terms of the FP model. In a simulation study, they find the CL approach to give substantially smaller RMSE for the parameter estimates. However, the effect of the CL approximation on the

accuracy of inference is not investigated, and the authors do not discuss how the approach can be extended to other observation distributions. Additionally, they approximate the geomasking strategy used in DHS data with the bivariate positional error model.

Wilson and Wakefield (2021) consider DHS data, and develop an estimation procedure which applies to latent Gaussian models under positional uncertainty. Furthermore, the exact geomasking strategy employed in DHS is modeled through the conditional distribution $\pi(\boldsymbol{s}|\boldsymbol{s}^*)$, which resembles the FP-model for positional error. They employ the SPDE approach and use an INLA within MCMC algorithm (Gómez-Rubio and Rue, 2018) to estimate the model which accounts for positional error. In each iteration of the algorithm, the unknown true locations are sampled, before INLA is used to construct marginal posteriors of what corresponds to $\boldsymbol{\phi}$ and $\boldsymbol{z}$ in this text (see Section 2.3.3), from which samples are drawn. This method uses a *masterframe* of sampling locations. Additionally, the computational time of the method inhibits fast inference. The paper reports a running time of 52 hours to run 1000 iterations of the INLA within MCMC method for each scenario, mainly owing to the estimation with INLA in each iteration.

The above presentation highlights the need for a computationally efficient and flexible approach to inference under positional uncertainty. The numerical approach presented by Altay et al. (2022a) achieves computationally efficient inference by using the SPDE approach and TMB, and can easily be extended to generalized linear geostatistical models. A thorough description of the numerical approach is given in Section 4.2.

3.1.2 Covariate-specific Approaches

The findings of Altay et al. (2022b) and Totland (2022) indicate that the foremost concern when doing geostatistical inference under jittering is to account for the added uncertainty in the spatial covariates, $\boldsymbol{x}(\boldsymbol{s})$, $\boldsymbol{s} \in \mathcal{D}$. Specifically, Altay et al. (2022b) obtain nearly identical results when only accounting for jittering in the covariates and when accounting for jittering in both covariates and the spatial field. This result is reasonable, as the GRF used to model spatial correlation is typically much smoother and less variable than raster-based covariates at the spatial scale of jittering. This connects to the discussion of Fanshawe and Diggle (2011), who find that the local gradient of the underlying surface which we are trying to model may be important in determining the impact of positional errors.

Some techniques to specifically adjust for positional error in covariates have been proposed in the context of DHS data. Perez-Heydrich et al. (2016) propose a regression calibration approach for distance-based covariates, and Warren et al. (2016) suggest that raster-based covariates can be exchanged with averages of neighborhood buffers. The latter approach clearly does not address the issue of attenuation in the covariate coefficients. Furthermore, applying these approaches to the covariates changes the meaning of the covariate coefficients, $\boldsymbol{\beta}$, making the model more difficult to interpret. Therefore, we do not consider these approaches in the following, and in stead focus on the numerical approach with adjustment for jittering in both the spatial covariates and the spatial field.

3.2 Positional Uncertainty from Geomasking

When geographical data is collected along with other features of, e.g., a household or an individual, there is often a conflict of interest between making the geo-referenced data widely available and protecting the privacy of data subjects. This dilemma is present in areas such as epidemiology (Cassa et al., 2006), environmental studies (Fronterrè et al., 2018) and demographic and health studies (Burgert et al., 2013). Geomasking terms the practice of concealing the exact locations of observations in a dataset by applying a stochastic or deterministic masking procedure. In that way, geomasking can be used to strike a balance between the competing interests of privacy protection and the benefits of making the data widely available, such as allowing more researchers to analyze and produce insights from the geo-referenced data. The specific choice of geomasking strategy depends on the nature of the data, for which the risk of disclosure should be compared to the potential consequences of disclosure.

Armstrong et al. (1999) study geomasking in the context of preserving health record confidentiality. They find that in stead of using the common approach of areal aggregation of records, the records can be masked by randomly displacing their locations within a disk, and that this may give a better trade-off between privacy protection and utilization of the data. Here, we will study this type of geomasking, referred to as *disk jittering* (also referred to as *random perturbation* and *random direction-random distance method* in the literature) in addition to *donut jittering* (also referred to as *donut masking* in the literature), which are closely described in the following paragraphs. There exists multiple other geomasking strategies to choose from, and we refer to Zandbergen (2014) for a review on this topic.

In disk jittering, the geographic coordinates of a data point are displaced randomly by uniformly generating an angle between 0 and 2π radians and displacing the point in the direction of that angle. The displacement distance is uniformly generated and is restricted to be less than the radius of some circle. This is illustrated on the left side of Figure 3.1, where the true location (purple point) is randomly perturbed to a new location (green point) within a disk with radius $R_2 > 0$. We model disk jittering according to the FP model, i.e., we specify

$$\pi(\boldsymbol{s}|\boldsymbol{s}^*) \propto \frac{\mathbb{I}(d(\boldsymbol{s}, \boldsymbol{s}^*) < R_2)}{d(\boldsymbol{s}, \boldsymbol{s}^*)},$$

where $d(\boldsymbol{s}, \boldsymbol{s}^*)$ is the distance between $\boldsymbol{s}$ and $\boldsymbol{s}^*$ in appropriate units. Since we do Bayesian statistics, we also need to specify a prior on the unknown true location, $\pi(\boldsymbol{s}^*)$, which is discussed in Section 4.1. Importantly, the DHS program uses a combination of disk jittering schemes in their geomasking strategy (Burgert et al., 2013).

Donut jittering is similar to disk jittering: an angle is uniformly drawn between 0 and 2π radians before the coordinate is displaced in the direction of that angle. The displacement distance is uniformly drawn such that the jittered point is restricted to lie within an annulus with inner radius $R_1 > 0$ and outer radius $R_2 > R_1$, as is illustrated on the right side of Figure 3.1. We also model donut jittering according to the FP model, i.e., we specify

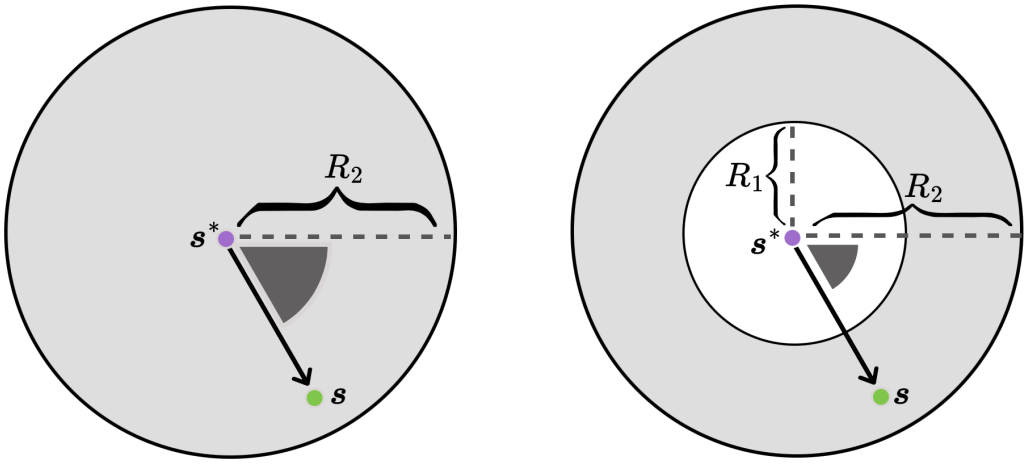


Figure 3.1: Illustration of disk jittering (left) and donut jittering (right). The true location (purple) is jittered to a new location (green point). In disk jittering, the displaced coordinate is restricted to lie within a disk with radius R_2, while in donut jittering, the displaced point is restricted to lie within an annulus with inner radius R_1 and outer radius R_2.

$$\pi(\boldsymbol{s}|\boldsymbol{s}^*) \propto \frac{\mathbb{I}(R_1 < d(\boldsymbol{s}, \boldsymbol{s}^*) < R_2)}{d(\boldsymbol{s}, \boldsymbol{s}^*)}.$$

Donut jittering has been advocated as a superior alternative to disk jittering on the basis of better privacy protection (Hampton et al., 2010; Allshouse et al., 2010). In these papers, privacy protection is quantified with the k-anonymity metric, which measures the number of individuals a de-identified study object cannot be distinguished from (Sweeney, 2002). However, the use of this metric by Hampton et al. (2010) and Allshouse et al. (2010) presupposes that the geomasking strategy is unknown to the analysts. Thus, their results do not directly apply to the situation studied in this thesis, where we want to utilize the *known* geomasking strategy to improve inference. Therefore, we in stead investigate the privacy protection properties of different geomasking strategies by considering the posterior of the unknown true locations (see Section 6.3).

We note that donut and disk jittering may be framed within the same class of positional error models, where disk jittering is just donut jittering with $R_1 = 0$. In the following, we refer to different combinations of R_1 and R_2 as different *jittering schemes*, and we may refer to the application of these schemes as *jittering* of the observation locations.

3.3 Theoretical Results

Some theoretical results have been developed for geostatistical inference in the presence of positional uncertainty. Fronterrè et al. (2018) consider the effect geomasking has on the spatial covariance structure, and calculate a theoretical variogram in the presence of positional uncertainty. Here, we carefully derive the same result.

First, the positional error model corresponds to the FP model described in Section 3.1, and is of the bivariate Gaussian form,

$$\boldsymbol{s}|\boldsymbol{s}^* \sim \mathcal{N}_2(\boldsymbol{s}^*, \delta^2\mathbf{I}_2), \tag{3.2}$$

with a positional error variance $\delta^2 > 0$. We consider the model given by (2.6), and let $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$ denote the observed locations while $\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_n^*$ denote the corresponding unknown true locations. Next, we introduce some notation, where we let

$$\begin{aligned} v_{ij} &= \frac{1}{2}\left[(y_i - \boldsymbol{x}(\boldsymbol{s}_i^*)^{\mathrm{T}}\boldsymbol{\beta})) - (y_j - \boldsymbol{x}(\boldsymbol{s}_j^*)^{\mathrm{T}}\boldsymbol{\beta}))\right]^2, \\ u_{ij} &= \|\boldsymbol{s}_i - \boldsymbol{s}_j\|, \quad u_{ij}^* = \|\boldsymbol{s}_i^* - \boldsymbol{s}_j^*\|, \quad i, j = 1, \ldots n,\ i \neq j. \end{aligned}$$

Additionally, we denote the components of the observed and true locations by $\boldsymbol{s}_i = (s_{i1}, s_{i2})^{\mathrm{T}}$ and $\boldsymbol{s}_i^* = (s_{i1}^*, s_{i2}^*)^{\mathrm{T}}$. With this notation, we define the variogram without positional uncertainty,

$$\gamma^*(u_{ij}^*) = \mathrm{E}[v_{ij}|u_{ij}^*] = \sigma_{\mathrm{N}}^2 + \sigma_{\mathrm{S}}^2(1 - \rho(u_{ij}^*)),$$

where $\rho(u_{ij}^*) = \sigma_{\mathrm{S}}^{-2}C(u_{ij}^*)$ is the isotropic correlation function of the GRF. With positional uncertainty, we do not know u_{ij}^*, and instead seek to find an expression for $\gamma(u_{ij}) = \mathrm{E}[v_{ij}|u_{ij}]$. We assume $v_{ij} \perp u_{ij} \mid u_{ij}^*$. Then, the pdf of of $v_{ij}|u_{ij}$ is given by

$$\pi(v_{ij}|u_{ij}) = \int_0^\infty \pi(v_{ij}|u_{ij}^*)\pi(u_{ij}^*|u_{ij})\mathrm{d}u_{ij}^*,$$

where $\pi(v_{ij}|u_{ij}^*)$ is the pdf of a scaled chi-squared distribution with 1 degree of freedom, i.e., $v_{ij}|u_{ij}^* \sim \gamma^*(u_{ij}^*)\chi_1^2$. On the other hand, $\pi(u_{ij}^*|u_{ij})$ is the pdf of a Rice distribution, $u_{ij}^*|u_{ij} \sim \mathrm{Rice}(u_{ij}, \sqrt{2}\delta)$. To see this, we note that, by the positional error model,

$$\begin{aligned} (s_{i1}^* - s_{j1}^*) \mid s_{i1}, s_{j1} &\sim \mathcal{N}(s_{i1} - s_{j1}, 2\delta^2) = \mathcal{N}(u_{ij}\cos\theta, 2\delta^2), \\ (s_{i2}^* - s_{j2}^*) \mid s_{i2}, s_{j2} &\sim \mathcal{N}(s_{i2} - s_{j2}, 2\delta^2) = \mathcal{N}(u_{ij}\sin\theta, 2\delta^2), \end{aligned}$$

where the angle $\theta \in (0, 2\pi)$ is illustrated in Figure 3.2. Thus, treating δ and θ as fixed, we can equivalently write

$$\begin{aligned} (s_{i1}^* - s_{j1}^*) \mid u_{ij} &\sim \mathcal{N}(u_{ij}\cos\theta, 2\delta^2), \\ (s_{i1}^* - s_{j1}^*) \mid u_{ij} &\sim \mathcal{N}(u_{ij}\sin\theta, 2\delta^2). \end{aligned}$$

Then, by Appendix A of Fronterrè et al. (2018), the random variable

$$u_{ij}^*|u_{ij} = \sqrt{(s_{i1}^* - s_{j1}^*)^2 + (s_{i1}^* - s_{j1}^*)^2} \mid u_{ij} \sim \mathrm{Rice}(u_{ij}, \sqrt{2}\delta).$$

Now, we can develop an expression for $\gamma(u_{ij})$ by employing the law of total expectation,

$$\begin{aligned} \gamma(u_{ij}) &= \mathrm{E}[v_{ij}|u_{ij}] = \mathrm{E}\left[\mathrm{E}[v_{ij}|u_{ij}^*] \mid u_{ij}\right] \\ &= \sigma_{\mathrm{N}}^2 + \sigma_{\mathrm{S}}^2\left(1 - \mathrm{E}\left[\rho(u_{ij}^*)|u_{ij}\right]\right). \end{aligned} \tag{3.3}$$

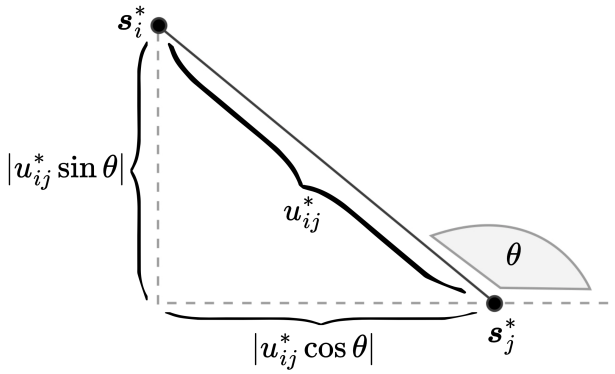


Figure 3.2: Illustration of how the angle θ is constructed in relation to two locations, $\boldsymbol{s}_i^*$ and $\boldsymbol{s}_j^*$.

From (3.3), we see that as the scale of the positional error diminishes, i.e., $\delta \to 0$ and $u_{ij} \to u_{ij}^*$, we retain the true variogram, $\gamma(u_{ij}) \to \gamma^*(u_{ij}^*)$. In general, $\mathrm{E}\left[\rho(u_{ij}^*|u_{ij})\right]$ is not available in closed form. However, we can derive a closed form expression for (3.3) when the covariance function is Gaussian. Specifically, the Gaussian correlation function is given by

$$\rho(u_{ij}^*) = \exp\left\{-\left(\frac{u_{ij}^*}{\rho_{\mathrm{S}}/\sqrt{8}}\right)^2\right\}.$$

Then, one can show (see Appendix A) that

$$\mathrm{E}\left[\rho(u_{ij}^*) \mid u_{ij}\right] = \frac{1}{1+4r^2} \exp\left\{-\left(\frac{u_{ij}}{\rho_{\mathrm{S}}/\sqrt{8}\cdot\sqrt{1+4r^2}}\right)^2\right\}, \tag{3.4}$$

where $r = \sqrt{8}\delta/\rho_{\mathrm{S}}$. The true variogram is retained if $r \to 0$, while if $r \to \infty$, the spatial correlation structure is destroyed, and $\gamma(u_{ij}) \to \sigma_{\mathrm{N}}^2 + \sigma_{\mathrm{S}}^2$. Furthermore, as $u_{ij} \to 0$, $\mathrm{E}\left[\rho(u_{ij}^*) \mid u_{ij}\right] \to (1+4r^2)^{-1} \leq 1$, so that $\gamma(0) \geq \sigma_{\mathrm{N}}^2$. Thus, we may infer that the positional uncertainty in this scenario leads to biased estimates of the spatial range and the variance of the nugget effect. Specifically, we would expect the spatial range to be underestimated, because of the diminishing spatial correlation structure, and the nugget effect to be overestimated, because of the added uncertainty.

We focus on disk and donut jittering in the following chapters as opposed to positional error models of the bivariate normal form. Additionally, we use a Matérn covariance function as opposed to a Gaussian one. Nonetheless, the jittering distributions are similar to the bivariate normal distribution: they are symmetric around the true location and decay radially outwards from it (from R_1 and outwards). Furthermore, the Gaussian covariance function is a limiting case of the Matérn covariance function, as $\nu \to \infty$. Therefore, we expect the ratio R_2/ρ_{S} to play a similar role in our simulations under disk and donut jittering as δ/ρ_{S} does here. Specifically, we expect that increasing the ratio R_2/ρ_{S} will induce a larger underestimation of ρ_{S} and a larger overestimation of σ_{N}^2 when jittering is ignored.

Chapter 4

Geostatistical Inference under Positional Uncertainty

In this chapter, we describe the Gaussian geostatistical model which incorporates positional uncertainty in Section 4.1. Then, two approaches to estimation and inference with this model are presented. Section 4.2 describes the numerical approach, and Section 4.3 describes and constructs the sample-based approach. We remark Sections 4.1 and 4.2 are largely based on the work of Totland (2022).

4.1 Model Description

In the presence of positional uncertainty, such as jittering, we have a set of true locations, denoted by $\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_n^* \in \mathcal{D} \subset \mathbb{R}^2$, and a corresponding set of jittered locations, denoted by $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$. Figure 4.1 illustrates the application of disk jittering to a set of (true) locations. Importantly, since the true locations are unknown, and the observed locations have been displaced randomly, we treat both $\boldsymbol{s}_i^*$ and $\boldsymbol{s}_i$, $i = 1, \ldots n$, as random quantities.

The data originates from the true locations, so the observation model given by

$$y_i | \eta(\boldsymbol{s}_i^*), \sigma_{\mathrm{N}}^2 \sim \mathcal{N}(\eta(\boldsymbol{s}_i^*), \sigma_{\mathrm{N}}^2), \quad i = 1, \ldots n,$$

with latent model

$$(\eta(\boldsymbol{s}_1^*), \ldots \eta(\boldsymbol{s}_n^*))^{\mathrm{T}} | \boldsymbol{\beta}, \sigma_{\mathrm{S}}^2, \rho_{\mathrm{S}}, \boldsymbol{s}_1^*, \ldots \boldsymbol{s}_n^* \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma}_{\mathrm{S}}(\sigma_{\mathrm{S}}^2, \rho_{\mathrm{S}})),$$

corresponds exactly to the standard geostatistical model described in Section 2.1.2. However, only the jittered locations, $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, are known in this setting, so we need an observation model which incorporates the uncertainty which this adds. To obtain this, we specify a model for the positional uncertainty through the conditional distribution of an observed location given its corresponding true location and a prior on the unknown true location,

$$\boldsymbol{s}_i | \boldsymbol{s}_i^* \sim \pi(\boldsymbol{s}_i | \boldsymbol{s}_i^*), \quad \boldsymbol{s}_i^* \sim \pi(\boldsymbol{s}_i^*), \quad i = 1, \ldots n. \tag{4.1}$$

For the prior on the unknown true locations, $\pi(\boldsymbol{s}_i^*)$, one could, e.g., use the population density if this is known *a priori*. Population density maps may be available through sources such as WorldPop et al. (2018), but these surfaces are modeled surfaces with inherent bias and uncertainty, which necessities an investigation of the impact

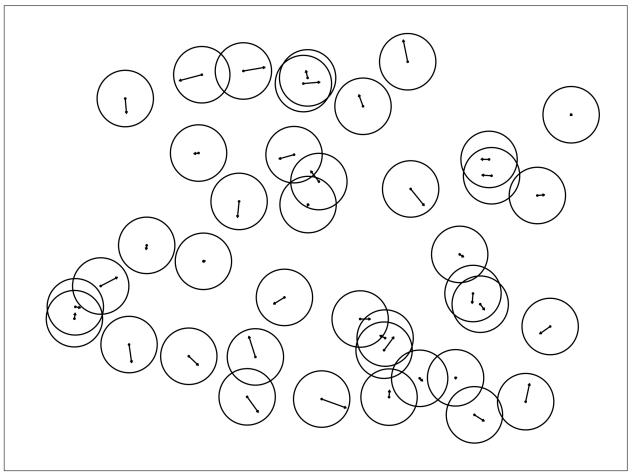


Figure 4.1: An illustration of disk jittering. The points are randomly displaced as indicated by the arrows. The circle around each point indicates the outer radius of the jittering disk.

their use would have on inference. We do not consider this further, and in stead place a uniform prior over the domain of interest, i.e., we let $\pi(\boldsymbol{s}_i^*) \sim \mathcal{U}(\mathcal{D})$. The positional error model in (4.1) corresponds to the FP model for positional error presented in Section 3.1. That is, we view the true locations, $\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_n^*$, to be the locations of some features we are interested in, e.g., clusters in a health study, whose true locations are reported with some error, so that we instead observe $\boldsymbol{s}_i = \boldsymbol{s}_i^* + \boldsymbol{\epsilon}_i,\ i = 1, \ldots, n$.

To summarize, the Gaussian model which incorporates positional uncertainty in the observation locations is given by

$$y_i|\eta(\boldsymbol{s}_i^*), \sigma_{\mathrm{N}}^2 \sim \mathcal{N}(\eta(\boldsymbol{s}_i^*), \sigma_{\mathrm{N}}^2), \quad \boldsymbol{s}_i|\boldsymbol{s}_i^* \sim \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*), \quad i = 1, \ldots n,$$

where the pairs $(y_1, \boldsymbol{s}_1), \ldots (y_n, \boldsymbol{s}_n)$ represent the observations. We assume that the observations are conditionally independent, given the latent field and the parameter(s) in the observation layer,

$$\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n|\eta, \sigma_{\mathrm{N}}^2) = \prod_{i=1}^{n} \pi(y_i, \boldsymbol{s}_i|\eta, \sigma_{\mathrm{N}}^2). \tag{4.2}$$

Both the numerical and sample-based approach rely on the SPDE approach presented in Section 2.3.2. That is, we approximate the latent process by

$$\eta(\boldsymbol{s}) \approx \tilde{\eta}(\boldsymbol{s}) = \boldsymbol{x}(\boldsymbol{s})^{\mathrm{T}}\boldsymbol{\beta} + \sum_{k=1}^{K} \phi_k(\boldsymbol{s}) w_k, \quad \boldsymbol{s} \in \mathcal{D}.$$

Using this approximation, the joint likelihood in Equation (4.2) becomes

$$\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n|\tilde{\eta}, \sigma_{\mathrm{N}}^2) = \prod_{i=1}^{n} \pi(y_i, \boldsymbol{s}_i|\boldsymbol{z}, \sigma_{\mathrm{N}}^2), \tag{4.3}$$

where $\boldsymbol{z} = (\boldsymbol{\beta}, \boldsymbol{w})^{\mathrm{T}}$ are the random effects from Section 2.3.3. The model structure after applying the SPDE approach is illustrated in Figure 4.2. Here, we follow the

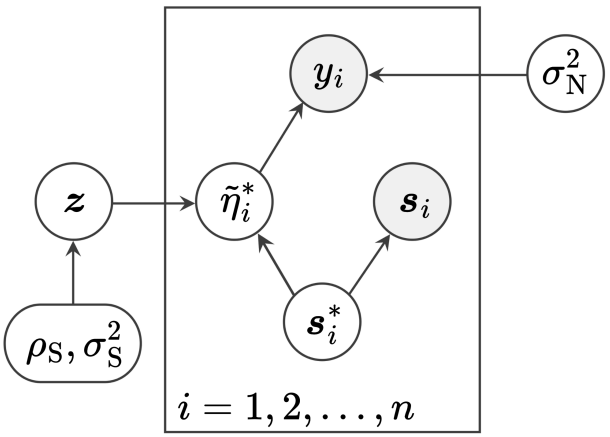


Figure 4.2: Diagram of the Gaussian geostatistical model with positional uncertainty, after the SPDE approach has been applied, where $\tilde{\eta}^*_i = \tilde{\eta}(\boldsymbol{s}^*_i)$. The directed edges show dependence relationships and emphasize the hierarchical structure of the model. Observed quantities are indicated with a grey background.

convention presented by (Cressie and Wikle, 2011, Chapter 2.4), where the directed edges in the figure emphasize the causal relationships and the modeled distributions. For example, $\boldsymbol{s}^*_i$ "causes" $\boldsymbol{s}_i$ in that $\boldsymbol{s}_i$ is a jittered version of $\boldsymbol{s}^*_i$, and we model the distribution $\pi(\boldsymbol{s}_i|\boldsymbol{s}^*_i)$.

