Statistical Methods for Geospatial Modeling with Stratified Cluster Survey Data

John Paige

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2020

Reading Committee:

Jon Wakefield, Chair

Adrian Dobra

Geir-Arne Fuglstad

Program Authorized to Offer Degree:
Department of Statistics

University of Washington

Abstract

Statistical Methods for Geospatial Modeling with Stratified Cluster Survey Data

John Paige

Chair of the Supervisory Committee:
Professor Jon Wakefield
Statistics and Biostatistics

The production of fine-scale, pixel level maps have become increasingly common in the current era of precision public health. This has led to the use of cluster level spatial models by major organizations such as WorldPop and the Institute for Health and Metrics Evaluation. However, many of these models were originally developed in the context of environmental applications, and, when estimating health and demographic indicators in low and middle income countries, they are frequently applied to demographic data from complex, multi-stage household surveys, leaving the potential for both biased and anticonservative estimates unless adjustment for the design is carried out. We highlight three potential problems. First, survey stratification and cluster level variation are often not accounted for. Second, the cluster level models either do not fully account for the population census frame, or completely ignore this aspect. This is made more problematic by confusion between the terms 'prevalence' and 'risk'. Third, even if stratification is accounted for in the cluster level spatial model, if that model is continuously indexed in space, then it often becomes necessary to infer what stratification level is associated with each spatial location or enumeration area (EA) when aggregating predictions, which is inevitably inexact. However, little work has been done to identify how stratification misclassification can impact predictions and how to produce predictions that are more robust to this problem.

In this thesis, we investigate a variety of issues relevant to the use of cluster level data for estimating demographic outcomes continuously through space, and also as aggregates over administrative areas. We focus on the three problems mentioned above when estimating the neonatal mortality rate and secondary education completion for women aged 20–29 in Kenya using the 2014 Kenya Demographic Health Survey (DHS). First, we explore methods in small area estimation that can account for survey stratification, proposing models that include stratum level fixed effects for cluster-indexed spatial models. Second, we propose a general framework capable of accounting for cluster level, population denominator, and population-level variation as well as some aspects of EA location uncertainty. We call a model in this framework a combined population aggregation model (CPAM) since they are formed by combining standard cluster level risk models with an aggregation model for producing areal estimates. We propose a CPAM that, for Admin2 level areal estimates, produces estimates with substantially more uncertainty in the resulting neonatal mortality population aggregates of prevalence, total deaths, and relative prevalence between urban and rural areas, and at only moderately increased computational expense. The proposed CPAM is the first continuous spatial model accounting for the population census frame, cluster level variation, and population numerator and denominator variation when estimating prevalence, total counts, and relative prevalence in urban versus rural parts of an area. Lastly, we develop a Bayesian extension to the popular LatticeKrig model, which we call extended LatticeKrig (ELK). ELK allows for flexible, multiscale spatial dependence and non-Gaussian responses. We show this model holds particular promise for predicting areal aggregates due to its ability to flexibly model the covariance structure, and is more robust than traditional stochastic partial differential equation methods when accounting for confounding factors such as urbanicity.

TABLE OF CONTENTS

LIST OF FIGURES

GLOSSARY

IID: Independent and identically distributed

DHS: Demographic Health Survey

KDHS: Kenya Demographic Health Survey

IHME: Institute for Health and Metrics Evaluation

SAE: Small area estimation

PPS: Probability proportional to size

SRS: Simple random sample

EA: Enumeration area

LMIC: Low and middle income country

VR: Vital registration

AIS: AIDS Indicator Surveys

MICS: Multiple indicator cluster survey

LSMS: Living Standard Measurement Survey

SDG: Sustainable Development Goal

U5MR: Under-5 mortality rate

NMR: Neonatal mortality rate

MCMC: Markov chain Monte Carlo

INLA: Integrated nested Laplace approximations

HT: Horvitz-Thompson

ICAR: Intrinsic conditional autoregressive

BYM: Besag-York-Mollié

BYM2: Besag-York-Mollié 2

GP: Gaussian process

SPDE: Stochastic partial differential equation

LK: LatticeKrig

ELK: Extended LatticeKrig

M-RA: Multi-resolution approximation

PC: Penalized complexity

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Professor Jon Wakefield, for his mentorship throughout my time working with him at the University of Washington. He helped me grow both as a professional and as a person, and often with the assistance of a remarkable suite of jokes and stories that always illustrate his point perfectly, and for which I will always be grateful.

I would also like to thank the members of my committee that have given me advice and feedback, and have been a tremendous help throughout my time in grad school: Professors Jon Wakefield, Geir-Arne Fuglstad, Adrian Dobra, Peter Guttorp, Jim Thorson, and Darryl Holman. Among the committee members, Geir-Arne has been like a second advisor, and always seems to have a unique perspective that changes the way I think about the problems we have worked on together. In addition to Geir-Arne, and Jon, Professor Adrian Dobra was also on the reading committee. I am grateful for all of the time they have put into assisting me with the editing process as well as their very thoughtful advice. Professor Peter Guttorp is the reason I joined this program, and I will always be thankful for his consistent encouragement and the inspiration he gives to students interested in spatial statistics. Professors Jim Thorson and Darryl Holman have always had thoughtful questions that have helped me to think about potential avenues for future work and the larger context of the problems I have worked on.