4.2 The Numerical Approach to Inference

4.2.1 Numerical Approximation of the Likelihood

To evaluate the likelihood of an individual observation, we express it as an integral over the unknown true locations,

$$\begin{aligned}\pi(y_i, \boldsymbol{s}_i|\eta, \sigma^2_{\mathrm{N}}) &\approx \pi(y_i, \boldsymbol{s}_i|\tilde{\eta}, \sigma^2_{\mathrm{N}}) \\ &= \int_{\mathbb{R}^2} \pi(y_i, \boldsymbol{s}_i, \boldsymbol{s}^*_i|\tilde{\eta}, \sigma^2_{\mathrm{N}}) \mathrm{d}\boldsymbol{s}^*_i \\ &= \int_{\mathbb{R}^2} \pi(y_i|\tilde{\eta}(\boldsymbol{s}^*_i), \sigma^2_{\mathrm{N}}) \pi(\boldsymbol{s}_i|\boldsymbol{s}^*_i) \pi(\boldsymbol{s}^*_i) \mathrm{d}\boldsymbol{s}^*_i,\end{aligned} \tag{4.4}$$

where we have used that $\boldsymbol{s}^*_i$ is independent of $\tilde{\eta}$ and σ^2_{N}, and that y_i is conditionally independent of $\boldsymbol{s}_i$ given $\tilde{\eta}(\boldsymbol{s}^*_i)$ (see Figure 4.2). The integral in (4.4) needs to be approximated numerically. The numerical integration scheme should be applicable to both disk and donut jittering schemes, which are described Section 3.2. This is achieved by dividing the integration area of observation i into a number of $j = 1, \ldots, R^{(i)}$ rings, each with m_{ij} *primary* integration points. The division into rings and primary integration points is illustrated for one observation location in Figure 4.3a. Continuing, the observation likelihood can be approximated as follows,

$$\int_{\mathbb{R}^2} \pi(y_i|\tilde{\eta}(\boldsymbol{s}^*_i), \sigma^2_{\mathrm{N}}) \pi(\boldsymbol{s}_i|\boldsymbol{s}^*_i) \pi(\boldsymbol{s}^*_i) \mathrm{d}\boldsymbol{s}^*_i \approx \sum_{j=1}^{R^{(i)}} \sum_{k=1}^{m_{ij}} \lambda_{ijk} \pi(y_i|\tilde{\eta}(\boldsymbol{s}^*_{ijk}), \sigma^2_{\mathrm{N}}),$$

where λ_{ijk} is the integration weight associated with point k in ring j of observation i, which is given by

$$\lambda_{ijk} \propto \int_{A_{ijk}} \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)\pi(\boldsymbol{s}_i^*)\mathrm{d}\boldsymbol{s}_i^*,$$

where A_{ijk} is the area associated with primary integration point $\boldsymbol{s}_{ijk}^*$. We may assume that $\sum_{ij}\sum_k \lambda_{ijk} = 1$, since the scaling of these weights cancels in the posterior. Additionally, since we take $\pi(\boldsymbol{s}_{ijk}^*)$ to be constant over the support of $\pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)$, we have

$$\lambda_{ijk} \propto \int_{A_{ijk}} \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)\mathrm{d}\boldsymbol{s}_i^*. \tag{4.5}$$

The weight λ_{ijk} is calculated numerically by subdividing A_{ijk} into a 10×10 grid of *secondary* integration regions, each with an associated secondary integration point and weight. This is illustrated in Figure 4.3b. Importantly, if an observation location is restricted to lie within some region, e.g., within the same administrative region as the true location, represented by $A[i]$, the weights become

$$\lambda_{ijk} \propto \int_{A_{ijk}\cap A[i]} \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)\mathrm{d}\boldsymbol{s}_i^*. \tag{4.6}$$

Consequently, if $\pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)$ has support outside of $A[i]$, i.e., observation i could have been displaced outside of $A[i]$ if there were no boundary, the affected weights must be adjusted and all the weights of observation i must be re-normalized.

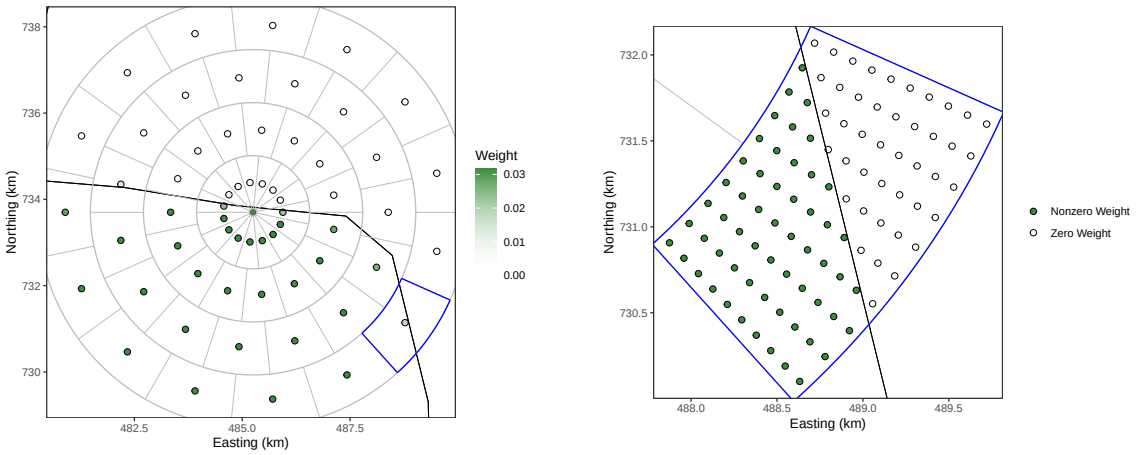
Both disk and donut jittering fit into this numerical integration scheme. A derivation of the integration weights under donut jittering is given in Section B of the Appendices. A corresponding derivation under disk jittering, and technical details on how primary and secondary integration points are generated under both disk and donut jittering can be found in Section E of the Appendices of Altay et al. (2022a). In the continuation, we use $m_{i1} = 1$ and $m_{ij} = 15$, $j > 1$, primary integration points, and $R^{(i)} = 5$ rings $\forall i$ for disk jittering schemes. For donut jittering schemes, we use $m_{ij} = 10$ primary integration points and $R^{(i)} = 5$ rings $\forall i, j$. With this choice the number of primary integration points per integration area is similar across across the donut and disk jittering schemes we compare for a given outer radius, R_2.

The numerical integration scheme enables evaluation of the negative log-likelihood which is used in TMB, which, in the situation with positional uncertainty, is given by

$$f(\boldsymbol{\phi}, \boldsymbol{z}) = -(\log \pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n|\boldsymbol{\phi}, \boldsymbol{z}) + \log \pi(\boldsymbol{z}|\boldsymbol{\phi}) + \log \pi(\boldsymbol{\phi})). \tag{4.7}$$

If the average number of primary integration points per observation is given by M_P and the average number of secondary integration points per observation is given by M_S, the evaluation of 4.7 still only requires $\mathcal{O}(M_\mathrm{P}M_\mathrm{S}n + K^{3/2})$, making computationally efficient inference feasible.

Besides increasing the cost of evaluating (4.7), increasing the number of primary integration points may also contribute to longer running times by resulting in a less sparse Hessian, $\mathbf{H}$, as defined in Section 2.3.3. This is because more integration points may lead to contributions from more of the basis functions $\{\phi_k\}_{k=1}^K$



(a) Primary Integration Points (b) Secondary Integration Points

Figure 4.3: An illustration of the integration scheme for a disk jittered observation location. (a) Primary integration points with corresponding weights and (b) the highlighted region from the left panel with its secondary integration points and corresponding weights. The black line in both panels is the border of an administrative area which the jittered location is restricted to lie within.

in the calculations of the observation likelihoods, $\pi(y_i, \boldsymbol{s}_i|\eta, \sigma_{\mathrm{N}}^2),\ i = 1, \ldots n$, since we potentially select integration points in additional triangles in the mesh, depending on its resolution. In turn, this means that $f(\boldsymbol{\phi}, \boldsymbol{z})$ may depend on more elements of $\boldsymbol{w}$, resulting in a less sparse Hessian.

4.2.2 Empirical Bayes Inference

Here, we describe the inference scheme in the numerical approach. The presentation focuses on the situation with positional uncertainty, but the inference scheme equally applies to the situation without positional uncertainty (or when it is ignored), where the conditioning on the observed locations would be removed.

We use TMB to estimate the model, where the data likelihood is evaluated via the numerical integration scheme described in Section 4.2.1. TMB calculates MMAP estimates of the parameters $\boldsymbol{\phi} = (\log \kappa, \log \tau, \log \sigma_{\mathrm{N}})^{\mathrm{T}}$, as described in Section 2.3.3. These estimates are transformed to create MMAP estimates of $\boldsymbol{\theta} = (\rho_{\mathrm{S}}, \sigma_{\mathrm{S}}^2, \sigma_{\mathrm{N}}^2)^{\mathrm{T}}$, through the relationships given in (2.11). Inference with the random effects is based on $\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \widehat{\boldsymbol{\phi}})$, where $\widehat{\boldsymbol{\phi}}$ contains the MMAP estimates of the parameters. That is, in stead of considering the exact posterior,

$$\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n) = \int \pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \boldsymbol{\phi})\pi(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n)\mathrm{d}\boldsymbol{\phi}, \tag{4.8}$$

we consider $\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \widehat{\boldsymbol{\phi}})$, which essentially amounts to replacing the integration in (4.8) with a maximization. This characterizes empirical Bayes (EB) inference, and means that the uncertainty in the parameters is not propagated to the estimates of the random effects (Carlin and Louis, 2000). Additionally, the approximate posterior of the random effects, $\boldsymbol{z} = (\boldsymbol{\beta}, \boldsymbol{w})^{\mathrm{T}}$, is assumed to be Gaussian with

mean equal to the final modal values of the random effects, $\widehat{\boldsymbol{z}} = (\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{w}})^{\mathrm{T}}$. Specifically, we base inference with the random effects on the approximate posterior given by

$$\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \widehat{\boldsymbol{\phi}} \sim \mathcal{N}_{p+K}\left(\widehat{\boldsymbol{z}},\ \mathbf{H}_{zz}^{-1}\right), \tag{4.9}$$

where $\mathbf{H}_{zz}$ is the random effects sub-matrix of $\mathbf{H}$ evaluated at $\boldsymbol{\phi} = \widehat{\boldsymbol{\phi}}$, $\boldsymbol{z} = \widehat{\boldsymbol{z}}$. The posterior is used to draw samples of the random effects, which are used to construct interval and point estimates for $\boldsymbol{z}$. Point estimates of the random effects are created by taking the median of the samples.

To measure predictive power, we are interested in predicting the latent process at a set of $n_{\mathrm{P}} \in \mathbb{N}^+$ prediction locations $\boldsymbol{l}_1, \ldots \boldsymbol{l}_{n_{\mathrm{P}}} \in \mathcal{D}$. That is, we want to create interval and point estimates of $\boldsymbol{\eta}_{\mathrm{P}} = (\eta(\boldsymbol{l}_1), \ldots, \eta(\boldsymbol{l}_{n_{\mathrm{P}}}))^{\mathrm{T}}$, which has an approximate EB posterior that can be derived from (4.9). Specifically, let $\mathbf{X}_{\mathrm{P}}$ be the $n_{\mathrm{P}} \times p$ covariate matrix and let $\mathbf{A}_{\mathrm{P}}$ be the $n_{\mathrm{P}} \times K$ projection matrix corresponding to the prediction locations. Let $\boldsymbol{\beta}' = \boldsymbol{\beta}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \widehat{\boldsymbol{\phi}}$ and $\boldsymbol{w}' = \boldsymbol{w}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \widehat{\boldsymbol{\phi}}$. Then, the approximate posterior of the discretized latent process is given by

$$\boldsymbol{\eta}_{\mathrm{P}}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \widehat{\boldsymbol{\phi}} \sim \mathcal{N}\left(\mathbf{X}_{\mathrm{P}}\widehat{\boldsymbol{\beta}} + \mathbf{A}_{\mathrm{P}}\widehat{\boldsymbol{w}},\ \mathrm{Cov}(\mathbf{X}_{\mathrm{P}}\boldsymbol{\beta}' + \mathbf{A}_{\mathrm{P}}\boldsymbol{w}')\right).$$

In practice, we draw samples from (4.9) and create point predictions and credible intervals of the latent process by respectively calculating the mean and quantiles of the samples of $\mathbf{X}_{\mathrm{P}}\boldsymbol{\beta}' + \mathbf{A}_{\mathrm{P}}\boldsymbol{w}'$.

4.3 The Sample-based Approach to Inference

Here, we present the sample-based approach to inference under positional uncertainty in terms of disk and donut jittering. The approach uses an MCMC algorithm to sample from the posterior of the model described in Section 4.1. The distribution of the samples from the MCMC algorithm will in theory converge to the exact posterior of the model. Thus, although computational efficiency may be an issue for the sample-based approach, it can still be used to investigate the inference of the numerical approach.

4.3.1 Derivation of Parameter Posterior

To construct the MCMC algorithm, we first define some new notation. Let $\boldsymbol{\theta}^* = (\rho_{\mathrm{S}}, \sigma_{\mathrm{S}}^2, \sigma_{\mathrm{N}}^2, \boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_n^*)^{\mathrm{T}}$. Then, the Gaussian geostatistical model with positional uncertainty may be presented as

$$\begin{aligned} \boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\theta}^* &\sim \mathcal{N}_n(\mathbf{S}\boldsymbol{z}, \mathbf{I}_n/\tau_{\mathrm{N}}), \\ \boldsymbol{z}|\boldsymbol{\theta} &\sim \mathcal{N}_{p+K}(\mathbf{0}, \mathbf{Q}_z^{-1}), \end{aligned}$$

where $\tau_{\mathrm{N}} = \sigma_{\mathrm{N}}^{-2}$ and $\mathbf{S} = [\mathbf{X}\ \mathbf{A}]$ is an $n \times (p+K)$ matrix which depends on the true locations.

Now, our strategy is to construct a collapsed Gibbs sampler (Liu, 1994), where we sample from the posterior of $\boldsymbol{\theta}^*$ in each iteration, and, once the chain has converged, sample the random effects, $\boldsymbol{z}$, from $\pi(\boldsymbol{z}|\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$. Thus, we need

to derive an expression for the parameter posterior, $\pi(\boldsymbol{\theta}^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)$, at least up to a normalizing constant. We divide the derivation of this into three steps.

Step 1: We derive the density of the full conditional distribution of the random effects,

$$\begin{aligned}\pi(\boldsymbol{z}|\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) = \pi(\boldsymbol{z}|\boldsymbol{\theta}^*, \boldsymbol{y}) &\propto \pi(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\theta}^*)\pi(\boldsymbol{z}|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*) \\ &\propto \exp\left\{-\frac{\tau_{\mathrm{N}}}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{z})^{\mathrm{T}}\mathbf{I}_n(\boldsymbol{y} - \mathbf{S}\boldsymbol{z}) - \frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\mathbf{Q}_z\boldsymbol{z}\right\} \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{C}})^{\mathrm{T}}\mathbf{Q}_{\mathrm{C}}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{C}})\right\}, \end{aligned} \tag{4.10}$$

where

$$\begin{aligned}&\mathbf{Q}_{\mathrm{C}} = \mathbf{Q}_z + \mathbf{S}^{\mathrm{T}}\mathbf{S} \cdot \tau_{\mathrm{N}}, \quad \text{and} \\ &\boldsymbol{\mu}_{\mathrm{C}} = \mathbf{Q}_{\mathrm{C}}^{-1}\mathbf{S}^{\mathrm{T}}\boldsymbol{y}.\end{aligned}$$

In the first transition of (4.10), we use the fact that the random effects are conditionally independent of the observed locations, given the true locations and the observations. We recognize the last line in (4.10) as a Gaussian distribution, i.e.,

$$\boldsymbol{z}|\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \sim \mathcal{N}_{K+p}(\boldsymbol{\mu}_{\mathrm{C}}, \mathbf{Q}_{\mathrm{C}}^{-1}). \tag{4.11}$$

Step 2: From the definition of conditional probability, we can write

$$\pi(\boldsymbol{z}|\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) = \frac{\pi(\boldsymbol{z}, \boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)}{\pi(\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)}.$$

Thus, we can integrate $\boldsymbol{z}$ out from the full joint distribution as follows,

$$\begin{aligned}\pi(\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) &= \frac{\pi(\boldsymbol{z}, \boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)}{\pi(\boldsymbol{z}, |\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)} \\ &= \frac{\pi(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\theta}^*)\pi(\boldsymbol{z}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n|\boldsymbol{\theta}^*)\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{z}|\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)},\end{aligned} \tag{4.12}$$

where we in the second transition use that the observations, $\boldsymbol{y}$, are conditionally independent of the observed locations given the true locations. Continuing, we assume the observed locations to be conditionally independent of the coefficients, $\boldsymbol{z}$, given the true locations, i.e., we assume

$$\pi(\boldsymbol{z}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n|\boldsymbol{\theta}^*) = \pi(\boldsymbol{z}|\boldsymbol{\theta}^*)\pi(\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n|\boldsymbol{\theta}^*).$$

Furthermore, there is no dependence between the true locations alone and $\boldsymbol{z}$, and the parameters $\boldsymbol{\theta}$ give no information about the observed locations. Thus, we can write Equation (4.12) as

$$\pi(\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) = \frac{\pi(\boldsymbol{y}|\boldsymbol{z}, \boldsymbol{\theta}^*)\pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}^*)}{\pi(\boldsymbol{z}, |\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n)} \cdot \prod_{i=1}^{n} \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*). \tag{4.13}$$

Using the result of the previous step, we can write (4.13) as

$$\begin{aligned}\pi(\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) \propto \frac{|\tau_{\mathrm{N}}\mathbf{I}_n|^{\frac{1}{2}}|\mathbf{Q}_z|^{\frac{1}{2}}}{|\mathbf{Q}_{\mathrm{C}}|^{\frac{1}{2}}} \exp\left\{-\frac{\tau_{\mathrm{N}}}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{z})^{\mathrm{T}}\mathbf{I}_n(\boldsymbol{y} - \mathbf{S}\boldsymbol{z})\right\} \\ \times \exp\left\{\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{C}})^{\mathrm{T}}\mathbf{Q}_{\mathrm{C}}(\boldsymbol{z} - \boldsymbol{\mu}_{\mathrm{C}})\right\} \\ \times \exp\left\{-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}\mathbf{Q}_z\boldsymbol{z}\right\} \cdot \pi(\boldsymbol{\theta}^*) \cdot \prod_{i=1}^{n} \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*).\end{aligned}$$

Step 3: Finally, we note that the left hand side in the first line of Equation (4.12) does not depend on $\boldsymbol{z}$. Thus, we can use any suitable value for $\boldsymbol{z}$, e.g., $\boldsymbol{z} = \boldsymbol{\mu}_{\mathrm{C}}$. With this choice, we know the posterior distribution of the parameters up to a normalizing constant,

$$\begin{aligned}\pi(\boldsymbol{\theta}^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) &\propto \pi(\boldsymbol{\theta}^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n) \\ &\propto \frac{\tau_{\mathrm{N}}^{\frac{n}{2}}|\mathbf{Q}_z|^{\frac{1}{2}}}{|\mathbf{Q}_{\mathrm{C}}|^{\frac{1}{2}}} \exp\left\{-\frac{\tau_{\mathrm{N}}}{2}(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_{\mathrm{C}})^{\mathrm{T}}\mathbf{I}_n(\boldsymbol{y} - \mathbf{S}\boldsymbol{\mu}_{\mathrm{S}})\right\} \\ &\quad \times \exp\left\{-\frac{1}{2}\boldsymbol{\mu}_{\mathrm{C}}^{\mathrm{T}}\mathbf{Q}_z\boldsymbol{\mu}_{\mathrm{C}}\right\} \cdot \pi(\boldsymbol{\theta}^*) \cdot \prod_{i=1}^{n} \pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*).\end{aligned}$$

4.3.2 MCMC Algorithm

Now that we have determined the posterior density of $\boldsymbol{\theta}^*$ up to a normalizing constant, we can construct an MCMC algorithm to sample from the posterior distribution of the model. Just as within TMB, we choose to work with $\boldsymbol{\phi} = (\log\kappa, \log\tau, \log\sigma_{\mathrm{N}})^{\mathrm{T}}$ in stead of $\boldsymbol{\theta} = (\rho_{\mathrm{S}}, \sigma_{\mathrm{S}}^2, \sigma_{\mathrm{N}}^2)^{\mathrm{T}}$, since priors are specified in terms of $\boldsymbol{\phi}$ and because we avoid constraints on the proposals in the sampler. The relationship between $\boldsymbol{\phi}$ and $\boldsymbol{\theta}$ are given in (2.11).

The algorithm consists of two Metropolis-within-Gibbs steps (Møller, 2003). First, new values for $\boldsymbol{\phi}$ are proposed and either accepted or rejected in an MH step. Then, the same is done for the true locations, $\boldsymbol{s}_1^*, \ldots \boldsymbol{s}_n^*$. At a regular thinning interval of $t > 0$ iterations, the conditional distribution of the random effects, $\boldsymbol{z}$, given the data and the current parameters is sampled from the exact distribution given in (4.11). The MCMC algorithm is outlined in Algorithm 2. When we apply the algorithm in a simulation setting, we use the parameter values used to simulate the data as initial values. In the following, we describe the proposal distributions used in the MH steps.

Sampling of $\rho_{\mathrm{S}}, \sigma_{\mathrm{S}}^2$ and σ_{N}^2

Let b denote the current iteration of the MCMC algorithm. Then, the acceptance probability of the new proposal for $\boldsymbol{\phi}$, $\boldsymbol{\phi}^{(b)}$, becomes

Algorithm 2 The MCMC algorithm used in the sample-based approach.

Assign integer values to B, the number of iterations, and t, the thinning interval
Assign initial values $\boldsymbol{\phi}^{(0)}, \boldsymbol{s}_1^{*\,(0)}, \ldots, \boldsymbol{s}_n^{*\,(0)}$
for $b = 1, \ldots, B$ **do**
  Sample $\boldsymbol{\phi}^{(b)}$ from $\pi(\boldsymbol{\phi}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \boldsymbol{s}_1^{*\,(b-1)}, \ldots, \boldsymbol{s}_n^{*\,(b-1)})$ in an MH step
  Sample $\boldsymbol{s}_1^{*\,(b)}, \ldots, \boldsymbol{s}_n^{*\,(b)}$ from $\pi(\boldsymbol{s}_1^*, \ldots \boldsymbol{s}_n^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \boldsymbol{\theta}^{(b)})$ in an MH step
  if $b \mod t = 0$ **then**
    Sample coefficients from $\pi(\boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \boldsymbol{\theta}^{*\,(b)})$
  end if
end for

$$\alpha(\boldsymbol{\phi}^{(b)}, \boldsymbol{\phi}^{(b-1)}) = \min\left\{1, \frac{\pi(\boldsymbol{\phi}^{(b)}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \boldsymbol{s}_1^{*\,(b-1)}, \ldots, \boldsymbol{s}_n^{*\,(b-1)}) \cdot q(\boldsymbol{\phi}^{(b-1)}|\boldsymbol{\phi}^{(b)})}{\pi(\boldsymbol{\phi}^{(b-1)}|\boldsymbol{y}, \boldsymbol{s}_1, \ldots \boldsymbol{s}_n, \boldsymbol{s}_1^{*\,(b-1)}, \ldots, \boldsymbol{s}_n^{*\,(b-1)}) \cdot q(\boldsymbol{\phi}^{(b)}|\boldsymbol{\phi}^{(b-1)})}\right\},$$

where $q(\boldsymbol{\phi}^{(b)}|\boldsymbol{\phi}^{(b-1)})$ is the density of the proposal distribution. We use random walk proposals for $\boldsymbol{\phi}$, i.e.,

$$\boldsymbol{\phi}^{(b)}|\boldsymbol{\phi}^{(b-1)} \sim \mathcal{N}_3(\boldsymbol{\phi}^{(b-1)}, \text{diag}(\boldsymbol{v})),$$

where $\boldsymbol{v}$ is a vector containing the marginal variances for each of the parameter proposals.

Sampling of $\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_n^*$

To construct a proposal distribution for the true locations, we use a *simplified* observation model,

$$y_i^{\text{S}} = \boldsymbol{x}(\boldsymbol{s}_i^*)^{\text{T}}\boldsymbol{b} + u_i + \varepsilon_i^{\text{S}}, \quad i = 1, \ldots, n,$$

where $\varepsilon_1^{\text{S}}, \ldots \varepsilon_n^{\text{S}} \mid \sigma_{\text{NS}}^2 \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_{\text{NS}}^2)$. Here, $\boldsymbol{b}$ plays the role of $\boldsymbol{\beta}$ and u_i plays the role of $u(\boldsymbol{s}_i^*)$, but they are treated as known constants. This implies that the observations of this simplified model are conditionally independent, given σ_{NS}^2. Let $\boldsymbol{y}^{\text{S}} = (y_1^{\text{S}}, \ldots, y_n^{\text{S}})^{\text{T}}$. The posterior of each true location in the simplified model is

$$\begin{aligned}\pi(\boldsymbol{s}_i^*|\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{y}^{\text{S}}, \sigma_{\text{NS}}^2) &\propto \pi(y_i^{\text{S}}|\boldsymbol{s}_i^*, \sigma_{\text{NS}}^2)\pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)\pi(\boldsymbol{s}_i^*) \\ &\propto \exp\left(-\frac{1}{2\sigma_{\text{NS}}^2}(y_i - \mu_i)^2\right) \cdot \frac{\mathbb{I}(R_1 \leq r_i \leq R_2)}{r_i}, \qquad (4.14)\end{aligned}$$

where $\mu_i = \boldsymbol{x}(\boldsymbol{s}_i^*)^{\text{T}}\boldsymbol{b} + u_i$, $r_i = d(\boldsymbol{s}_i, \boldsymbol{s}_i^*)$ is the euclidean distance between $\boldsymbol{s}_i$ and $\boldsymbol{s}_i^*$ in appropriate units, and R_1 and R_2 are the inner and outer radius of the jittering scheme, respectively. We may also restrict the jittered locations to lie within some administrative region, but we will not study that situation with the sample-based approach.

We sample from the density in (4.14) with rejection sampling, and use it to generate proposals for the unknown true locations. That is, the true locations are independently proposed according to $q(\boldsymbol{s}_i^{*(b)}|\boldsymbol{s}_i^{*(b-1)}) = \pi(\boldsymbol{s}_i^*|\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{y}^{\text{S}}, \sigma_{\text{NS}}^2)$. In practice, we set $\boldsymbol{b} = \boldsymbol{\beta}$, $u_i = u(\boldsymbol{s}_i^*)$ and $\sigma_{\text{NS}}^2 = \sigma_{\text{N}}^2$, since we have access to these values in a simulation setting.

Validation of Algorithm Implementation

To validate the implementation of the MCMC algorithm, we apply it to the standard geostatistical model in (2.6) when there is no positional uncertainty, and compare its sample-distributions to the posteriors created with INLA for the same model. The results are illustrated in Section C of the Appendices, and give us confidence that the MCMC algorithm is correctly implemented.

4.3.3 Sample-based Inference

Inference with the sample-based approach is based on the parameter and random effect samples generated by the MCMC algorithm, which are approximately distributed according to the posteriors

$$\boldsymbol{\theta}^*|\boldsymbol{y}, \boldsymbol{s}_1, \dots \boldsymbol{s}_n \quad \text{and} \quad \boldsymbol{z}|\boldsymbol{y}, \boldsymbol{s}_1, \dots, \boldsymbol{s}_n, \boldsymbol{\theta}^*,$$

respectively, assuming that the algorithm has converged. To assess convergence, we inspect trace-plots. Based, on the trace-plots, we discard the first $n_{\text{B}} > 0$ samples as *burn-in* samples. After discarding the burn-in samples, we create point estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\beta}$, by taking the median of their generated posterior samples, minimizing the absolute error.

As described in Section 4.2.2, we evaluate predictive power by predicting the latent process at a set of locations. In the sample-based approach, samples of the posterior latent process at these locations are constructed from the posterior samples of the random effects (after removing burn-in samples). Point predictions of the latent process are acquired by taking the mean of samples, minimizing the mean squared error.

Chapter 5

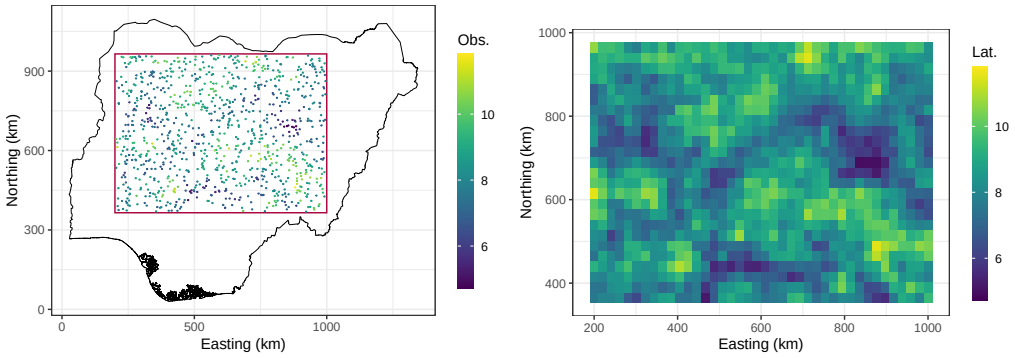
Investigating the Accuracy of the Numerical Approach

Chapter 4 presented two distinct approaches for inference in the presence of positional uncertainty: the numerical approach and the sample-based approach. Upon implementation, it became evident that the sample-based approach is considerably more computationally demanding than the numerical approach, and practically constrained to scenarios with limited impact of jittering. However, it can be used to investigate the accuracy of the numerical approach in some selected scenarios, which is the subject of this chapter. We begin, in Section 5.1, with a description of the scenarios we have selected for the investigation. Results are presented in Section 5.2. Finally, Section 5.3 gives a discussion of the sample-based approach in relation the investigation and the path forward.

5.1 Scenario Descriptions

We consider two scenarios, *Scenario 1* and *Scenario 2*. First, we describe the setup which is shared in both scenarios. We simulate $n = 1000$ observations from the model described in Section 4.1 on a rectangular spatial domain, $\mathcal{D}$, measuring 800 km $\times$ 600 km. The true locations are sampled uniformly on $\mathcal{D}$, before jittering is applied. The rectangle, $\mathcal{D}$, is considered to be cut out of Nigeria, as illustrated in Figure 5.1a, so that we can use existing rasters for Nigeria to construct spatial covariates. The spatial covariates we consider are gathered from WorldPop et al. (2018) and are min-max scaled to be between 0 and 1. Specifically, we consider four covariates that are respectively related to rasters of population count[1] (PopC), terrain elevation [2] (Elev), distance to inland water[3] (DistW) and distance to major road intersections[4] (DistR). They are displayed in Figure 6.1 along with histograms of their pixel values. A further description of the covariates are given in Table 6.1. In this chapter, we only consider models with covariates Elev, DistW and DistR. The reason why PopC is excluded, is that its inclusion made adequate sampling of the true locations infeasible. This is likely because of the covariate's highly non-smooth behaviour, which may give posteriors for the unknown true locations that

[1]https://hub.worldpop.org/geodata/summary?id=6409 (Downloaded: 10-27-2022)
[2]https://hub.worldpop.org/geodata/summary?id=23424 (Downloaded: 10-27-2022)
[3]https://hub.worldpop.org/geodata/summary?id=24171 (Downloaded: 10-27-2022)
[4]https://hub.worldpop.org/geodata/summary?id=17706 (Downloaded: 03-14-2023)



(a) Observations (b) Latent process

Figure 5.1: A set of observations (Obs.) over the domain (red rectangle) is displayed in (a), while the underlying latent process which we want to predict (Lat.) is displayed in (b).

are strongly peaked and perhaps multimodal, making efficient sampling difficult. Thus, we consider a vector of covariate coefficients given by

$$\boldsymbol{\beta} = (\mu, \beta_{\text{Elev}}, \beta_{\text{DistW}}, \beta_{\text{DistR}})^{\text{T}},$$

where μ is the intercept and the covariate corresponding to each coefficient is given in the subscript. We consider equally strong covariates, i.e., $\boldsymbol{\beta} = \beta(1, 1, 1, 1)^{\text{T}}$. Furthermore, we note that the covariate rasters are represented in the World Geodetic System (WGS84), and to represent locations in Nigeria, e.g., in the domain $\mathcal{D}$, we use the "Minna / Nigeria Mid Belt" projected coordinate system (EPSG, n.d.) throughout the text.

The spatial marginal variance and the variance of the nugget effect are fixed such that $\sigma_{\text{S}}^2 = 1$ and $\sigma_{\text{N}}^2 = 0.1$. That is, we study a situation in which the contribution of the spatial field to the response is most likely in the range ± 2, and the variance of the measurement error is 10 % of the marginal variance of the spatial field. In both scenarios we take $\beta = 4$, which gives covariates that account for 37 % of the variation in the data on average. This average variation percentage is calculated by simulating 20 sets of locations uniformly over $\mathcal{D}$ and evaluating $S^2/(S^2 + \sigma_{\text{S}}^2 + \sigma_{\text{N}}^2)$, where S^2 is the empirical variance of $\mathbf{X}\boldsymbol{\beta}$, given by

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left[\boldsymbol{x}(\boldsymbol{s}_i^*)^{\text{T}} \boldsymbol{\beta} - \frac{1}{n} \sum_{j=1}^{n} \boldsymbol{x}(\boldsymbol{s}_j^*)^{\text{T}} \boldsymbol{\beta} \right]^2.$$

The spatial range and the jittering scheme differ in the two scenarios. In Scenario 1, we consider a spatial range of $\rho_{\text{S}} = 160$ (km) and a disk jittering scheme with $R_1 = 0$ (km) and $R_2 = 2$ (km). In Scenario 2, we consider a spatial range of $\rho_{\text{S}} = 240$ and a donut jittering scheme with $R_1 = 1$ and $R_2 = 2$. The scenarios are summarized in Table 5.1.

The range was set to $\rho_{\text{S}} = 160$ in Scenario 1, because this corresponds to 20% of the domain width. However, in Scenario 2, this choice of range made the true locations too computationally intensive to sample. Therefore, we increased the spatial range to 30% of the domain width, setting $\rho_{\text{S}} = 240$, which yielded an

acceptance probability and an estimated *effective sample size* (ESS) for the true locations comparable to that in Scenario 1. The reason why we needed to increase the range is most likely because it reduces the impact of the jittering (see Section 3.3), making the true locations easier to sample.

Scenario	(R_1, R_2)	β	ρ_{S}	σ^2_{S}	σ^2_{N}
1	$(0, 2)$	4	160	1	0.1
2	$(1, 2)$	4	240	1	0.1

Table 5.1: The two scenarios for which we compare the numerical and sample-based approach.

In both scenarios, we use a PC prior on the spatial field, with $\mathrm{P}(\rho_{\mathrm{S}} < 160) = 0.5$ and $\mathrm{P}(\sigma_{\mathrm{S}} > 1) = 0.5$. That is, we place the median of the range on 20% of the domain width, while we place the median of the spatial marginal standard deviation on 1. This amounts to an *a priori* expectation that the spatial field's contribution to the response is in the range ± 2. The measurement error variance is also equipped with a PC prior, $\mathrm{P}(\sigma_{\mathrm{N}} > 0.986) = 0.01$. This choice makes the marginal standard deviation of the measurement errors approximately equal to 0.3 Simpson et al. (2017). That is, we *a priori* expect the measurement error to exhibit a marginal variance that is approximately 10% of the marginal variance of the spatial field. Furthermore, for the covariate coefficients we use the prior $\boldsymbol{\beta} \sim \mathcal{N}_p(\mathbf{0}, 25\mathbf{I}_p)$, since we *a priori* expect the covariates, which are min-max scaled, to contribute in the range of ± 10 to the response. We note that the true parameter values are known *a priori* in this simulation setting. However, it should be possible to justify priors in a similar manner in a real application, based on domain knowledge.

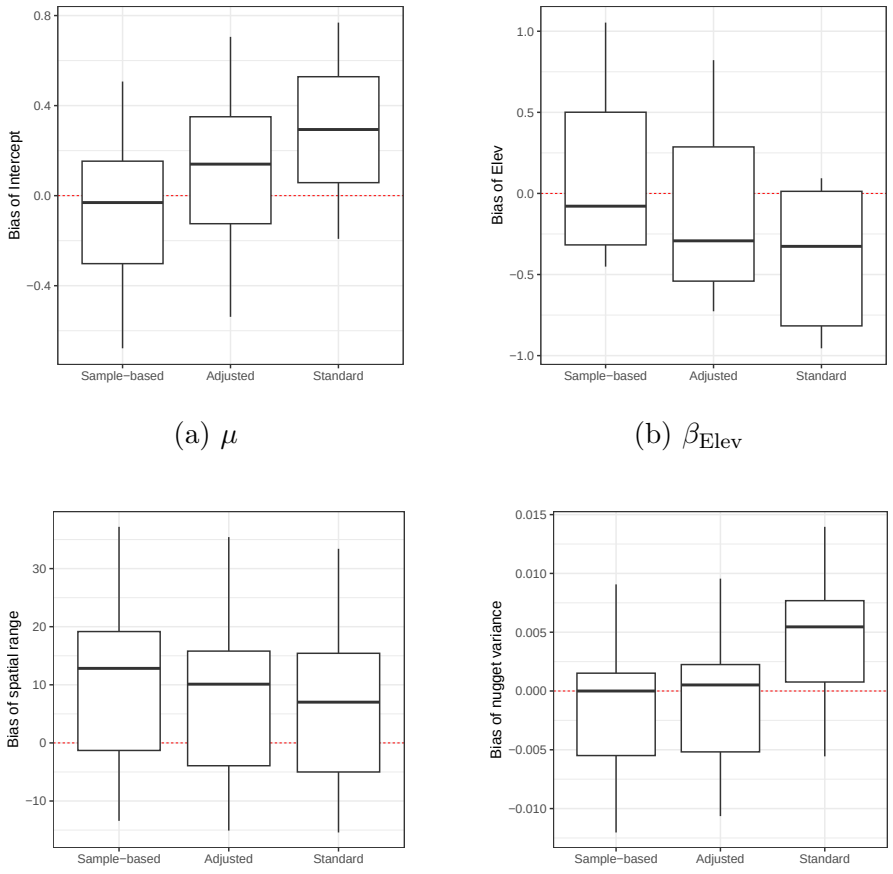
Finally, we underscore that the selection of scenarios is primarily driven by the feasibility of constructing an efficient sampler for the true locations. In other words, our choice is largely influenced by the limitations of the MCMC algorithm employed in the sample-based approach, where sampling the true locations poses the primary challenge. As we increase either the covariate signal or the outer jittering radius, or decrease the spatial range in the described scenarios, the sampling of true locations becomes exceedingly difficult. Additionally, we found that the choice of $\boldsymbol{v}$ in the random walk proposals for the parameters had little impact on the sampling of true locations, and after some testing to maximize the estimated ESS, we ultimately set $\boldsymbol{v} = (0.25^2, 0.10^2, 0.05^2)^{\mathrm{T}}$ in both scenarios.

5.2 Results

In the following, we consider three estimated models. The numerical approach is used to estimate an *adjusted* model, denoted by $\mathbf{M}_{\mathrm{Adj.}}$, which adjusts for jittering, and a *standard* model, denoted by $\mathbf{M}_{\mathrm{Std.}}$, which ignores jittering. Additionally, we estimate a model with the sample-based approach, which comprises the posterior samples generated with the MCMC algorithm and any predictions or point estimates we derive from these samples. The sample-based model is denoted by $\mathbf{M}_{\mathrm{SB}}$.

Parameter	$\log \kappa$	$\log \tau$	$\log \sigma_{\mathrm{N}}$	$\boldsymbol{s}^*$
ESS	1002	871	912	158

Table 5.2: The mean ESS of parameters sampled by the MCMC algorithm across datasets for Scenario 1. The column denoted $\boldsymbol{s}^*$ refers to the mean ESS over both the first and second coordinate over all the true locations.



(a) μ (b) β_{Elev}

(c) ρ_{S} (d) σ^2_{N}

Figure 5.2: Cross-dataset bias of point estimates of (a) μ, (b) β_{Elev}, (c) the spatial range and (d) the nugget variance for $\mathbf{M}_{\mathrm{Std.}}$, $\mathbf{M}_{\mathrm{Adj.}}$ and $\mathbf{M}_{\mathrm{SB}}$ in Scenario 1.

5.2.1 Scenario 1

We simulated 10 datasets, for which $\mathbf{M}_{\text{Adj.}}$, $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{SB}}$ were estimated. For each dataset, the MCMC-algorithm was executed with $20,000$ iterations, which took around 21 hours to complete on average, where the main computational burden came from the rejection sampling of the true locations. Samples from the first 1000 iterations were discarded as burn-in samples, based on inspection of trace-plots, some of which are displayed in Figure D.1 and discussed in Section D.1 in the Appendices. The acceptance probability of the parameters was on average 19.1%, while the acceptance probability of the true locations was on average 2.9%. The estimated ESSs of the parameters and the true locations are summarized in Table 5.2, which were calculated with the `effectiveSize` function from the `coda` R-package (Plummer et al., 2006).

The biases across datasets for the estimates of μ, β_{Elev}, ρ_{S} and σ^2_{N} are displayed for $\mathbf{M}_{\text{Adj.}}$, $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{SB}}$ in Figure 5.2. For μ and β_{Elev}, $\mathbf{M}_{\text{SB}}$ exhibits the least biased estimates across datasets, while $\mathbf{M}_{\text{Adj.}}$ has less biased estimates than $\mathbf{M}_{\text{Std.}}$. There is little difference between the models with respect to the biases of β_{DistW} and β_{DistR}, which are displayed in Figure D.2. For the estimated nugget variance, the cross-dataset bias is visibly lower for $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{SB}}$, compared to $\mathbf{M}_{\text{Std.}}$. The same can not be said for the spatial parameters, where we observe little difference between the biases corresponding to ρ_{S} and σ^2_{S}, where the cross-dataset bias of the latter is displayed in Figure D.2.

To measure predictive performance, we predict the latent process at set of regularly spaced points over $\mathcal{D}$, illustrated as a square grid in Figure 5.1b. Specifically, we treat $\mathbf{M}_{\text{SB}}$ as a base-line, and consider the pair-wise relative difference in CRPS of $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$, relative to $\mathbf{M}_{\text{SB}}$. That is, we consider

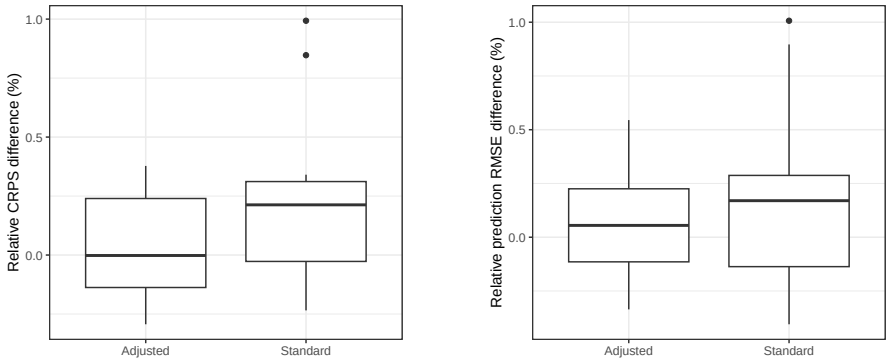
$$100\% \times \frac{\text{CRPS}_{\text{Adj.}} - \text{CRPS}_{\text{SB}}}{\text{CRPS}_{\text{SB}}}$$

for $\mathbf{M}_{\text{Adj.}}$ and likewise for $\mathbf{M}_{\text{Std.}}$. The variation in this CRPS-measure across datasets is displayed in Figure 5.3a, from which we observe a similar spread for both models (except for two extreme values for $\mathbf{M}_{\text{Std.}}$) and superior performance for $\mathbf{M}_{\text{Adj.}}$ with respect to the median. We similarly consider the the pair-wise relative difference in prediction RMSE given by

$$100\% \times \frac{\text{RMSE}_{\text{Adj.}} - \text{RMSE}_{\text{SB}}}{\text{RMSE}_{\text{SB}}}$$

for $\mathbf{M}_{\text{Adj.}}$ and likewise for $\mathbf{M}_{\text{Std.}}$. The variation in this RMSE-measure across datasets is displayed in Figure 5.3b, from which we observe a slightly larger spread for $\mathbf{M}_{\text{Std.}}$ and superior performance for $\mathbf{M}_{\text{Adj.}}$ with respect to the median.

To explicitly illustrate how $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ compare to $\mathbf{M}_{\text{SB}}$, we have selected one of the 10 simulated datasets, for which we display some qualitative plots. Figure 5.4 displays the histograms of spatial range, spatial variance and nugget variance produced with the sample-based approach, in addition to the true parameter values and the MMAP estimates of $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$. The figure is illustrative of the cross-dataset results in that, for the spatial parameters, $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ produce similar estimates, while for the variance of the nugget effect we see that $\mathbf{M}_{\text{Adj.}}$ is better at "finding" the mode of the distribution, thereby giving less biased estimates on average.



(a) CRPS (b) RMSE

Figure 5.3: (a) Pair-wise relative differences in CRPS and (b) Pair-wise relative differences in prediction RMSE across datasets for $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ relative to $\mathbf{M}_{\text{SB}}$ in Scenario 1.

For the same dataset, Figure 5.5 displays the posterior distribution of μ and β_{Elev} in $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$, along with corresponding sample-histograms from the MCMC algorithm. The figure illustrates that $\mathbf{M}_{\text{Adj.}}$ may give posterior distributions that are more in-line with that of $\mathbf{M}_{\text{SB}}$, whose posteriors converge to the exact posteriors in theory. Specifically, this indicates that the posteriors of $\mathbf{M}_{\text{Adj.}}$ are more accurate in terms of mean and variance compared to the posteriors of $\mathbf{M}_{\text{Std.}}$. A corresponding display for β_{DistW} and β_{DistR} is given in Figure D.3, for which the models give very similar results.

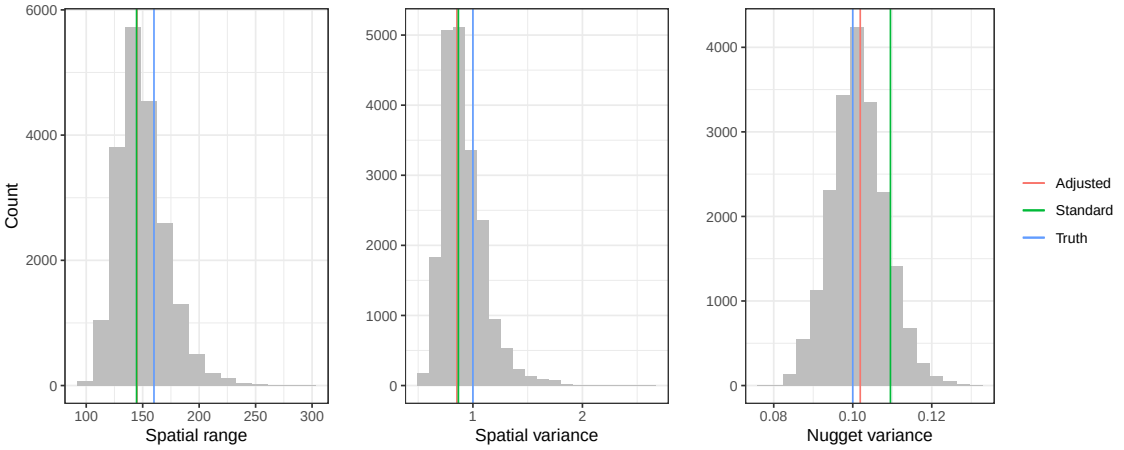
5.2.2 Scenario 2

We simulated 10 datasets, for which $\mathbf{M}_{\text{Adj.}}$, $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{SB}}$ were estimated. For each dataset, the MCMC-algorithm was executed with $20,000$ iterations, which took around 20 hours to complete on average. Samples from the first 1000 iterations were discarded as burn-in samples, based on inspection of trace-plots, some of which are displayed in Figure D.4 and discussed in Section D.2 in the Appendices. The average acceptance probability for proposals of $\boldsymbol{\theta}$ was 23.2%, while the average acceptance probability of the true location proposals was 1.8%. The estimated ESSs of the parameters and the true locations are summarized in Table 5.3.

Parameter	$\log \kappa$	$\log \tau$	$\log \sigma_{\text{N}}$	$\boldsymbol{s}^*$
ESS	930	931	962	105

Table 5.3: The mean ESS of parameters sampled by the MCMC algorithm across datasets for Scenario 2. The column denoted $\boldsymbol{s}^*$ refers to the mean ESS over both the first and second coordinate over all the true locations.

In Figure 5.6, the cross-dataset biases for estimates of μ, β_{Elev}, ρ_{S} and σ^2_{N} are displayed in a boxplot, showing the same pattern as in Scenario 1, namely that the two covariate coefficients and the nugget variance are estimated with less bias in $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{SB}}$ compared to in $\mathbf{M}_{\text{Std.}}$. For the spatial parameters, the models



(a) ρ_S (b) σ_S^2 (c) σ_N^2

Figure 5.4: Sample-histograms of $\mathbf{M}_{SB}$ for (a) the spatial range (b) the spatial variance and (c) the variance of the nugget effect, for a given dataset in Scenario 1. The true parameter values are given in blue, while the estimates of $\mathbf{M}_{Std.}$ and $\mathbf{M}_{Adj.}$ are given in green and red, respectively.

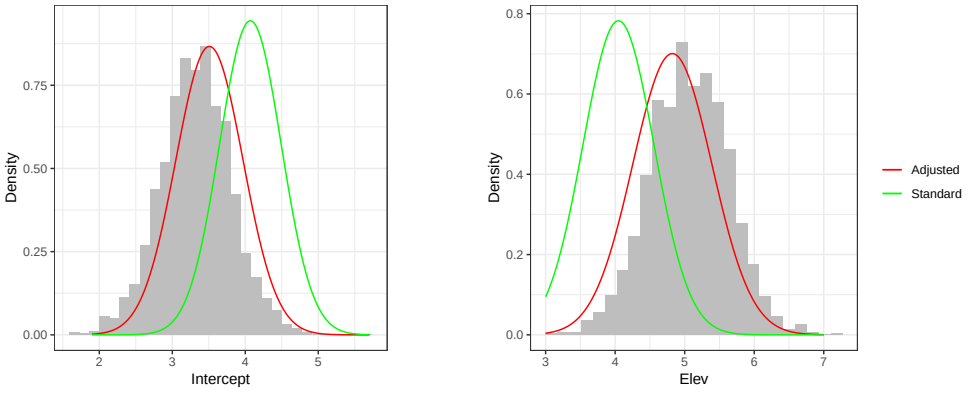
perform similarly across datasets. The bias across datasets corresponding to β_{DistR}, β_{DistW} and σ_S^2 displayed in Figure D.5, which shows similar results for all the models.

Figure 5.7a shows the pair-wise relative CRPS across datasets of $\mathbf{M}_{Adj.}$ and $\mathbf{M}_{Std.}$, relative to $\mathbf{M}_{SB}$. For this CRPS-measure, we observe superior performance for $\mathbf{M}_{Adj.}$ compared to $\mathbf{M}_{Std.}$, which clearly has higher scores on average. Figure 5.7b shows the pair-wise relative prediction RMSE of $\mathbf{M}_{Adj.}$ and $\mathbf{M}_{Std.}$, relative to $\mathbf{M}_{SB}$, which also shows improved performance for $\mathbf{M}_{Adj.}$ over $\mathbf{M}_{Std.}$.

Just as for Scenario 1, we have chosen one of the ten datasets for which we present some qualitative plots, to explicitly illustrate how $\mathbf{M}_{Adj.}$ and $\mathbf{M}_{Std.}$ compare to $\mathbf{M}_{SB}$. Figure 5.8 displays the histograms of the MCMC samples for ρ_S, σ_S^2 and σ_N^2 along with vertical lines showing the true parameter values and MMAP estimates of $\mathbf{M}_{Std.}$ and $\mathbf{M}_{Adj.}$. We observe that the point estimates of $\mathbf{M}_{Std.}$ and $\mathbf{M}_{Adj.}$ are similar for ρ_S and σ_S^2, while they differ more for σ_N^2, where the $\mathbf{M}_{Adj.}$ is closer to the mode of the distribution of MCMC-samples. Figure 5.9 displays the posterior distributions of μ and β_{Elev} for $\mathbf{M}_{Std.}$, $\mathbf{M}_{Adj.}$ and $\mathbf{M}_{SB}$, where we see that the $\mathbf{M}_{Adj.}$ gives posteriors that are closer to that of $\mathbf{M}_{SB}$, which should resemble the exact posteriors. Corresponding plots for β_{DistW} and β_{DistR} are given in Figure D.6, for which the models give very similar posteriors.

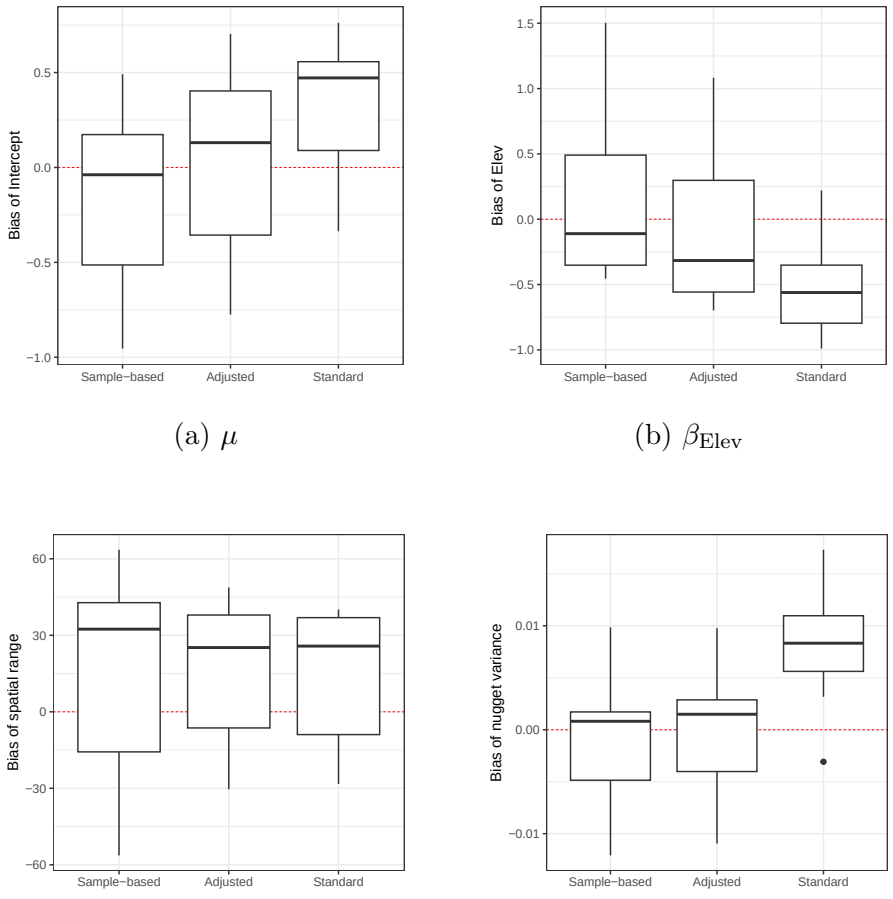
5.3 Discussion of Results

In this chapter, we have compared parameter estimates, posteriors and predictive measures of $\mathbf{M}_{Std.}$, $\mathbf{M}_{Adj.}$ and $\mathbf{M}_{SB}$. Due to the difficulty of implementing an efficient sampler when jittering is present, we restricted our comparison to two scenarios with moderately strong covariates and a modest jittering scale. Bias-plots across all datasets show that there is little difference between $\mathbf{M}_{Adj.}$ and $\mathbf{M}_{Std.}$ with regard to the spatial parameters, while for the nugget variance and two of the covariate coefficients, $\mathbf{M}_{Adj.}$ has visibly lower bias compared to $\mathbf{M}_{Std.}$. In terms of the predictive



(a) μ (b) β_{Elev}

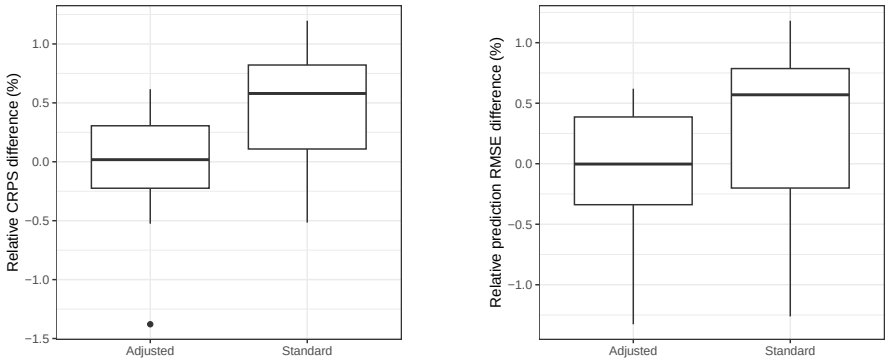
Figure 5.5: The estimated posterior distributions for two of the covariate coefficients in $\mathbf{M}_{\text{Std.}}$ (green) and $\mathbf{M}_{\text{Adj.}}$ (red) for a given dataset in Scenario 1. The histograms display the sample-based posteriors of $\mathbf{M}_{\text{SB}}$.



(a) μ (b) β_{Elev}

(c) ρ_{S} (d) σ^2_{N}

Figure 5.6: Cross-dataset biases of point estimates of (a,b) covariate coefficient, (c) the spatial range and (d) the nugget variance for $\mathbf{M}_{\text{Std.}}$, $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{SB}}$ in Scenario 2.



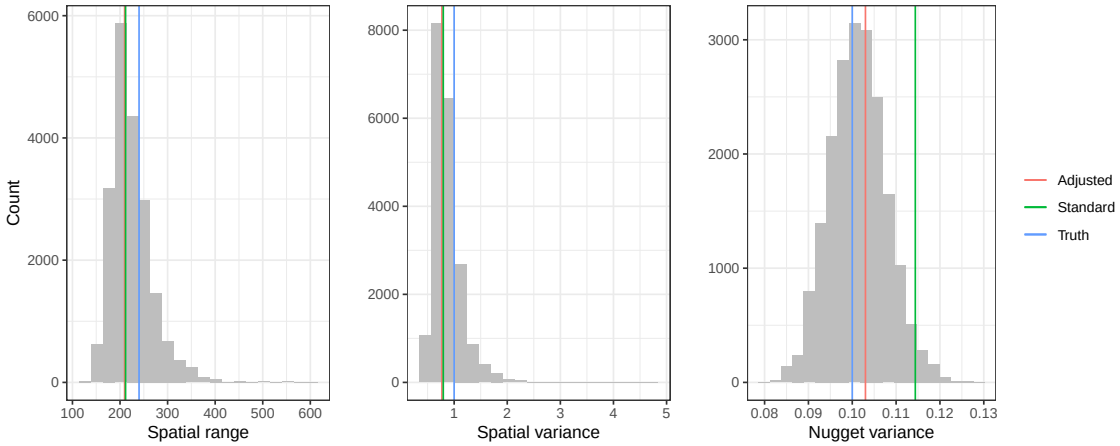
(a) CRPS (b) RMSE

Figure 5.7: (a) Pair-wise relative differences in CRPS and (b) Pair-wise relative differences in prediction RMSE across datasets for $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ relative to $\mathbf{M}_{\text{SB}}$ in Scenario 2.

measures, $\mathbf{M}_{\text{Adj.}}$ outperforms $\mathbf{M}_{\text{Std.}}$ in both scenarios. We expect the difference in parameter bias and predictive measures between $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Adj.}}$ to increase with larger jittering scales and stronger covariates (Altay et al., 2022b; Totland, 2022)

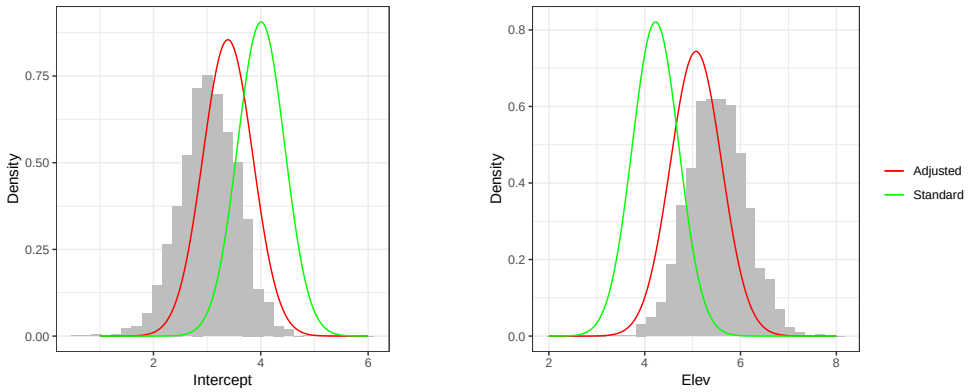
The qualitative plots of parameter posteriors suggest that their exact posteriors are unimodal in the scenarios considered, which makes the optimization in TMB suitable to find MMAP estimates. Furthermore, the qualitative plots of covariate coefficient posteriors display unimodal and close to symmetric distribution, which indicates that the Gaussian approximation used in the numerical approach is reasonable in these scenarios.

To summarize, the comparison of $\mathbf{M}_{\text{Std.}}$, $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{SB}}$ demonstrates that adjusting for jittering in the numerical approach may lead to more accurate inference in terms of more accurate posterior distributions of the covariate coefficients and less biased estimates of the measurement error variance. Additionally, we found the sample-based approach too computationally intensive for practical use in scenarios with larger effects of jittering in terms of, e.g., a larger outer jittering radius or stronger covariates. Consequently, we only consider the numerical approach when we conduct a simulation study in the next chapter.



(a) ρ_{S} (b) σ^2_{S} (c) σ^2_{N}

Figure 5.8: Sample-histograms of $\mathbf{M}_{\mathrm{SB}}$ for (a) the spatial range (b) the spatial variance and (c) the variance of the nugget effect, for a given dataset in Scenario 2. The true parameter values are given in blue, while the estimates of $\mathbf{M}_{\mathrm{Std.}}$ and $\mathbf{M}_{\mathrm{Adj.}}$ are given in green and red, respectively.



(a) μ (b) β_{Elev}

Figure 5.9: The estimated posterior distribution for two of the covariate coefficients in $\mathbf{M}_{\mathrm{Std.}}$ (green) and $\mathbf{M}_{\mathrm{Adj.}}$ (red) for a given dataset in Scenario 2. The histograms display the sample-based posteriors of $\mathbf{M}_{\mathrm{SB}}$.

Chapter 6

Simulation Study

In Chapter 5, we demonstrated that the numerical approach can be used to approximate the inference of the sample-based approach. Additionally, the numerical approach is fast and its application is not restricted to scenarios with limited effect of jittering. Given these qualities, we continue with the numerical approach in this chapter, and compare $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ under varying jittering scales and with strong covariates. Additionally, we want to explore how the jittering schemes can be evaluated with respect to some measure of privacy protection, so that the choice of scheme can be properly justified. We proceed by describing the setup of the simulation study in Section 6.1. Results from the study are presented in Section 6.2. Lastly, we explore how privacy protection can be evaluated in Section 6.3.

6.1 Simulation Setup

First, we describe the setup which is fixed across simulation scenarios. We simulate observations from the geostatistical model described in Section 4.1 on the same rectangular domain $\mathcal{D}$ as described in Section 5.1. The true locations are sampled uniformly on $\mathcal{D}$, before jittering is applied. We simulate $n = 1000$ observations in each dataset, because DHS datasets are typically of this magnitude (in clusters). We set the spatial range to 20% of the domain width, namely $\rho_{\text{S}} = 160$, while taking $\sigma^2_{\text{S}} = 1$ and $\sigma^2_{\text{N}} = 0.1$. That is, we want the contribution of the spatial field to the response to most likely be in the range ± 2, and the variance of the measurement error to be 10% of the marginal variance of the spatial field. Furthermore, we use the full set of covariates presented in Table 6.1, which are displayed in Figure 6.1. Then, the vector of covariate coefficients is given by

$$\boldsymbol{\beta} = (\mu, \beta_{\text{PopC}}, \beta_{\text{Elev}}, \beta_{\text{DistW}}, \beta_{\text{DistR}})^{\text{T}}.$$

All covariates have equal strength, i.e., $\boldsymbol{\beta} = \beta(1, 1, 1, 1, 1)^{\text{T}}$, where we set $\beta = 8$, which gives covariates that account for 68% of the variation in the response according to the calculation procedure described in Section 5.1. That is, we consider a situation wherein the covariates explain the majority of the variation in the response, which is motivated by the results of Totland (2022), who finds that the covariates should explain more than 50% of the variation for the jittering to have a deciding impact on inference, while very strong covariates explaining $\sim 80\%$ of the variation in the response could lead to stability issues in the numerical approach. We use the same priors as described at the end of Section 5.1, for the same reasons as presented there.

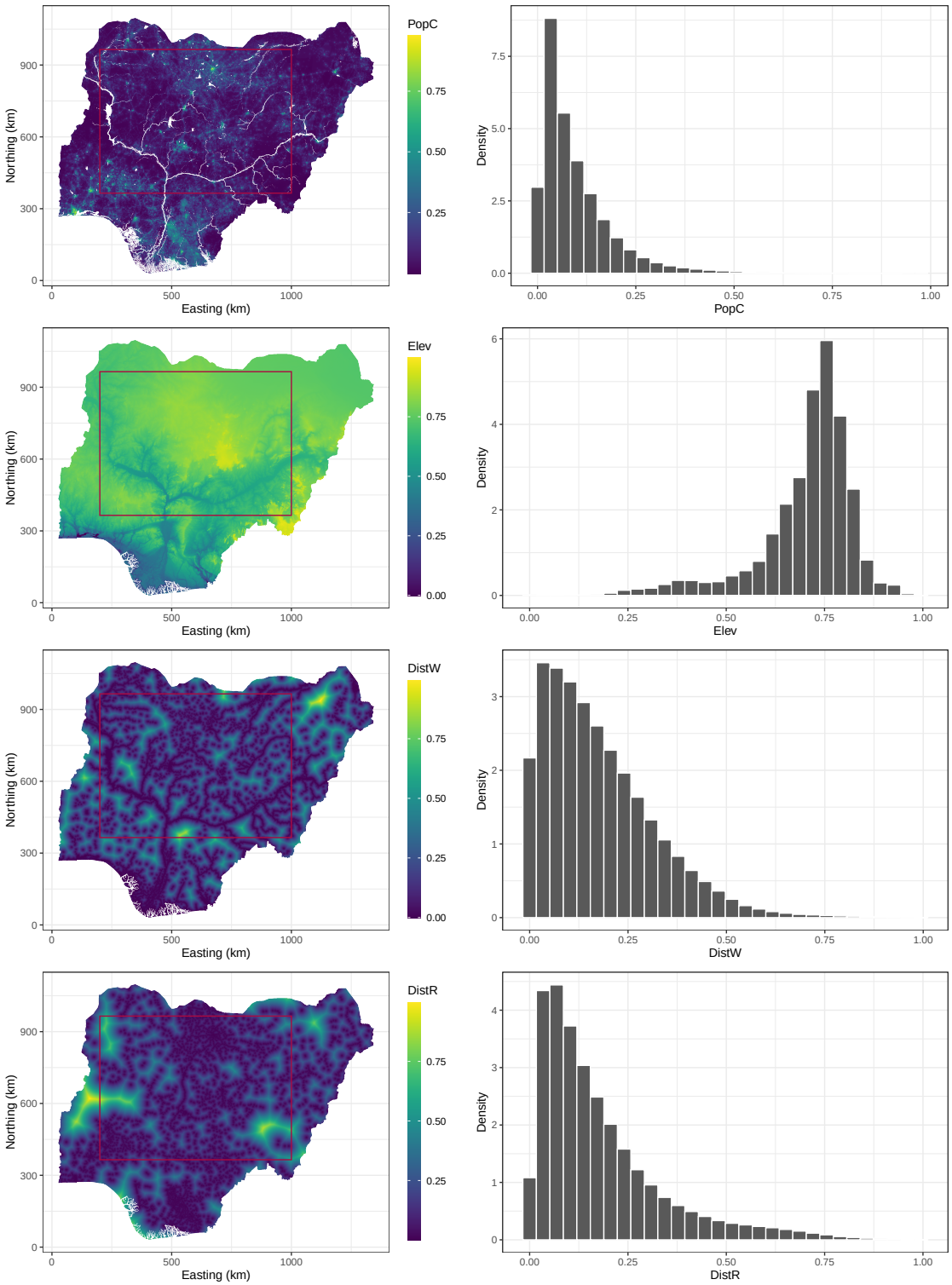


Figure 6.1: Rasters of the spatial covariates used in the simulation study, plotted along with histograms of their pixel values. The first row displays PopC, the second row displays Elev, the third displays DistW, and the third row displays DistR. The red rectangles indicate the areas used in the simulation study.

Covariate	Description	Source
PopC	Min-max scaled $\log(1+x)$ transformation of the population count.	Link
Elev	Min-max scaled $\log(1+x)$ transformation of terrain elevation once negative values are set to 0.	Link
DistW	Min-max scaled $\log(1+x)$ transformation of distance to inland water.	Link
DistR	Min-max scaled distance to major road intersections.	Link

Table 6.1: Description of covariates used throughout the text. All covariates are min-max scaled to be between 0 and 1. A link to the source is available in the digital version of this text.

To study and compare how different models for positional uncertainty impact the inference with $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$, we consider both disk and donut jittering on a *moderate* and *substantial* jittering scale. Specifically, we consider disk jittering with $(R_1, R_2) \in \{(0,3),(0,9)\}$ and donut jittering with $(R_1, R_2) \in \{(1,3),(3,9)\}$. That is, we let $R_1 = R_2/3$ for donut jittering. By comparison, the geomasking strategy of the 2018 Nigeria DHS (NDHS2018) dataset (which is closely described in Section 7.1) uses disk jittering with $R_2 = 2$ for locations classified as urban, and $R_2 = 5$ locations classified as rural, with a 1% chance of rural clusters being displaced up to 10 km. With this in mind, we argue that the choice of $R_2 = 3$ adequately represents the case of moderate jittering, while $R_2 = 9$ adequately represents the case of substantial jittering for the spatial scale considered here.

6.2 Results

We simulated 50 datasets with the setup described in Section 6.1 and applied all four jittering schemes to each dataset, giving $50 \times 4 = 200$ datasets with positional uncertainty. Then, we estimated $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ with each of the 200 jittered datasets. In total, this amounted to estimating $200 \times 2 = 400$ models, which took around 26 hours in total. Estimating $\mathbf{M}_{\text{Adj.}}$ took 4.5 minutes on average, while estimating $\mathbf{M}_{\text{Std.}}$ took 8.3 seconds on average.

6.2.1 Parameter Estimates

To evaluate $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ with respect to parameter estimation, we consider both the relative bias, $(\hat{\theta}-\theta)/\theta$ and the absolute bias, $\hat{\theta}-\theta$, where $\hat{\theta}$ is the posterior mean for covariate coefficients and MMAP estimates for ρ_{S}, σ^2_{S} and σ^2_{N}. Specifically, the median relative bias corresponding to ρ_{S}, σ^2_{S} and σ^2_{N} and the median absolute bias corresponding to the elements of $\boldsymbol{\beta}$ are reported in Table 6.2 for both models under all jittering schemes. Additionally, Figure 6.2 displays the variation in absolute bias across datasets in $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ for ρ_{S}, σ^2_{S} and σ^2_{N} and β_{Elev} under the four

	Median Bias							
	$\mathbf{M}_{\text{Adj.}}$				$\mathbf{M}_{\text{Std.}}$			
Scheme	Disk		Donut		Disk		Donut	
R_2	3	9	3	9	3	9	3	9
ρ_{S}	−0.5%	−3.4%	−4.3%	0.7%	−3.6%	−14.7%	−7.6%	−5.8%
σ^2_{S}	−3.5%	−7.5%	−5.9%	−8.2%	−3.9%	0.3%	−2.2%	7.3%
σ^2_{N}	4.3%	34.5%	15.5%	61.1%	135.9%	342.5%	184.7%	443.6%
μ	0.52	0.32	0.27	0.51	1.17	1.58	1.34	2.64
β_{PopC}	−0.15	−1.08	−0.53	−1.54	−2.19	−3.59	−2.62	−5.27
β_{Elev}	−0.75	−0.09	−0.24	−0.16	−1.08	−0.93	−1.16	−2.19
β_{DistR}	−0.19	−0.47	−0.13	−0.56	−0.63	−1.10	−0.71	−1.68
β_{DistW}	0.40	−0.42	0.17	−0.49	0.16	−1.15	−0.15	−1.38

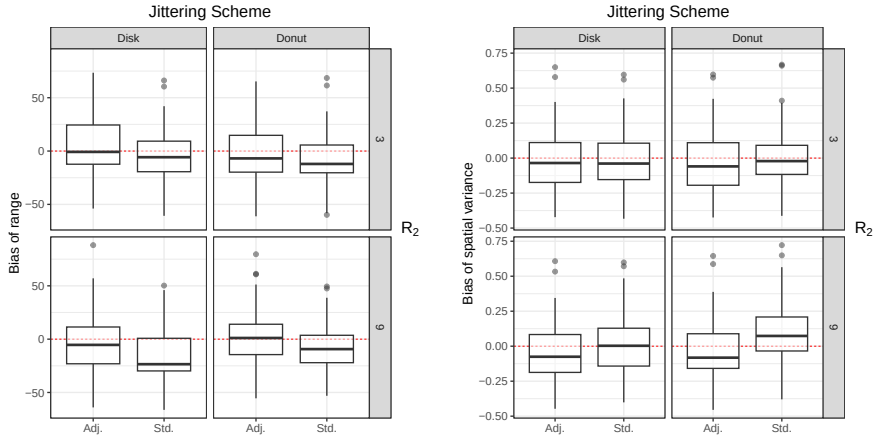
Table 6.2: The median relative bias (in percent) corresponding to estimates of ρ_{S}, σ^2_{S} and σ^2_{N} and the median absolute bias corresponding to the elements of $\boldsymbol{\beta}$ in $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ for all jittering schemes.

jittering schemes.

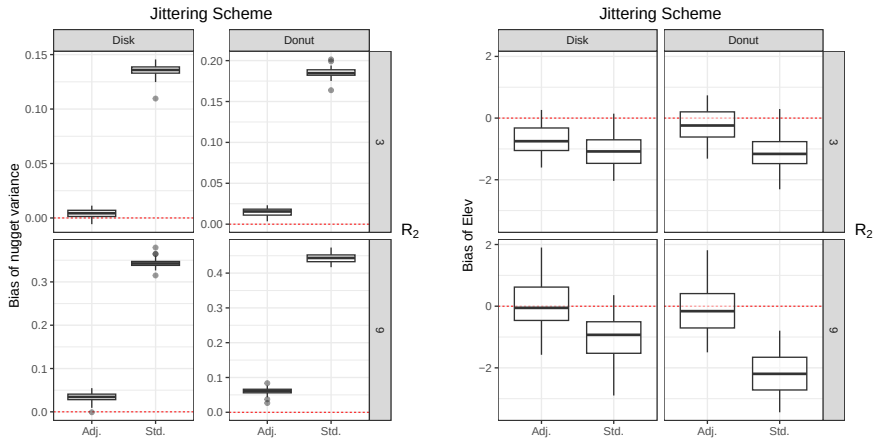
The cross-dataset biases of the spatial parameters, ρ_{S} and σ^2_{S}, are respectively displayed in Figures 6.2a and 6.2b, which show that the spread and the median bias are similar for both models under all jittering schemes. However, we do observe from Table 6.2 that the spatial range is always underestimated in $\mathbf{M}_{\text{Std.}}$ with more bias than in $\mathbf{M}_{\text{Adj.}}$ with respect to the median. This is in line with our expectations from the theory discussed in Section 3.3, and we see that the effect is amplified with more jittering ($R_2 = 9$). Nonetheless, the difference may not be of practical importance. On the other hand, Table 6.2 and Figure 6.2c show a striking difference between the models in the bias corresponding to σ^2_{N}, where $\mathbf{M}_{\text{Adj.}}$ has substantially less biased estimates than $\mathbf{M}_{\text{Std.}}$. For example, the median bias in σ^2_{N} is almost 10 times larger in $\mathbf{M}_{\text{Std.}}$ when compared to $\mathbf{M}_{\text{Adj.}}$ under disk jittering with $R_2 = 9$. This result also agrees with our expectations from the theory discussed in Section 3.3.

For β_{Elev}, whose bias across datasets is plotted in Figure 6.2d, $\mathbf{M}_{\text{Adj.}}$ clearly exhibits superior performance compared to $\mathbf{M}_{\text{Std.}}$ across all jittering schemes, where the difference increases with more jittering ($R_2 = 9$). The biases corresponding to the other elements of $\boldsymbol{\beta}$ are displayed in Figure E.1, wherein we see the same pattern as for β_{Elev}, except for the estimates of β_{DistW} when $R_2 = 3$, which are less biased in $\mathbf{M}_{\text{Std.}}$ with respect to the absolute value of the median. We observe that both models tend to underestimate the covariate coefficients (excluding the intercept), where $\mathbf{M}_{\text{Std.}}$ underestimates the coefficients more strongly. This suggests that when jittering is disregarded, it results in more attenuated estimates of the covariate coefficients, and that using $\mathbf{M}_{\text{Adj.}}$ can help mitigate the attenuation. This result aligns with the findings of Totland (2022) and Altay et al. (2022b).

Importantly, some outliers have been removed in the bias plots of the covariate coefficient estimates in $\mathbf{M}_{\text{Adj.}}$ to make the figures interpretable. Specifically, two outliers have been removed for donut jittering with $(R_1, R_2) = (3, 9)$ and one for disk jittering with $(R_1, R_2) = (0, 9)$. We delay the discussion of these outliers to Chapter 8.



(a) ρ_{S} (b) σ^2_{S}



(c) σ^2_{N} (d) β_{Elev}

Figure 6.2: The cross-dataset absolute bias in $\mathbf{M}_{\mathrm{Adj.}}$ and $\mathbf{M}_{\mathrm{Std.}}$ corresponding to (a) the spatial range, ρ_{S}, (b) the spatial variance, σ^2_{S}, (c) the nugget variance, σ^2_{N}, and (d) β_{Elev} under all jittering schemes in the simulation study.

	Median Score							
	$\mathbf{M}_{\text{Adj.}}$				$\mathbf{M}_{\text{Std.}}$			
Scheme	Disk		Donut		Disk		Donut	
R_2	3	9	3	9	3	9	3	9
CRPS	0.266	0.282	0.266	0.286	0.281	0.315	0.285	0.344
RMSE	0.458	0.491	0.461	0.503	0.490	0.553	0.500	0.613

Table 6.3: The median CRPS and prediction RMSE of $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ for all jittering schemes.

6.2.2 Predictions

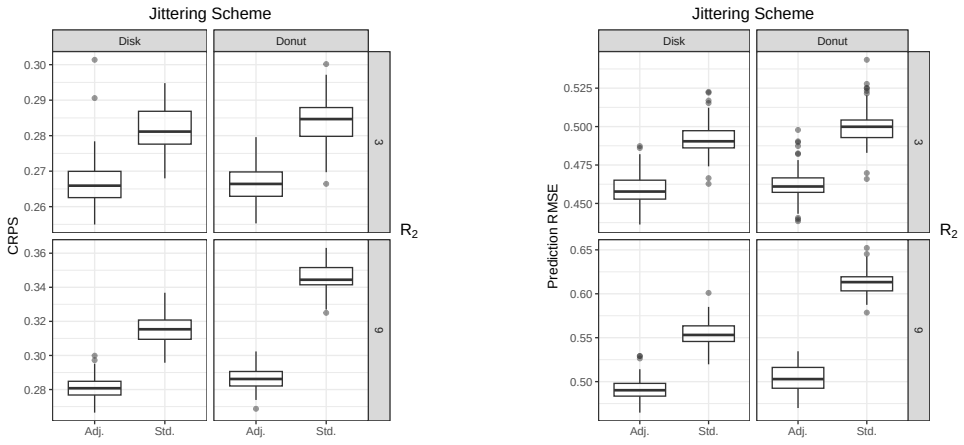
We evaluate the predictive power of $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ by drawing samples of the posterior latent process as described in Section 4.2.2 at a set of regularly spaced points over the domain, $\mathcal{D}$. The samples are used to calculate CRPS and prediction RMSE. Table 6.3 shows that the median CRPS is lower for $\mathbf{M}_{\text{Adj.}}$ for all jittering schemes, where the difference between the models is largest for donut jittering with $R_2 = 9$, where the median CRPS is 17% lower for $\mathbf{M}_{\text{Adj.}}$ relative to the median CRPS of $\mathbf{M}_{\text{Std.}}$. The variation in CRPS across datasets for $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ are displayed in a boxplot in Figure 6.3a, which shows superior performance for $\mathbf{M}_{\text{Adj.}}$.

We create point predictions by taking the mean over posterior samples of the latent process, and calculate the prediction RMSE of the models. Table 6.3 shows that the median prediction RMSE is lower for $\mathbf{M}_{\text{Adj.}}$ for all jittering schemes, where the difference between the models is largest for donut jittering with $R_2 = 9$, where the median RMSE is 18% lower for $\mathbf{M}_{\text{Adj.}}$ relative to the median RMSE of $\mathbf{M}_{\text{Std.}}$. The variation in prediction RMSE across datasets is illustrated in Figure 6.3b as a boxplot, which shows that $\mathbf{M}_{\text{Adj.}}$ outperforms $\mathbf{M}_{\text{Std.}}$. For both prediction RMSE and CRPS, the difference between the models increases with more jittering ($R_2 = 9$), and the differences are slightly larger under donut jittering compared to under disk jittering.

We underscore that some outliers have been removed for $\mathbf{M}_{\text{Adj.}}$ in Figure 6.3 to make the plot interpretable. For both CRPS and prediction RMSE, one outlier was removed for donut jittering with $(R_1, R_2) = (1, 3)$, two were removed for donut jittering with $(R_1, R_2) = (3, 9)$, three for disk jittering with $(R_1, R_2) = (0, 3)$ and six for disk jittering with $(R_1, R_2) = (1, 3)$, which amounts to 12 outliers in total. That is, $12/200 = 6\%$ of the simulated datasets resulted in outliers for $\mathbf{M}_{\text{Adj.}}$. Based on these numbers, more outliers occur when the jittering scale increases from $R_2 = 3$ to $R_2 = 9$, and more outliers occur for disk jittering than for donut jittering. We delay the discussion of these outliers to Chapter 8.

6.3 Evaluating Privacy Protection

The results of the simulation study in Section 6.2 demonstrate how jittering may lead to erroneous inference if ignored. Furthermore, we observe that the impact of jittering depends on the specific jittering scheme, where, unsurprisingly, increasing the jittering scale increases the negative impact on inference, epsecially when it is ignored. Therefore, the jittering scheme, or more broadly, the geomasking strategy,



(a) CRPS (b) RMSE

Figure 6.3: Predictive measures of $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ where (a) displays the CRPS and (b) displays the prediction RMSE under all jittering schemes in the simulation study.

should be chosen with care in a specific application, where privacy protection should be weighted against the scale of the masking. However, to our knowledge, there exists no widely-agreed-upon approach to compare different geomasking strategies with respect to privacy protection, at least not when the geomasking strategy is known. For this reason, we explore some possible approaches for evaluating and comparing the privacy protection of donut and disk jittering schemes.

6.3.1 Approximating the Complete Posteriors

The k-anonymity metric has been used to evaluate privacy protection of geomasking strategies (Hampton et al., 2010; Allshouse et al., 2010), but as discussed in Section 3.2, this metric is unsuitable in the situation we consider, where the positional error model is known. In stead, we try to evaluate the privacy protection of different jittering schemes by considering the marginal posterior of each true location, which we refer to as the *complete* posterior. For $i = 1, \ldots, n$, the complete posterior of a true location is given by

$$\begin{aligned}
\pi(\boldsymbol{s}_i^* | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) &= \int \pi(\boldsymbol{s}_i^*, \boldsymbol{z} | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{z} \\
&= \int \pi(\boldsymbol{s}_i^* | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}, \boldsymbol{z}) \pi(\boldsymbol{z} | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{z} \\
&= \int \left[\frac{\pi(\boldsymbol{s}_i^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}, \boldsymbol{z})}{\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}, \boldsymbol{z})} \right] \pi(\boldsymbol{z} | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{z} \quad (6.1)
\end{aligned}$$

To simplify the fraction in the square brackets in the integrand of (6.1), we use the definition of conditional probability in the numerator and the denominator, and the dependence relations among the model components, which are illustrated in Figure 4.2. This gives

$$\frac{\pi(\boldsymbol{s}_i^*, \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}, \boldsymbol{z})}{\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}, \boldsymbol{z})} = \frac{\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n | \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{s}_i^*)\pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})\pi(\boldsymbol{s}_i^*)}{\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n | \boldsymbol{\theta}, \boldsymbol{z})\pi(\boldsymbol{z}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}. \tag{6.2}$$

Furthermore, we assume as before that the observation pairs $(y_i, \boldsymbol{s}_i),\ i = 1, \ldots, n$, are conditionally independent given $\boldsymbol{\theta}$, $\boldsymbol{z}$. We assume that the same is true when we additionally condition on $\boldsymbol{s}_i^*$, and that $\pi(y_j, \boldsymbol{s}_j | \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{s}_i^*) = \pi(y_j, \boldsymbol{s}_j | \boldsymbol{\theta}, \boldsymbol{z})$ when $i \neq j$. This is reasonable, because the jittering happens independently for each observation and one true location gives no information about any other observed or true location other than its observed counterpart. Thus, the right hand side of Equation (6.2) can be further simplified as

$$\begin{aligned}\frac{\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n | \boldsymbol{\theta}, \boldsymbol{z}, \boldsymbol{s}_i^*)\pi(\boldsymbol{s}_i^*)}{\pi(\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n | \boldsymbol{\theta}, \boldsymbol{z})} &= \frac{\pi(y_i, \boldsymbol{s}_i | \boldsymbol{s}_i^*, \boldsymbol{\theta}, \boldsymbol{z})\pi(\boldsymbol{s}_i^*) \cdot \prod_{j \neq i} \pi(y_j, \boldsymbol{s}_j | \boldsymbol{\theta}, \boldsymbol{z})}{\prod_{j=1}^n \pi(y_j, \boldsymbol{s}_j | \boldsymbol{\theta}, \boldsymbol{z})} \\ &= \frac{\pi(y_i | \boldsymbol{s}_i^*, \boldsymbol{\theta}, \boldsymbol{z})\pi(\boldsymbol{s}_i | \boldsymbol{s}_i^*)\pi(\boldsymbol{s}_i^*)}{\pi(y_i, \boldsymbol{s}_i | \boldsymbol{\theta}, \boldsymbol{z})},\end{aligned} \tag{6.3}$$

where $\pi(y_i, \boldsymbol{s}_i | \boldsymbol{z}, \boldsymbol{\theta})$ can be approximated with the numerical integration scheme described in Section 4.2.1. Now, we insert the result in the last line of (6.3) into the square brackets in the integrand of (6.1) and use the fact that $\pi(\boldsymbol{s}_i^*)$ is constant over $\mathcal{D}$ to obtain

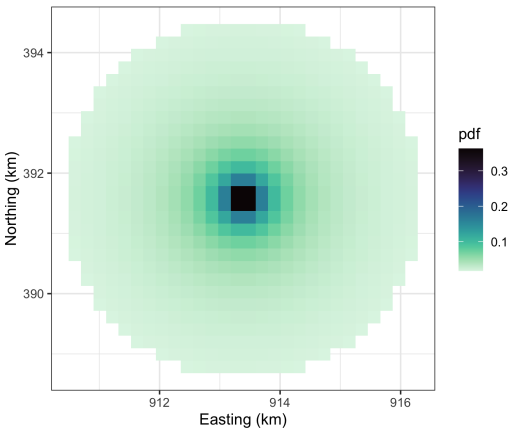
$$\pi(\boldsymbol{s}_i^* | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) \propto \pi(\boldsymbol{s}_i | \boldsymbol{s}_i^*) \cdot \int \left[\frac{\pi(y_i | \boldsymbol{s}_i^*, \boldsymbol{\theta}, \boldsymbol{z})}{\pi(y_i, \boldsymbol{s}_i | \boldsymbol{z}, \boldsymbol{\theta})} \right] \pi(\boldsymbol{z} | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{z}. \tag{6.4}$$

The integral in Equation (6.4) can be evaluated with Monte Carlo integration, where we sample from the EB posterior of the random effects as described in Section 4.2.2. Thus, the complete posterior of a true location can be approximated by

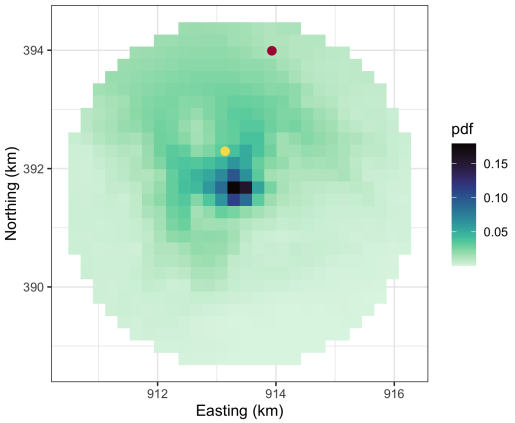
$$\pi(\boldsymbol{s}_i^* | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}) \approx C \cdot \pi(\boldsymbol{s}_i | \boldsymbol{s}_i^*) \cdot \frac{1}{J} \sum_{j=1}^{J} \frac{\pi(y_i | \boldsymbol{s}_i^*, \boldsymbol{\theta}, \boldsymbol{z}^{(j)})}{\pi(y_i, \boldsymbol{s}_i | \boldsymbol{z}^{(j)}, \boldsymbol{\theta})}.$$

where C is a normalizing constant and J is the number of samples used in the Monte Carlo integration. To calculate the complete posterior (approximately), we set $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, i.e., the MMAP estimates from $\mathbf{M}_{\text{Adj.}}$. The complete posterior is approximated over a square grid, on which we compare it to the *partial* posterior, given by $\pi(\boldsymbol{s}_i^* | \boldsymbol{s}_i)$, where we only condition on the corresponding observed location. That is, we compare $\pi(\boldsymbol{s}_i^* | \boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta})$ to $\pi(\boldsymbol{s}_i^* | \boldsymbol{s}_i)$.

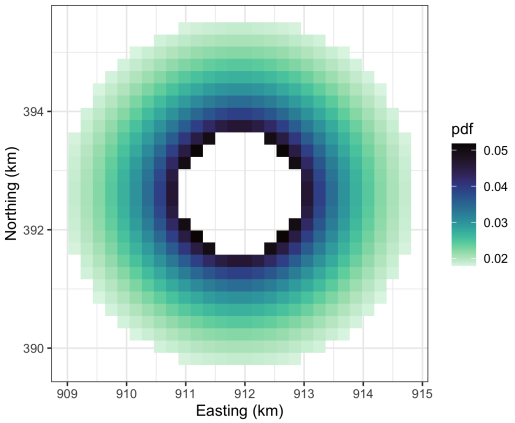
The complete and partial posterior densities are displayed for one observation in Figure 6.4 for disk jittering with $(R_1, R_2) = (0, 3)$ and donut jittering with $(R_1, R_2) = (1, 3)$. For both of these jittering schemes, we simulate 10 datasets according to the setup described in Section 6.1 and calculate the complete posterior of all true locations in each datasets. The results of this simulation are used in the following section.



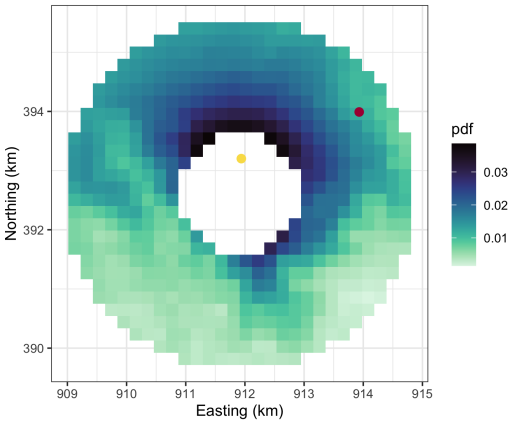
(a) $\pi(\boldsymbol{s}_i^*|\boldsymbol{s}_i)$



(b) $\pi(\boldsymbol{s}_i^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta})$



(c) $\pi(\boldsymbol{s}_i^*|\boldsymbol{s}_i)$



(d) $\pi(\boldsymbol{s}_i^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta})$

Figure 6.4: Panel (a) and (c) display the The distribution of a true location conditional on the observed locations for $R_1 = 0$ and $R_1 = 1$, respectively. In panels (b) and (d), we additionally condition on the observations and parameters. The red point in each plot shows the true location, $\boldsymbol{s}_i^*$, while the yellow point shows the expected true location, $\overline{\boldsymbol{s}_i^*}$.

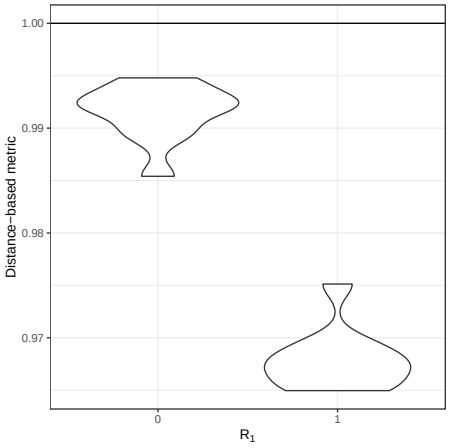


Figure 6.5: Violin plot of the distance-based metric across datasets for disk jittering ($R_1 = 0$) and donut jittering ($R_1 = 1$).

6.3.2 Investigating Two Metrics

To evaluate the privacy protection of the two jittering schemes, we consider the expectation of the partial and complete posterior for a given true location. For the partial posterior, $\mathrm{E}[\boldsymbol{s}_i^*|\boldsymbol{s}_i] = \boldsymbol{s}_i$, since the distribution is symmetric around the observed location. For the complete posterior, we numerically calculate

$$\overline{\boldsymbol{s}_i^*} = \mathrm{E}[\boldsymbol{s}_i^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta}]$$

from the approximated complete posterior. Figures 6.4b and 6.4d display the expected true location of the complete posterior (yellow point) for one observation. Further, we consider the distance from these posterior expectations to the unknown true location. That is, we compare $\|\overline{\boldsymbol{s}_i^*} - \boldsymbol{s}_i^*\|$ and $\|\boldsymbol{s}_i - \boldsymbol{s}_i^*\|$. For this purpose we consider the median of the fraction

$$\frac{\|\overline{\boldsymbol{s}_i^*} - \boldsymbol{s}_i^*\|}{\|\boldsymbol{s}_i - \boldsymbol{s}_i^*\|}, \quad i = 1, \ldots, n, \tag{6.5}$$

over all true locations, which we refer to as the *distance-based* metric. Figure 6.5 displays the distance-based metric across datasets for the two jittering schemes. Based on this plot, we learn more about the unknown true location under the donut jittering scheme than under the disk jittering scheme, with respect to the distance-based metric. However, we note that fairly little is learned: under the donut jittering scheme, Figure 6.5 shows that the expectation of the complete posterior is only around 3% closer to the unknown true location than the expectation of the partial posterior, with respect to the median. However, the complete posteriors may have multiple local areas with higher mass that are far apart, as Figure 6.4d illustrates. This may render the distance-based metric unsuitable to accurately evaluate privacy protection, since it only considers the global (average) behaviour of the complete posteriors.

As an alternative approach to evaluate the privacy protection of jittering schemes, we calculate the KLD of the complete posterior, $p(\boldsymbol{s}_i^*) = \pi(\boldsymbol{s}_i^*|\boldsymbol{y}, \boldsymbol{s}_1, \ldots, \boldsymbol{s}_n, \boldsymbol{\theta})$, from the partial posterior $q(\boldsymbol{s}_i^*) = \pi(\boldsymbol{s}_i^*|\boldsymbol{s}_i)$, which we approximate as follows,

$$D_{\mathrm{KL}}(p \parallel q) = \int_{\mathbb{R}^2} p(\boldsymbol{s}_i^*) \log\left(\frac{p(\boldsymbol{s}_i^*)}{q(\boldsymbol{s}_i^*)}\right) \mathrm{d}\boldsymbol{s}_i^* \approx \sum_{j \in \mathcal{G}_i} p(\boldsymbol{s}_{ij}^*) \log\left(\frac{p(\boldsymbol{s}_{ij}^*)}{q(\boldsymbol{s}_{ij}^*)}\right) \Delta\boldsymbol{s}_{ij}^*,$$

where $\mathcal{G}_i$ is an index set corresponding to a regular grid of square cells over the support of $\pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)$, $\boldsymbol{s}_{ij}^*$ is the center point of the j'th cell in the grid and $\Delta\boldsymbol{s}_{ij}^*$ is the area of the j'th cell in the grid. The KLD can be interpreted as the amount of information lost when using q to approximate p (Burnham and Anderson, 2002, p. 51). Thus, a positive KLD indicates that information has been gained about the unknown true locations when we transition from the partial to the complete posterior. For a given dataset, we consider the average KLD over all true locations, $\overline{D}_{\mathrm{KL}}$. Across the 10 simulated datasets, we obtain that $\overline{D}_{\mathrm{KL}}$ is 0.057 on average for disk jittering and 0.045 on average for donut jittering. That is, we obtain that, on average, $\overline{D}_{\mathrm{KL}}$ is 27% higher for the disk jittering scheme relative to donut jittering scheme. This could indicate that more is learned about the unknown true locations under the disk jittering scheme, which is in contrast to the findings with the distance-based metric, for which donut jittering performed worse.

Using KLD to directly compare jittering schemes in the sense described above may be unreasonable, because we consider the KLD from different distributions, i.e., the distribution q differs between the jittering schemes. Nonetheless, the disagreement between the distance-based and the KLD-based metric may suggest that more sophisticated methods are needed to adequately compare the privacy protection provided by different jittering schemes.

Chapter 7

Case Study

The practical application of geomasking in DHS data is part of the main motivation for studying positional uncertainty in the context of geostatistical modeling. Previous work support the use of the numerical approach to adjust for positional uncertainty in DHS data (Altay et al., 2022a,b; Totland, 2022). Our results speak for the same. For example, the simulation study in Chapter 6, which considered similar jittering scales to those employed in many DHS datasets, demonstrated that failing to adjust for jittering may lead to more attenuated estimates of covariate coefficients and a substantial underestimation of the nugget variance. In this chapter, we therefore demonstrate a practical use case for the numerical approach by applying it to geomasked DHS data in a study of vaccination coverage in Nigeria. The data is described in Section 7.1, before we present the model and results in Section 7.2.

7.1 Data Description

We consider the NDHS2018 dataset, and model the vaccination coverage of the first dose of measles-containing-vaccine (MCV1) among children under 5 years of age. With its estimated 223 million residents (UN, Department of Economic and Social Affairs, 2022), Nigeria has one of the highest measles burdens in the world (Sato et al., 2022), and collection of data combined with modeling of key indicators such as vaccination coverage can be crucial for effectively designing policies and distributing vaccination resources to combat the spread of the illness. Fuglstad et al. (2022) also analyse MCV1 in Nigeria in a spatial statistics context, and focus on how complex survey designs can be better utilized in model-based approaches.

The NDHS2018 dataset was sampled in tow stages. First, clusters consisting of multiple households were sampled, before a fixed number of 30 households were sampled for each selected cluster. Fieldwork was completed for 1389 of the sampled clusters (National Population Commission and ICF, 2019). We model MCV1 at the cluster level, and the locations of the 1379 clusters that were eligible for analysis are displayed in Figure 7.1. The reported cluster locations contain positional uncertainty from geomasking (Burgert et al., 2013). The geomasking strategy consists of a combination of different disk jittering schemes. Specifically, clusters classified as urban are disk jittered with an outer radius of $R_2 = 2$ (km), while clusters classified as rural are displaced up to 5 km ($R_2 = 5$), and have a 1% chance of being displaced up to 10 km ($R_2 = 10$). Additionally, the jittered locations are restricted to lie within the same administrative region as their true counterparts. For this dataset,

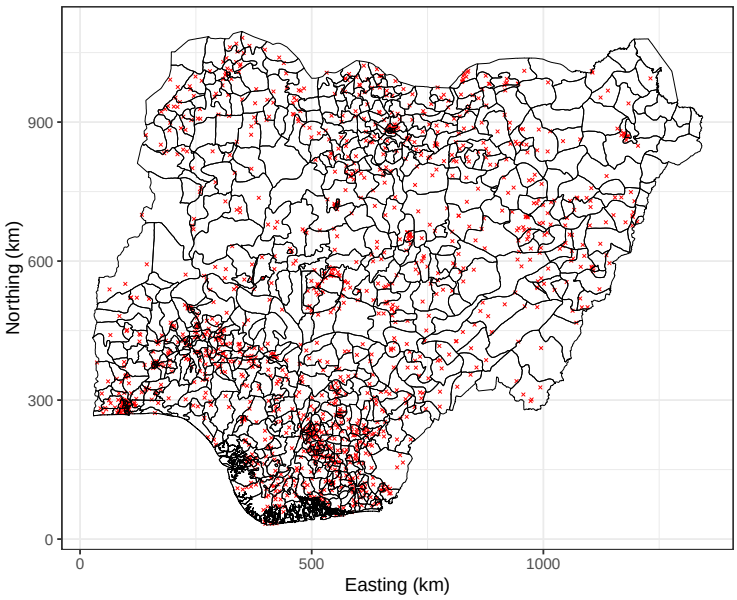


Figure 7.1: The borders of the 774 mainland admin2 regions in Nigeria. The red points show the locations of clusters used to study MCV1.

the second administrative level is used, which we refer to as the *admin2* level. That is, a jittered location, $\boldsymbol{s}$ is restricted to lie within the same admin2 region as its corresponding unknown true location, $\boldsymbol{s}^*$. The admin2 level of Nigeria is divided into 774 regions, whose boundaries are displayed in Figure 7.1. We use the shape file of the admin2 level boundaries which is provided by GADM (2022).

To describe the geomasking strategy mathematically, recall that $\boldsymbol{s}_1^*, \ldots, \boldsymbol{s}_n^* \in \mathcal{D}$ denote the unknown true locations, while $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n \in \mathcal{D}$ denote the corresponding jittered locations, which we observe. Let $\text{Urb}[i] \in \{\text{U}, \text{R}\}$ indicate whether location $\boldsymbol{s}_i$ is urban (U) or rural (R). Then, the positional uncertainty is modeled independently for each location as

$$\boldsymbol{s}_i | \boldsymbol{s}_i^* \sim \pi_{\text{Urb}[i]}(\boldsymbol{s}_i | \boldsymbol{s}_i^*), \quad i = 1 \ldots n,$$

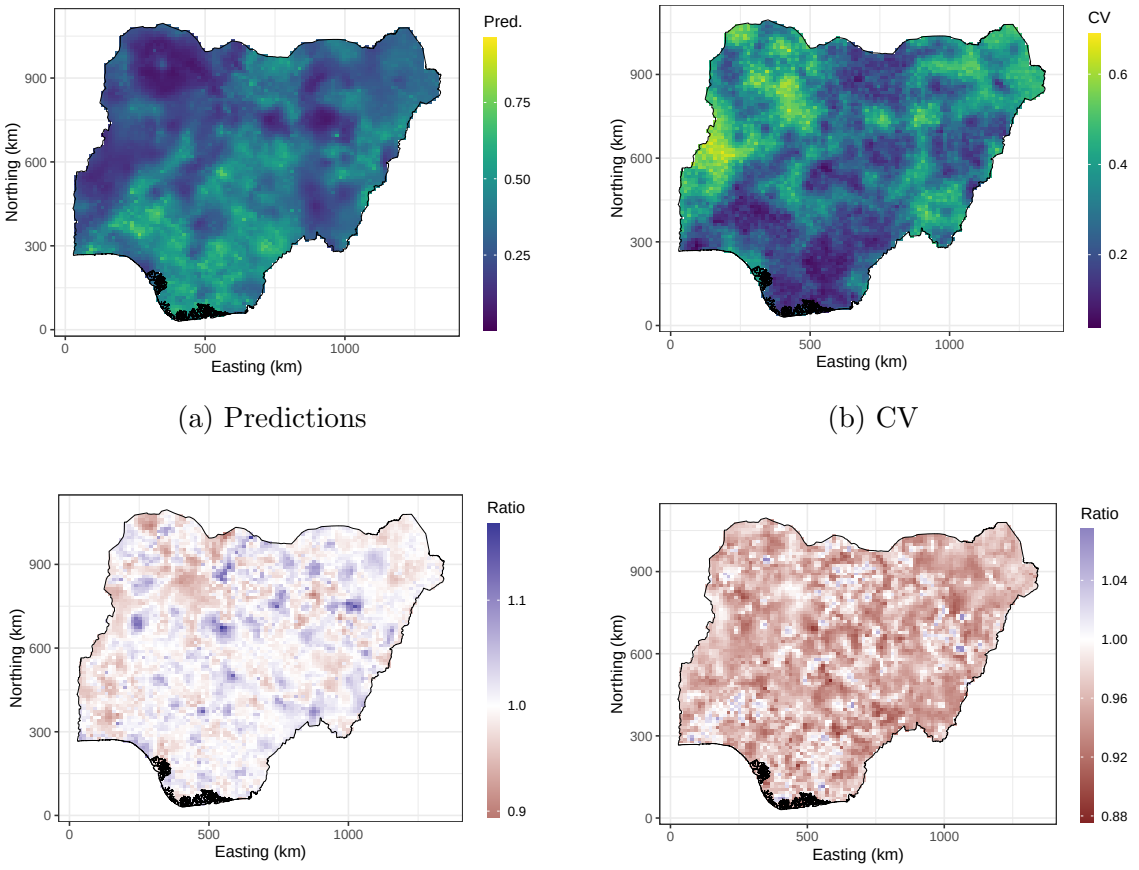
where urban locations are geomasked with disk jittering with $R_2 = 2$,

$$\pi_{\text{U}}(\boldsymbol{s}_i | \boldsymbol{s}_i^*) \propto \frac{\mathbb{I}(A(\boldsymbol{s}_i) = A(\boldsymbol{s}_i^*)) \cdot \mathbb{I}(d(\boldsymbol{s}_i, \boldsymbol{s}_i^*) < 2)}{d(\boldsymbol{s}_i, \boldsymbol{s}_i^*)}, \quad \text{Urb}[i] = \text{U}.$$

Here, $A(\boldsymbol{s})$ is the admin2 region corresponding to location $\boldsymbol{s}$ and $d(\boldsymbol{s}, \boldsymbol{s}^*)$ is the distance in kilometers between locations $\boldsymbol{s}$ and $\boldsymbol{s}^*$. Rural clusters, on the other hand, are geomasked with a "mixture" of two disk jittering schemes,

$$\pi_{\text{R}}(\boldsymbol{s}_i | \boldsymbol{s}_i^*) \propto \frac{\mathbb{I}(A(\boldsymbol{s}_i) = A(\boldsymbol{s}_i^*)) \cdot \mathbb{I}(d(\boldsymbol{s}_i, \boldsymbol{s}_i^*) < D_{\text{R}}^{(i)})}{d(\boldsymbol{s}_i, \boldsymbol{s}_i^*)}, \quad \text{Urb}[i] = \text{R},$$

where the $D_{\text{R}}^{(i)}$'s are i.i.d. random variables taking the value 5 with a probability of 99% and the value 10 with a probability of 1%. A technical description of the



(a) Predictions

(b) CV

(c) Ratio of Predictions

(d) Ratio of Uncertainties

Figure 7.2: (a) Predictions of MCV1 of $\mathbf{M}_{\text{Adj.}}$ (Pred.) and (b) corresponding CVs. (c) Ratios (Adj./Std.) of point predictions and (d) ratios (Adj./Std.) of CVs for $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$

integration scheme which is used for rural clusters can be found in Section E of the Appendices of Altay et al. (2022a).

7.2 Modeling Vaccination Coverage

7.2.1 Model Description

We model MCV1 with the same latent and parameter layer (excluding the nugget effect) as in the model described in Section 4.1, but, to study prevalence, we use a binomial observation likelihood and a logit link, such that

$$y_i|r_i, n_i \sim \text{Binomial}\left(n_i,\ r_i\right), \quad \boldsymbol{s}_i|\boldsymbol{s}_i^* \sim \pi_{\text{Urb}[i]}(\boldsymbol{s}_i|\boldsymbol{s}_i^*),$$
$$r_i|\boldsymbol{s}_i^* = \text{logit}^{-1}(\eta(\boldsymbol{s}_i^*)), \quad \boldsymbol{s}_i \in \mathcal{D}, \quad i = 1, \dots n,$$

where n_i is the number of children in cluster i, r_i is the risk in cluster i and $\mathcal{D}$ denotes the geographical area of Nigeria in the projected coordinate system given by the "Minna / Nigeria East Belt" projection. Furthermore, we use the same

	$\mathbf{M}_{\text{Adj.}}$		$\mathbf{M}_{\text{Std.}}$	
Parameter	Estimate	95% CI	Estimate	95% CI
ρ_{S}	84.95	-	77.85	-
σ^2_{S}	0.69	-	0.75	-
μ	-0.97	$(-1.56, -0.37)$	-0.91	$(-1.49, -0.34)$
β_{PopC}	1.40	$(1.08, 1.72)$	1.12	$(0.82, 1.40)$
β_{Elev}	0.27	$(-0.55, 1.09)$	0.27	$(-0.52, 1.05)$
β_{DistW}	0.004	$(-0.52, 0.52)$	-0.01	$(-0.53, 0.50)$
β_{DistR}	-0.74	$(-1.34, -0.13)$	-0.90	$(-1.50, -0.29)$

Table 7.1: The parameter estimates of $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ with 95% credible intervals (CIs) for the covariate coefficients.

covariates as in the simulation study in Chapter 6, which are described in Table 6.1 and displayed in Figure 6.1.

We use PC priors on the spatial field given by $\text{P}(\rho_{\text{S}} < 240) = 0.5$ and $\text{P}(\sigma_{\text{S}} > 1) = 0.5$, which are respectively chosen as such because 20% of Nigeria's width at its widest (from west to east) is approximately 240 km, and we do not expect the spatial field to contribute with much more than ± 2 to the latent process. Furthermore, we let $\boldsymbol{\beta} \sim \mathcal{N}_4(\mathbf{0}, 25\mathbf{I}_4)$, because we expect the contribution from the min-max scaled covariates to the latent process to be within the range ± 10.

7.2.2 Predictions and Parameter Estimates

Preparing input and estimating both $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{Adj.}}$ took around 19 minutes in total, where the estimation of $\mathbf{M}_{\text{Std.}}$ took half a minute and the estimation of $\mathbf{M}_{\text{Adj.}}$ took around 14 minutes. The predictions of MCV1 for $\mathbf{M}_{\text{Adj.}}$, i.e., point predictions of the transformed latent process, $\text{logit}^{-1}(\eta)$, at a set of regularly spaced locations over Nigeria are displayed in Figure 7.2a along with corresponding coefficients of variation (CVs) in Figure 7.2b. Corresponding figures for $\mathbf{M}_{\text{Std.}}$ are not shown, because they are very similar visually. Additionally, Figure 7.2c displays the ratio (Adj./Std.) of point predictions, showing that the models give very similar predictions. Figure 7.2d displays the ratio of CVs (Adj./Std.), which reveals that $\mathbf{M}_{\text{Std.}}$ generally has larger uncertainty in the predictions.

The MMAP estimates of the spatial parameters, along with point estimates and credible intervals for the covariate coefficients are given in Table 7.1 for both $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$. The table shows that the estimates of ρ_{S} and σ^2_{S} are quite similar, though $\mathbf{M}_{\text{Adj.}}$ gives a higher range than $\mathbf{M}_{\text{Std.}}$, which could indicate that spatial correlation is lost due to jittering when it is not adjusted for. The estimates and credible intervals of $\boldsymbol{\beta}$ vary markedly between the models. For both models, only β_{PopC} and β_{DistR} are considered significantly different from 0 at the 5% level (excluding the intercept), because their credible intervals do not contain 0. The largest covariate (in absolute value) is PopC, which has a point estimate that is 25% larger in $\mathbf{M}_{\text{Adj.}}$ relative to the estimate in $\mathbf{M}_{\text{Std.}}$. Based on the findings in the simulation study, we interpret this difference as extra attenuation of the covariate coefficient estimate due to ignoring jittering in $\mathbf{M}_{\text{Std.}}$.

Chapter 8

Discussion

Section 3.1.1 presented the existing approaches to account for positional uncertainty in geostatistical modeling, prior to the numerical approach. The presentation highlighted the subtle difference between CP and FP models for positional error, and made clear the need for a computationally efficient and flexible approach to adjust for positional error. Section 3.1.2 presented approaches to specifically account for positional uncertainty in the spatial covariates, which may be the most important concern in this setting as illustrated by Altay et al. (2022b). Future work should investigate this topic further for various jittering scales and observation likelihoods.

The numerical approach, described in Section 4.2, was extended to handle donut jittering by deriving a new expression for the integration weights and implementing it in code. Inference with the numerical approach is based on several approximations, such as Laplace approximations in TMB, the approximation of the likelihood with numerical integration and treating the posterior of the random effect as a multivariate Gaussian distribution. This motivated an investigation of its accuracy by using a sample-based method.

A sample-based approach to inference was constructed and implemented as described in Section 4.3. We quickly discovered that the main challenge in this setting was to efficiently sample the unknown true locations. This set restrictions on the the number of observations, the spatial range, and the outer and inner jittering radii which we were able to consider in practice. In Chapter 5, we therefore selected two specific scenarios for the investigation, one with moderate disk jittering and the other with moderate donut jittering. Acquiring over 100 ESS of the unknown true locations in these scenarios still necessitated nearly $20,000$ iterations, taking around 20 hours to complete. The investigation demonstrated that the numerical approach may effectively approximate exact inference, and that adjusting for jittering may lead to more accurate inference in terms of the posterior distributions of covariate coefficients, less biased point estimates of the nugget variance and superior predictive measures.

Because of the practical restrictions on the sample-based approach, we continued with the numerical approach to study how moderate and substantial jittering impact inference in a simulation study in Chapter 6. The results were similar for disk and donut jittering and aligned with previous work on DHS data. More jittering, in terms of a larger outer radius, generally gave rise to stronger attenuation in the point estimates of the covariate coefficients in the standard and adjusted model, with considerably more attenuated estimates in the standard model. Addition-

ally, the nugget variance was substantially overestimated in the standard model, which was expected from the theory described in Section 3.3. For example, under substantial donut jittering, the median bias was nearly 10 times greater in the standard model than in the adjusted model. The spatial range was more biased and always underestimated in the standard model with respect to the median, which also agrees with the expectations from theory. The predictive measures were, with the exception of some outlier values, superior for the adjusted model. For example, under substantial donut jittering, the median CRPS was found to be 17% lower and the median prediction RMSE was found to be 18% lower when jittering was adjusted for, relative to ignored.

In the simulation study, certain outlier values for the estimated covariate coefficients and the predictive measures were removed for the adjusted model. The occurrence of these outliers can be attributed to large estimated uncertainty in the covariate coefficients for certain datasets. The uncertainty is estimated with the Hessian of the likelihood which TMB uses in its optimization, implying that the outliers may arise from a flat likelihood surface, leading to a Hessian close to a zero matrix. It is plausible that the exact posteriors of the covariate coefficients exhibit a flat distribution, where multiple values are nearly equally probable. Such scenarios represent a potential problem for the numerical approach, as it could make the optimization unstable. Exploring the exact posteriors using the sample-based approach could have provided valuable insights on this problem, but we were not able to practically sample the unknown true locations with the setup considered in the simulation study. Nevertheless, we believe that for a given dataset, numerical instabilities such as these outliers may be avoided in many cases by increasing the resolution of the mesh or adjusting priors.

Future work is needed to construct an efficient sampler for the unknown true locations when the impact of jittering is substantial, e.g., due to a large jittering scale or strong covariates. We were not able to investigate how the exact posteriors behave in such scenarios, which is important to assess the suitability of the numerical approach in, e.g., the context of DHS data. We therefore encourage more work on developing an efficient sampler for the unknown true locations. Additionally, a natural line of investigation is to compare the inference of the sample-based approach, which uses the SPDE approach to approximate the spatial field, to the inference of a sampler which uses the exact geostatistical model.

We explored how privacy protection can be evaluated by approximating the full posterior of the unknown true locations with Monte Carlo integration and the integration procedure employed in the numerical approach. To evaluate privacy protection, we considered a distance-based and a KLD-based metric. The distance-based metric may not capture the important local behaviour of the full posterior, while it is unclear how the KLD-based metric should be calibrated in order to fairly compare the privacy protection of different jittering schemes. Thus, we encourage more work on this issue, so that jittering schemes employed for geomasking purposes can be justified based on a metric measuring privacy protection.

Based on our results from the investigation with the sample-based approach and the simulation study, in combination with findings of previous work (Altay et al., 2022a,b; Totland, 2022), we recommend that the numerical approach is used to adjust for jittering in geostatistical analyses of DHS data and in other applications with positional uncertainty of a similar spatial scale when there are $\sim 10^3$ obs-

ervations and raster-based covariates are used. Even if this is not the case, adjusting for jittering is fast and easy to apply for arbitrary geographies, as we demonstrated in the case study of vaccination coverage in Chapter 7. Thus, even if the impact of positional uncertainty is believed to be negligible, the numerical approach may be used to check if accounting for it leads to considerably different inference. Furthermore, an R-package has recently been developed (Altay et al., 2023), which makes the numerical approach with adjustment for positional uncertainty readily available, minimizing the amount of code and technical knowledge required from the user.

For future lines of work, the sensitivity of inference to the number of primary and secondary points in the integration scheme should be investigated, as this could substantially reduce computation times by reducing the number of required operations and giving a sparser Hessian in the optimization routine. Importantly, this needs to be investigated in relation to the smoothness of covariates, as this is likely to have a deciding impact on the adequate number of integration points per area. Another line of future work is to extend the numerical approach to space-time models. DHS datasets are typically sampled at regular time intervals, and utilizing the temporal structure of the data could be valuable for forecasting the development of, e.g., health indicators.

Bibliography

W. B. Allshouse, M. K. Fitch, K. H. Hampton, D. C. Gesink, I. A. Doherty, P. A. Leone, M. L. Serre, and W. C. Miller. Geomasking sensitive health data and privacy protection: an evaluation using an E911 database. *Geocarto international*, 25(6):443–452, 2010. doi: https://doi.org/10.1080/10106049.2010.496496.

U. Altay, J. Paige, A. Riebler, and G.-A. Fuglstad. Fast geostatistical inference under positional uncertainty: Analysing DHS household survey data, 2022a. URL `https://arxiv.org/abs/2202.11035`.

U. Altay, J. Paige, A. Riebler, and G.-A. Fuglstad. Jittering Impacts Raster- and Distance-based Geostatistical Analyses of DHS Data, 2022b. URL `https://arxiv.org/abs/2211.07442`.

U. Altay, J. Paige, A. Riebler, and G.-A. Fuglstad. GeoAdjust: Adjusting for Positional Uncertainty in Geostatistial Analysis of DHS Data, 2023. URL `https://arxiv.org/abs/2303.12668`.

M. P. Armstrong, G. Rushton, and D. L. Zimmerman. Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, 18(5):497–525, 1999. doi: https://doi.org/10.1002/(sici)1097-0258(19990315)18:5⟨497::aid-sim45⟩3.0.co;2-#. URL `https://pubmed.ncbi.nlm.nih.gov/10209808/`.

J. J. Barber, A. E. Gelfand, and J. A. Silander. Modelling Map Positional Error to Infer True Feature Location. *The Canadian Journal of Statistics / La Revue Canadienne de Statistique*, 34(4):659–676, 2006. URL `http://www.jstor.org/stable/20445227`.

T. Bayes. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, 53:370–418, 1763.

C. R. Burgert, J. Colston, T. Roy, and B. Zachary. Geographic displacement procedure and georeferenced datarelease policy for the Demographic and Health Surveys. 2013. URL `https://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf`. DHS Spatial Analysis Reports No. 7.

K. P. Burnham and D. R. Anderson, editors. *Model Selection and Multimodel Inference*. Springer New York, NY, 2002. doi: https://doi.org/10.1007/b97636.

B. P. Carlin and T. A. Louis. Empirical Bayes: Past, Present and Future. *Journal of the American Statistical Association*, 95(452):1286–1289, 2000. URL `http://www.jstor.org/stable/2669771`.

R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective*. Chapman and Hall/CRC, New York, 2006. doi: https://doi.org/10.1201/9781420010138.

C. A. Cassa, S. J. Grannis, J. M. Overhage, and K. D. Mandl. A context-sensitive approach to anonymizing spatial surveillance data: impact on outbreak detection. *Journal of the American Medical Informatics Association*, 13(2):160–5, 2006. doi: https://doi.org/10.1197/jamia.M1920.

N. Cressie and J. Kornak. Spatial Statistics in the Presence of Location Error with an Application to Remote Sensing of the Environment. *Statistical Science*, 18(4): 436–456, 2003. doi: https://doi.org/10.1214/ss/1081443228.

N. Cressie and C. K. Wikle. *Statistics for Spatio-temporal Data*. John Wiley Sons, Inc., Hoboken, New Jersey, 2011.

P. J. Diggle and P. J. Ribeiro. *Model-based Geostatistics*. Springer New York, NY, 2007. doi: https://doi.org/10.1007/978-0-387-48536-2.

EPSG. EPSG:26393 Minna / Nigeria East Belt. `https://epsg.io/26393`, n.d. Accessed: 2023-02-24.

European Parliament, Council of the European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). *Official Journal of the European Union*, (119):1–88, 2016. URL `https://eur-lex.europa.eu/eli/reg/2016/679/oj`.

T. R. Fanshawe and P. J. Diggle. Spatial Prediction in the Presence of Positional Error. *Environmetrics*, 22(2):109–122, 2011. doi: https://doi.org/10.1002/env.1062. URL `https://onlinelibrary.wiley.com/doi/abs/10.1002/env.1062`.

C. Fronterrè, E. Giorgi, and P. J. Diggle. Geostatistical inference in the presence of geomasking: A composite-likelihood approach. *Spatial Statistics*, 28:319–330, 2018. doi: https://doi.org/10.1016/j.spasta.2018.06.004. URL `https://www.sciencedirect.com/science/article/pii/S2211675317302737`.

G.-A. Fuglstad, D. Simpson, F. Lindgren, and H. Rue. Constructing Priors that Penalize the Complexity of Gaussian Random Fields. *Journal of the American Statistical Association*, 114(525):445–452, 2019. doi: https://doi.org/10.1080/01621459.2017.1415907.

G.-A. Fuglstad, Z. R. Li, and J. Wakefield. The Two Cultures for Prevalence Mapping: Small Area Estimation and Spatial Statistics, 2022. URL `https://arxiv.org/abs/2110.09576`.

J. Gabrosek and N. Cressie. The Effect on Attribute Prediction of Location Uncertainty in Spatial Data. *Geographical Analysis*, 34(3):262–285, 2002. doi: https://doi.org/10.1111/j.1538-4632.2002.tb01088.x.

GADM. GADM (version 4.1). `https://gadm.org/download_country.html`, 2022. Accessed: 2022-08-24.

T. Gneiting and A. Raftery. Strictly Proper Scoring Rules, Prediction, and Estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007. doi: https://doi.org/10.1198/016214506000001437.

V. Gómez-Rubio and H. Rue. Markov chain Monte Carlo with the Integrated Nested Laplace Approximation. *Statistics and Computing*, 28:1033–1051, 2018. doi: https://doi.org/10.1007/s11222-017-9778-y.

K. H. Hampton, M. K. Fitch, I. A. Doherty W. B. Allshouse, D. C. Gesink, P. A. Leone, M. L. Serre, and W. C. Miller. Mapping health data: improved privacy protection with donut method geomasking. *American Journal of Epidemiology*, 172(9):1062–1069, 2010. doi: https://doi.org/10.1093/aje/kwq248.

W. K. Hastings. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika*, 57(1):97–109, 1970. URL `http://www.jstor.org/stable/2334940`.

A. Jordan, F. Krüger, and S. Lerch. Evaluating Probabilistic Forecasts with scoringRules. *Journal of Statistical Software*, 90(12):1–37, 2019. doi: https://www.doi.org/10.18637/jss.v090.i12.

K. Kristensen, A. Nielsen, C. W. Berg, H. Skaug, and B. M. Bell. TMB: Automatic Differentiation and Laplace Approximation. *Journal of Statistical Software*, 70 (5):1–21, 2016. doi: https://www.doi.org/10.18637/jss.v070.i05. URL `https://www.jstatsoft.org/index.php/jss/article/view/v070i05`.

S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951. doi: https://doi.org/10.1214/aoms/1177729694.

P. C. Kyriakidis and A. G. Journel. Geostatistical Space–Time Models: A Review. *Mathematical Geology*, 31(6):651–684, 1999. doi: https://doi.org/10.1023/A:1007528426688.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(4):423–498, 2011. doi: https://doi.org/10.1111/j.1467-9868.2011.00777.x. URL `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x`.

G. Lindgren. *Stationary Stochastic Processes: Theory and Applications (1st ed.)*. Chapman and Hall/CRC, Boca Raton, FL, 2012. doi: https://doi.org/10.1201/b12171.

J. S. Liu. The Collapsed Gibbs Sampler in Bayesian Computations with Applications to a Gene Regulation Problem. *Journal of the American Statistical Association*, 89(427):958–966, 1994. URL `http://www.jstor.org/stable/2290921`.

N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 12 2004. doi: 10.1063/1.1699114. URL https://doi.org/10.1063/1.1699114.

J. Møller, editor. *Spatial Statistics and Computational Methods*. Springer New York, NY, 2003. doi: https://doi.org/10.1007/978-0-387-21811-3.

National Population Commission and ICF. Nigeria Demographic and Health Survey 2018 - Final Report. 2019. URL http://dhsprogram.com/pubs/pdf/FR359/FR359.pdf.

A. Osgood-Zimmerman and J. Wakefield. A Statistical Introduction to Template Model Builder: A Flexible Tool for Spatial Modeling, 2021. URL https://arxiv.org/abs/2103.09929.

C. Perez-Heydrich, J. L. Warren, C. R. Burgert, and M. E. Emch. Influence of Demographic and Health Survey Point Displacements on Raster-Based Analyses. *Spatial Demography*, 4(2):135–153, 2016. doi: https://doi.org/10.1007/s40980-015-0013-1.

M. Plummer, N. Best, K. Cowles, and K. Vines. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*, 6(1):7–11, 2006. URL https://journal.r-project.org/archive/.

C. P. Robert and G. Casella. *Monte Carlo Statistical Methods*. Springer New York, NY, 2004. doi: https://doi.org/10.1007/978-1-4757-4145-2.

G. O. Roberts and A. F. M. Smith. Simple conditions for the convergence of the Gibbs sampler and Metropolis-Hastings algorithms. *Stochastic Processes and their Applications*, 49(2):207–216, 1994. doi: https://doi.org/10.1016/0304-4149(94)90134-1. URL https://www.sciencedirect.com/science/article/pii/0304414994901341.

H. Rue and L. Held. *Gaussian Markov Random Fields: Theory and Applications*. Chapman and Hall/CRC, Boca Raton, FL, 2005. doi: https://doi.org/10.1201/9780203492024.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian Inference for Latent Gaussian Models by Using Integrated Nested Laplace Approximations. *Journal of the Royal Statistical Society Series B*, 71:319–392, 04 2009. doi: https://www.doi.org/10.1111/j.1467-9868.2008.00700.x.

R. Sato, O. A. Makinde, K. C. Daam, and B. Lawal. Geographical and time trends of measles incidence and measles vaccination coverage and their correlation in Nigeria. *Human Vaccines & Immunotherapeutics*, 18(6), 2022. doi: https://doi.org/10.1080/21645515.2022.2114697.

D. Simpson, F. Lindgren, and H. Rue. Think continuous: Markovian Gaussian models in spatial statistics. *Spatial Statistics*, 1:16–29, 2012. doi: https://doi.org/10.1016/j.spasta.2012.02.003. URL https://www.sciencedirect.com/science/article/pii/S2211675312000048.

D. Simpson, H. Rue, A. Riebler, T. G. Martins, and S. H. Sørbye. Penalising Model Component Complexity: A Principled, Practical Approach to Constructing Priors. *Statistical Science*, 32(1):1–28, 2017. doi: https://doi.org/10.1214/16-STS576.

L. Sweeney. K-Anonymity: A Model for Protecting Privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002. doi: https://doi.org/10.1142/S0218488502001648.

The DHS Program. DHS Overview. `https://dhsprogram.com/Methodology/Survey-Types/DHS.cfm`, n.d. Accessed: 2023-05-08.

J. Totland. *Inference under Jittering in Geostatistical Models*. Department of Mathematical Sciences, NTNU. Project thesis, 2022.

UN. Transforming Our World: The 2030 Agenda for Sustainable Development, 2015. URL `https://wedocs.unep.org/20.500.11822/9814`.

UN, Department of Economic and Social Affairs. World Population Prospects: The 2022 Revision. `https://population.un.org/wpp/`, 2022. Accessed: 2023-05-26.

J. L. Warren, C. Perez-Heydrich, C. R. Burgert, and M. E. Emch. Influence of Demographic and Health Survey Point Displacements on Distance-Based Analyses. *Spatial Demography*, 4(2):155–173, 2016. doi: https://doi.org/10.1007/s40980-015-0014-0.

K. Wilson and J. Wakefield. Estimation of health and demographic indicators with incomplete geographic information. *Spatial and Spatio-temporal Epidemiology*, 37:100421, 2021. doi: https://doi.org/10.1016/j.sste.2021.100421. URL `https://www.sciencedirect.com/science/article/pii/S1877584521000204`.

WorldPop, Center for International Earth Science Information Network (CIESIN), and Columbia University. Global High Resolution Population Denominators Project - Funded by The Bill and Melinda Gates Foundation (OPP1134076). `https://hub.worldpop.org/doi/10.5258/SOTON/WP00645`, 2018.

P. A. Zandbergen. Ensuring Confidentiality of Geocoded Health Data: Assessing Geographic Masking Strategies for Individual-Level Data. *Advances in Medicine*, Article 567049, 2014. doi: https://doi.org/10.1155/2014/567049.

Appendices

A Derivation of Variogram under Geomasking

Here, we derive the expression for $\mathrm{E}\left[\rho(u^*_{ij}) \mid u_{ij}\right]$ presented in Equation (3.4). For ease of notation, let $\phi = \rho_{\mathrm{S}}/\sqrt{8}$. We use a change of variables, and define

$$\tilde{u}_{ij} = \frac{\phi}{\sqrt{4\delta^2 + \phi^2}}, \quad \tilde{u}^*_{ij} = \frac{\sqrt{4\delta^2 + \phi^2}}{\phi}$$

.

Note that $\tilde{u}^*_{ij}\tilde{u}_{ij} = u^*_{ij}u_{ij}$. Then,

$$\begin{aligned}\mathrm{E}\left[\rho(u^*_{ij}) \mid u_{ij}\right] &= \int_0^\infty \rho(u^*_{ij}) \cdot \pi(u^*_{ij}|u_{ij})\mathrm{d}u^*_{ij} \qquad \text{(A.1)}\\ &= \int_0^\infty \exp\left\{-\frac{(u^*_{ij})^2}{\phi^2}\right\} \cdot \exp\left\{-\frac{(u^*_{ij})^2 + u^2_{ij}}{4\delta^2}\right\} \frac{u^*_{ij}}{2\delta^2}\, I_0\left(\frac{u^*_{ij}u_{ij}}{2\delta^2}\right) \mathrm{d}u^*_{ij},\end{aligned}$$

where I_0 is the modified Bessel function of the first kind, order zero. Here, we have inserted the Gaussian correlation function and the pdf of a Rice$(u_{ij}, \sqrt{2}\delta)$, whose expression can be found in Appendix A of Fronterrè et al. (2018). Continuing, we perform the change of variables in (A.1), which gives

$$\mathrm{E}\left[\rho(u^*_{ij}) \mid u_{ij}\right] = \tilde{C} \cdot \int_0^\infty \exp\left\{-\frac{(\tilde{u}^*_{ij})^2 + \tilde{u}^2_{ij}}{4\delta^2}\right\} \frac{\tilde{u}^*_{ij}}{2\delta^2}\, I_0\left(\frac{\tilde{u}^*_{ij}\tilde{u}_{ij}}{2\delta^2}\right) \mathrm{d}\tilde{u}^*_{ij}, \qquad \text{(A.2)}$$

where

$$\tilde{C} = \frac{\phi^2}{\phi^2 + 4\delta^2} \exp\left\{-\frac{u^2_{ij}}{(\phi^2 + 4\delta^2)}\right\} = \frac{1}{1 + 4r^2} \exp\left\{-\frac{u^2_{ij}}{\phi^2(1 + 4r^2)}\right\},$$

where $r = \delta/\phi$. The integral in (A.2) integrates the pdf of a Rice$(\tilde{u}_{ij}, 2\delta^2)$ distribution over its support, which evaluates to one. Thus,

$$\mathrm{E}\left[\rho(u^*_{ij}) \mid u_{ij}\right] = \frac{1}{1 + 4r^2} \exp\left\{-\frac{u^2_{ij}}{\phi^2(1 + 4r^2)}\right\}.$$

B Derivation of Integration Weights under Donut Jittering

We derive an expression for the integration weights under donut jittering. For generality, we allow the inner donut radius, $R^{(i)}_1 > 0$, and the outer donut radius,

$R_2^{(i)} > R_2^{(i)}$, to vary between observations, although we treat these as common for all observations elsewhere in the text. The integration weights depend on the probability mass of the donut jittering distribution within their corresponding integration area. Therefore, first, we define the area of the ijk-th integration point as

$$A_{ijk} = \{\boldsymbol{s}_i + (r\cos a, r\sin a)^{\mathrm{T}} : r_{i(j-1)} \leq r < r_{ij},\ a_{ij(k-1)} \leq a < a_{ijk}\},$$

where $r_{i0} = R_1^{(i)}$ and $r_{iR^{(i)}} = R_2^{(i)}$, and we let $L_i = R_2^{(i)} - R_1^{(i)}$. We take the integration area boundaries as equispaced,

$$a_{ijk} = \begin{cases} \frac{2\pi}{m_{ij}}(k-1) + \frac{\pi}{m_{ij}}, & j \bmod 2 = 1, \\ \frac{2\pi}{m_{ij}}(k-1), & \text{otherwise}, \end{cases}$$

where $\frac{\pi}{m_{ij}}$ intersperses the integration points for every other ring based on m_{ij}. Next, we determine the radii r_{ij} so the integration weights are equal for each of the integration points. Specifically, for the outer-most ring, since the density $\pi(\boldsymbol{s}_i|\boldsymbol{s}_i^*)$ is uniform in radial coordinates on $(a, r) \in [0, 2\pi] \times [0, L_i]$, we have

$$\frac{2\pi L_i}{\sum_{j'=1}^{R^{(i)}} m_{ij'}} = \frac{2\pi(R_2^{(i)} - r_{i(R^{(i)}-1)})}{m_{iR^{(i)}}} \implies r_{i(R^{(i)}-1)} = \frac{R_1^{(i)} m_{iR^{(i)}} + R_2^{(i)} \sum_{j'=1}^{R^{(i)}-1} m_{ij'}}{\sum_{j'=1}^{R^{(i)}} m_{ij'}}.$$

Then, one can show, e.g., by induction, that this leads to

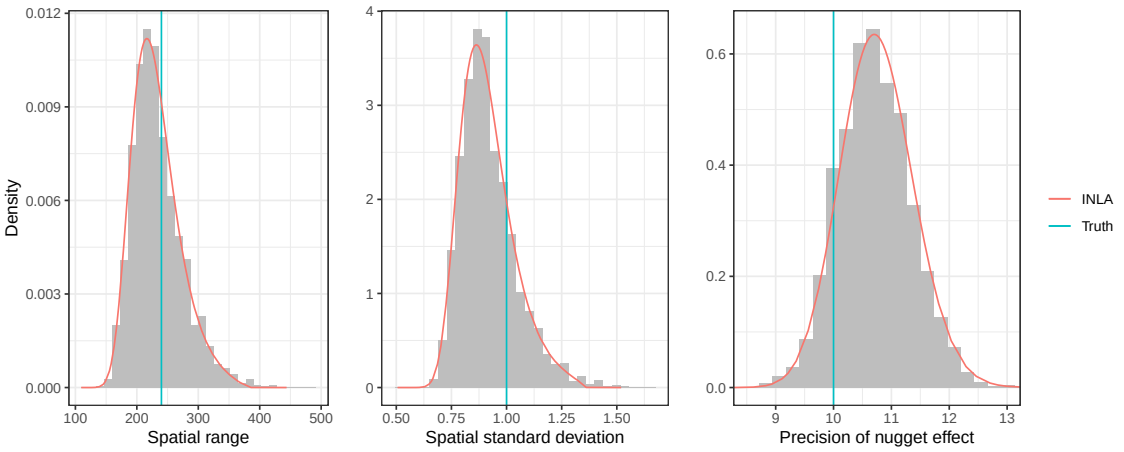
$$r_{ij} = \frac{R_1^{(i)} \sum_{j'=j+1}^{R^{(i)}} m_{ij'} + R_2^{(i)} \sum_{j'=1}^{j} m_{ij'}}{\sum_{j'=1}^{R^{(i)}} m_{ij'}}.$$

The integration weights are given by

$$\lambda_{ijk} \propto \frac{r_{ij} - r_{i(j-1)}}{L_i} \frac{a_{ijk} - a_{ij(k-1)}}{2\pi}.$$

C Verifying the MCMC Algorithm Implementation with INLA

Implementing MCMC samplers often requires a moderate amount of programming, and it can be challenging to validate the programmed algorithm, since the true posteriors are, naturally, not available analytically. To verify the implementation of the MCMC sampler described in Section 4.3, we used the R-package INLA to estimate marginal posteriors of the model parameters when there is no positional uncertainty.



(a) ρ_{S} (b) σ_{S} (c) $1/\sigma_{\mathrm{N}}^2$

Figure C.1: Marginal posteriors of INLA (red) and the corresponding sample-distributions of the MCMC algorithm (grey). The true values used to simulate the data are shown as blue vertical lines.

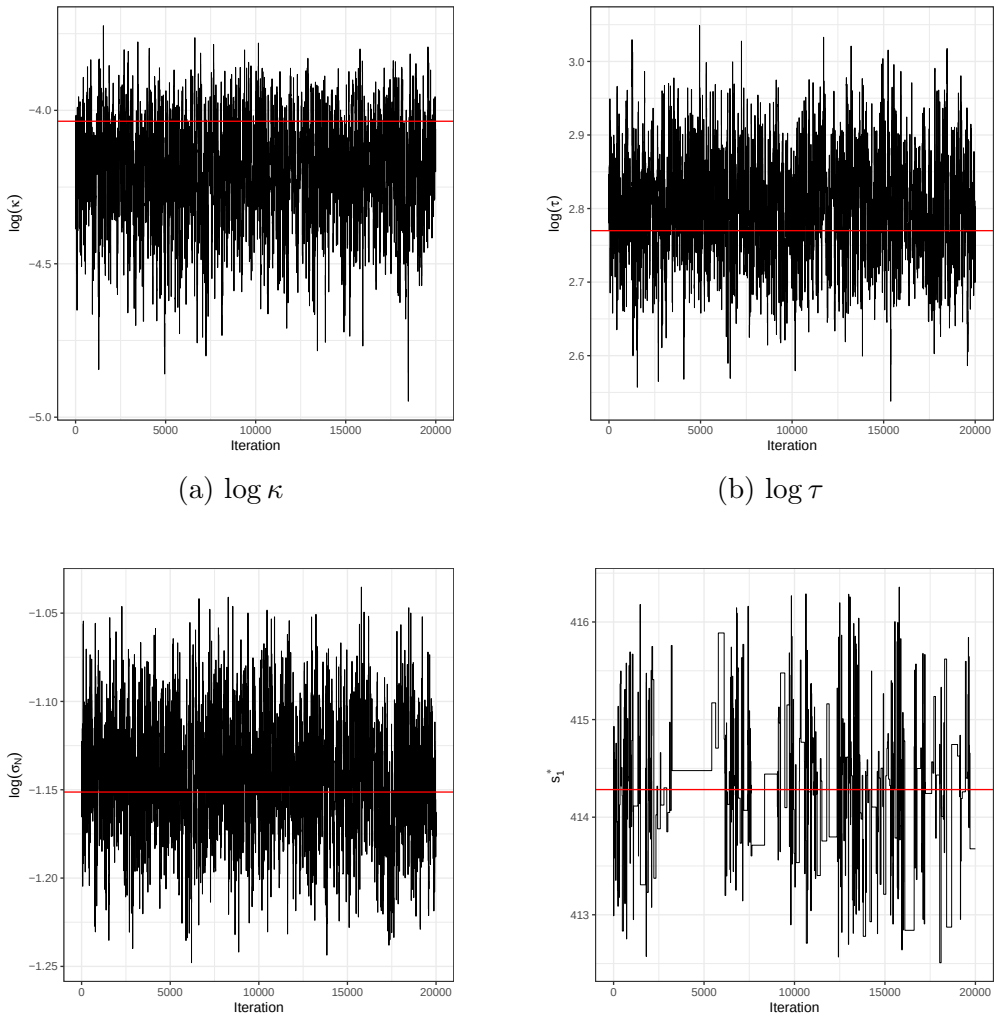
We considered the geostatistical model given by (2.6) with only an intercept, $\boldsymbol{\beta} = 1$, and otherwise the same parameters and domain as that in Scenario 2 in Chapter 5. The MCMC algorithm was run for $15,000$ iterations and 1000 burn-in samples were used. Figure C.1 displays the marginal parameter posteriors produced by INLA along with the corresponding sample-distributions MCMC algorithm. We observe that the histograms of the MCMC algorithm match the marginal posteriors of INLA, which strengthens our confidence in that the implementation of the MCMC algorithm in the sample-based approach is correct.

D Supplementary Figures to Chapter 5

D.1 Scenario 1

Figure D.1 displays trace-plots for $\log \kappa$, $\log \tau$, $\log \sigma_{\mathrm{N}}$ and the first coordinate of a true location, $\boldsymbol{s}^* = (s_1^*, s_2^*)^{\mathrm{T}}$ for one dataset used in Scenario 1. The trace-plot in Figure D.1d of the true coordinate illustrates poor mixing, which makes long chains necessary in order to achieve accurate posteriors. Trace-plots for the other true coordinates were similar, and are therefore not shown. Nonetheless, the trace-plots do not give any evidence that the chain has not converged, and 1000 burn-in samples seems to be more than enough. We do not show trace-plots for more datasets, because they exhibit very similar behaviour.

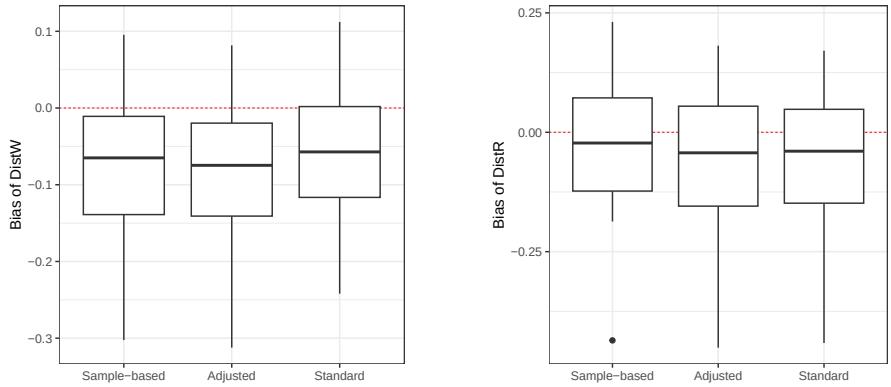
Figure D.2 displays the cross-dataset biases of β_{istW}, β_{DistR} and σ_{S}^2 in Scenario 1, for which there are no large difference between $\mathbf{M}_{\mathrm{Std.}}$, $\mathbf{M}_{\mathrm{Adj.}}$ and $\mathbf{M}_{\mathrm{SB}}$. Figure D.3 displays the posterior distributions of β_{DistW} and β_{DistR} for $\mathbf{M}_{\mathrm{Adj.}}$, $\mathbf{M}_{\mathrm{Std.}}$ and $\mathbf{M}_{\mathrm{SB}}$ for a given dataset, where we, as expected, observe that there is little difference between the models.



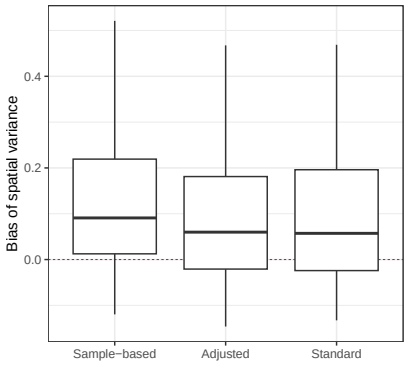
(a) $\log \kappa$ (b) $\log \tau$

(c) $\log \sigma_{\mathrm{N}}$ (d) s_1^*

Figure D.1: Trace-plots from the MCMC algorithm for one dataset in Scenario 1 for (a) $\log \kappa$, (b) $\log \tau$, (c) $\log \sigma_{\mathrm{N}}$ and (d) the first coordinate of a true location, s_1^*. The red horizontal lines indicate the true values used to simulate the dataset.

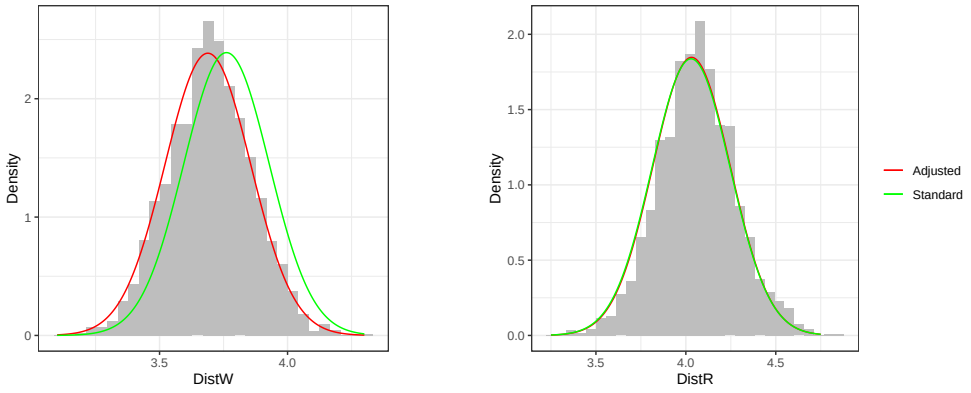


(a) β_{DistW} (b) β_{DistR}



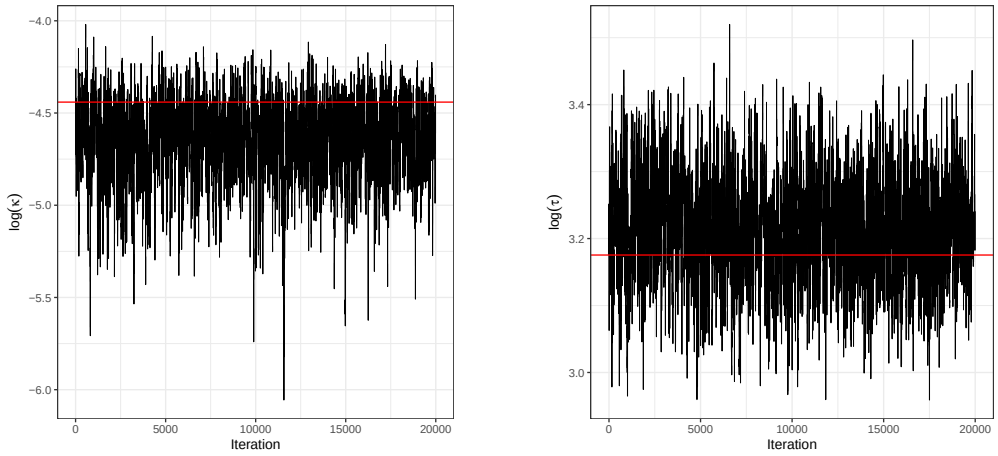
(c) σ_{S}^2

Figure D.2: Cross-dataset biases of point estimates of (a,b) covariate coefficient estimates and (c) the spatial marginal variance in $\mathbf{M}_{\text{Adj.}}$, $\mathbf{M}_{\text{Std.}}$ and $\mathbf{M}_{\text{SB}}$ in Scenario 1.



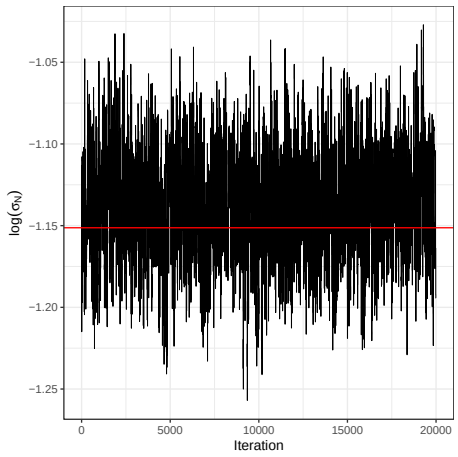
(a) β_{DistW} (b) β_{DistR}

Figure D.3: The estimated posterior distribution for two of the covariate coefficients in $\mathbf{M}_{\text{Std.}}$ (green) and $\mathbf{M}_{\text{Adj.}}$ (red) for a given dataset in Scenario 1. The histograms display the sample-based posteriors of $\mathbf{M}_{\text{SB}}$.

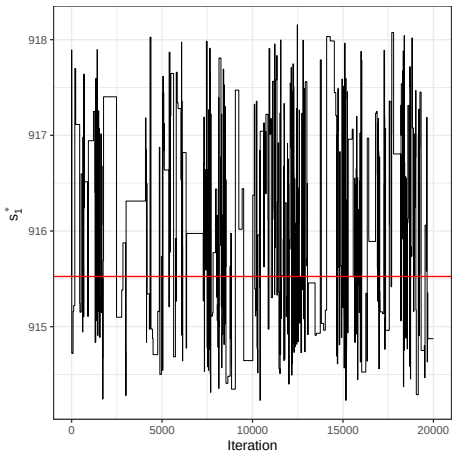


(a) $\log \kappa$

(b) $\log \tau$



(c) $\log \sigma_{\mathrm{N}}$



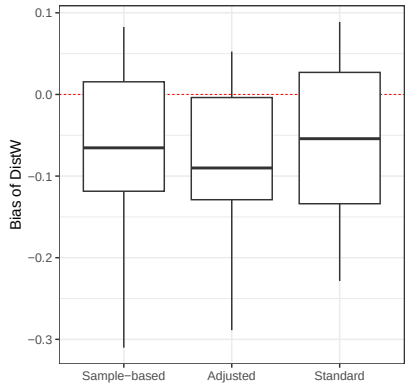
(d) s_1^*

Figure D.4: Trace-plots from the MCMC algorithm for one dataset in Scenario 2 for (a) $\log \kappa$, (b) $\log \tau$, (c) $\log \sigma_{\mathrm{N}}$ and (d) the first coordinate of a true location, s_1^*. The red horizontal lines indicate the true values used to simulate the dataset.

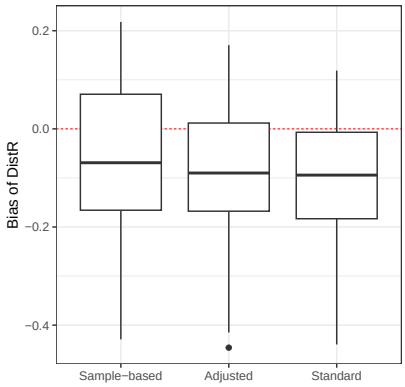
D.2 Scenario 2

Figure D.1 displays trace-plots for $\log \kappa$, $\log \tau$, $\log \sigma_{\mathrm{N}}$ and the first coordinate of one true location, $\boldsymbol{s}^* = (s_1^*, s_2^*)^{\mathrm{T}}$ for one dataset used in Scenario 2. The trace-plot in Figure D.4d of the true coordinate illustrates poor mixing, which makes long chains necessary in order to achieve accurate posteriors. Trace-plots for the other true coordinates were similar, and are therefore not shown. The trace-plots do not give any evidence that the chain has not converged, and 1000 burn-in samples appears to be more than enough. We do not show trace-plots for more datasets, because they display very similar behaviour.

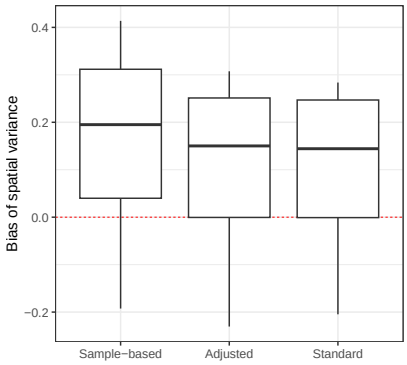
Figure D.5 displays the cross-dataset biases of β_{istW}, β_{DistR} and σ_{S}^2 in Scenario 2, for which there are no large difference between $\mathbf{M}_{\mathrm{Std.}}$, $\mathbf{M}_{\mathrm{Adj.}}$ and $\mathbf{M}_{\mathrm{SB}}$. Figure D.6 displays the posterior distributions of β_{DistW} and β_{DistR} for $\mathbf{M}_{\mathrm{Std.}}$, $\mathbf{M}_{\mathrm{Adj.}}$ and $\mathbf{M}_{\mathrm{SB}}$ for a given dataset, where we observe that there is little difference between the models.



(a) β_{DistW}

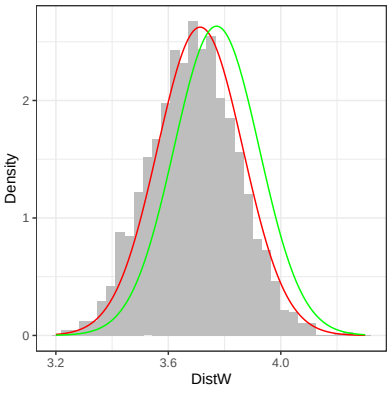


(b) β_{DistR}

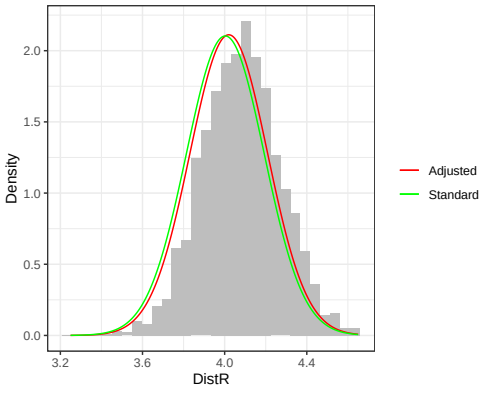


(c) σ_{S}^2

Figure D.5: Cross-dataset biases of point estimates of (a,b) covariate coefficient estimates and (c) spatial marginal variance for $\mathbf{M}_{\text{Std.}}$, $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{SB}}$ in Scenario 2.

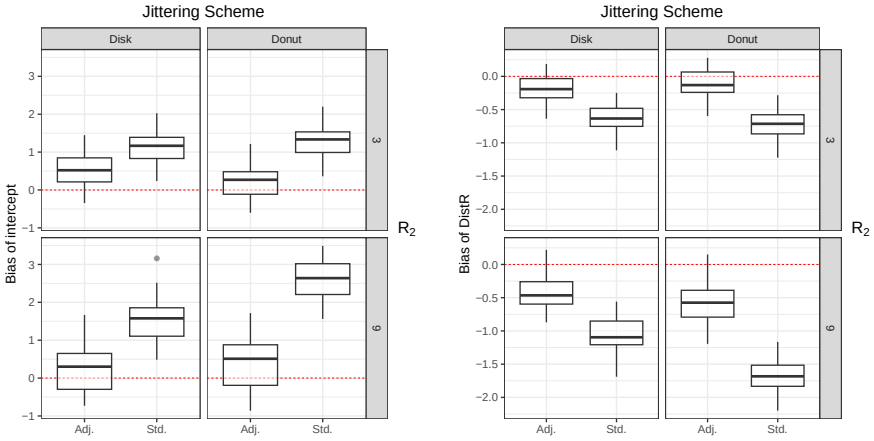


(a) β_{DistW}



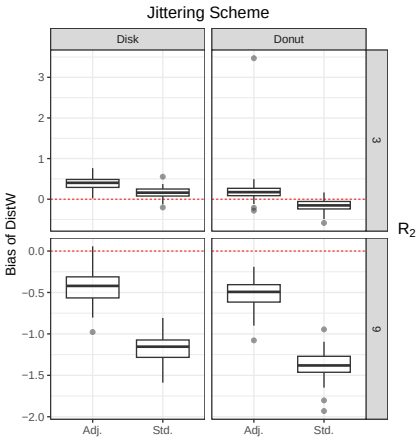
(b) β_{DistR}

Figure D.6: The estimated posterior distribution for two of the covariate coefficients in $\mathbf{M}_{\text{Std.}}$ (green) and $\mathbf{M}_{\text{Adj.}}$ (red) for a given dataset in Scenario 2. The histograms display the sample-based posteriors of $\mathbf{M}_{\text{SB}}$.

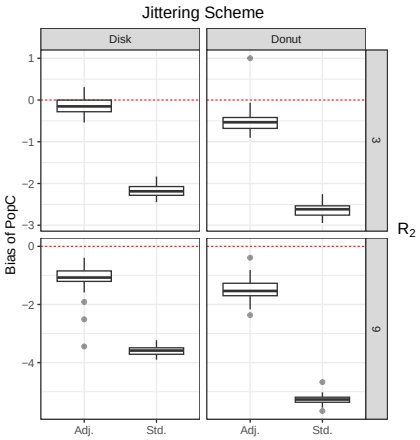


(a) μ

(b) β_{DistR}



(c) β_{DistW}

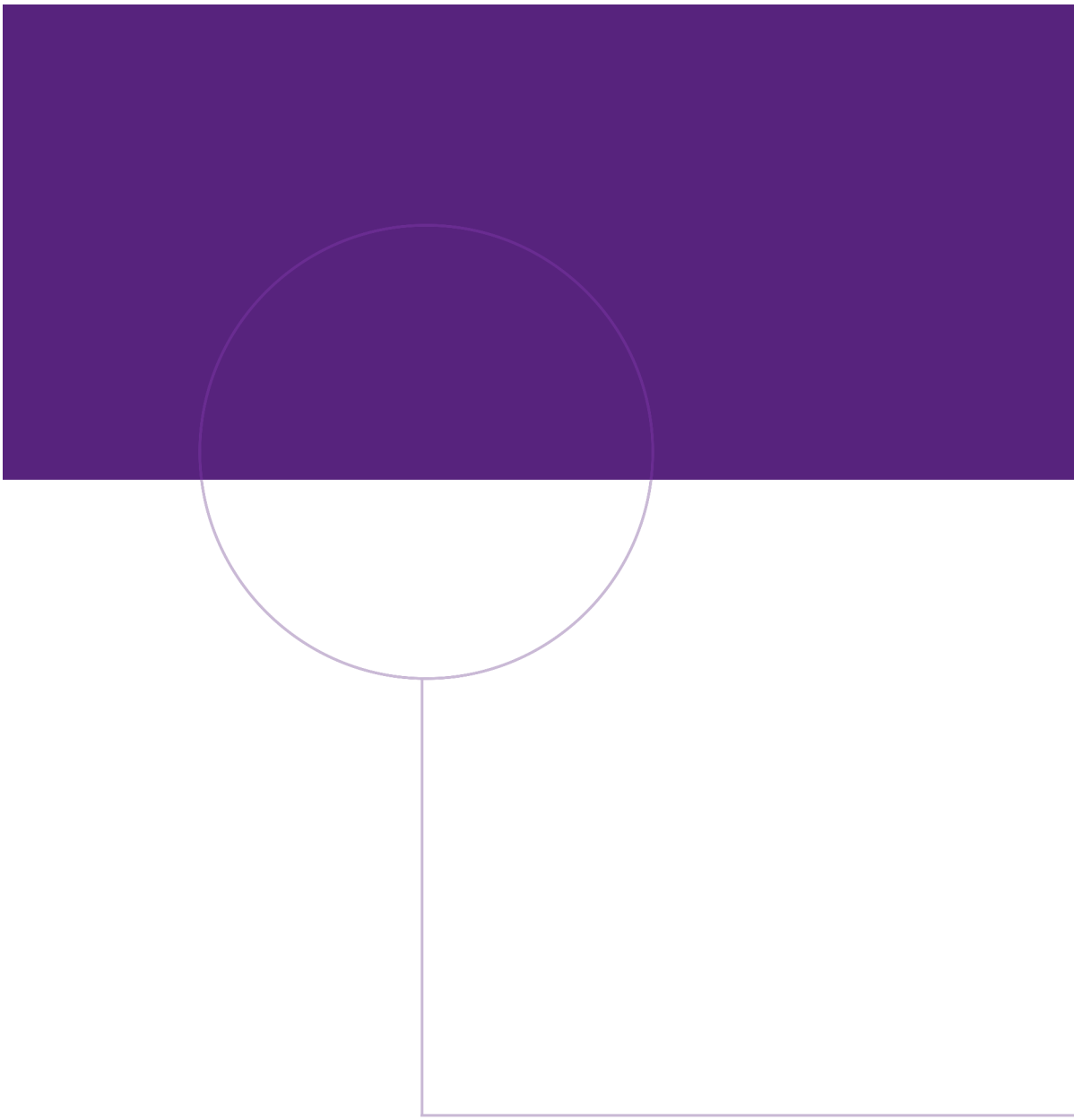


(d) β_{PopC}

Figure E.1: The cross-dataset biases in $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ for the covariate coefficients given by (a) μ, (b) β_{DistR}, (c) β_{DistW} and (d) β_{PopC}, under all jittering schemes in the simulation study.

E Supplementary Figures to Chapter 6

Figure E.1 displays the cross-dataset bias corresponding to the covariate coefficient in $\mathbf{M}_{\text{Adj.}}$ and $\mathbf{M}_{\text{Std.}}$ that were not shown in the simulation study in Chapter 6. With the exception of β_{DistW} when $R_2 = 3$, we observe that the covariate coefficients (excluding the intercept) are underestimated, and more so in $\mathbf{M}_{\text{Std.}}$ than in $\mathbf{M}_{\text{Adj.}}$. This effect is amplified for increased jittering ($R_2 = 9$).



NTNU

Norwegian University of
Science and Technology