In addition to Jon and Geir-Arne, Professor Andrea Riebler has been like an advisor to me, and I am very grateful for the opportunities she have extended to me at the Norwegian University of Science and Technology. I am excited to continue to work with her and Geir-

Arne in Norway.

At the University of Washington, I would like to thank Professor David Schmidt of the Department of Earth and Space Sciences for his assistance (in addition to Peter Guttorp) before working with Jon.

I would also like to thank soon to be doctors Bryan Martin, Aaron Osgood-Zimmerman, and Jessica Godwin for their assistance with my miscellaneous questions and thought experiments throughout my time at the University of Washington, whether they were related to statistics, computation, or demography.

I would like to extend my gratitude to Professors Victor Addona and Alicia Johnson of Macalester College for inspiring me to go into the field of Statistics. I will never forget your excellent teaching, and your openness to both friendly conversation and academic questions, and it has changed the course of my career considerably. I would also like to give credit and thanks to Professors Doug Nychka and Dorit Hammerling of the Colorado School of Mines and Dr. Daniel Feldman of Lawrence Berkeley National Lab for inspiring me to study spatial statistics and statistical computation, two interests that have remained important to me since our time working together.

Lastly, I would like to thank my friends and family, who have consistently supported and encouraged me throughout times both good and bad. You have helped to shape the value I place in all things, including the importance in the pursuit and exploration of truth, helping to make the world a better place, and honesty and humility. For that I will always be grateful.

DEDICATION

to my late father, Robert

Chapter 1

INTRODUCTION

1.1 Motivation

The United Nations have outlined a number of goals for countries known as the Sustainable Development Goals (SDGs) (United Nations, 2020) for countries to achieve by 2030. These particular goals have been set since they are vital for the countries' development from health, economic, social, and other standpoints, and also outline a number of metrics by which progress towards these goals can be measured and assessed. Moreover, it is important for inequality in progress towards these goals across spatial and demographic lines to be eliminated.

This dissertation will focus on neonatal (first 28 days of life) mortality rate (NMR) and secondary education prevalence, two indicators relevant for SDGs 3 and 4. SDG 3 targets health indicators, with target 3.2 calling for an end to preventable deaths to neonatals and under 5 year olds (United Nations, 2020). Specifically, SDG 3.2 calls for NMR to be reduced to less than 12 neonatals per 1,000 live births, and for under-5 mortality rates (U5MR) to be reduced to less than 25 deaths per live births. SDG 4 targets education, calling for equitable and quality education for all. In particular, primary and secondary education completion for all boys and girls is emphasized in SDG target 4.1 (United Nations, 2020).

Considerable spatial variation exists in under 5 mortality and secondary education prevalence, and education disparities between men and women persist throughout the world (Graetz et al., 2018; Golding et al., 2017; Giorgi et al., 2018; Gething et al., 2016). This has

found to be the case in Kenya, for instance, where differences between estimates in years of education, and uncertainties in neonatal mortalities between subnational areas, have been identified (Graetz et al., 2018; Golding et al., 2017).

In low and middle income countries (LMIC), vital registration (VR) data and official government statistics are often of inadequate quality for estimating quantities relevant for the SDGs, necessitating the use of sparse, multilevel household surveys for unbiased NMR estimates (Li et al., 2019; Wagner et al., 2018; Sandefur and Glassman, 2015; Jerven, 2013; Devarajan, 2013). We introduce one such survey used throughout this thesis in Section 1.3.2. Official government statistics suggest growth in primary education enrollment to be approximately a third higher on average from 1991 to 2011 in 21 African countries than as suggested by Demographic Health Surveys (DHS) (USAID, 2019) data, with Kenya and Rwanda having particularly large discrepancies (Sandefur and Glassman, 2015). Examples of such surveys include, aside from DHS: Multiple Indicator Cluster Surveys (MICS) (UNICEF - Statistics and Monitoring, 2012), AIDS Indicator Surveys (AIS) (DHS Program, 2019), and Living Standard Measurement Surveys (LSMS) (The World Bank, 2019). Sandefur and Glassman (2015) suggest that official demographic and health data may be biased, systematically exaggerating progress towards development goals, while independent household surveys such as DHS surveys are likely more reliable.

Although multistage household surveys such as DHS are typically trusted more than VR systems, they typically involve complex stratified cluster sampling designs that may oversample urban or rural areas as described in detail in the 2012 DHS Sampling and Household Listing Manual (ICF International, 2012, Sec. 5.2, p. 80–85). Emphasis is usually placed on producing estimates for administrative areas. For instance, Admin0 is the entire country, Admin1 (counties, in the case of Kenya) is one level smaller, and Admin2 (constituencies in the case of Kenya) is a level smaller than Admin1. Typically, DHS surveys are stratified two-stage cluster sampling schemes, where strata are Admin1 areas crossed with official ur-

ban/rural designations. In the first stage of the sampling design, enumeration areas (EAs) are selected using probability proportional to size (PPS) sampling, where the "size" used to determine the probability of any given EA is determined by the number of listed households. The number of households in each EA used to determine the probability of selecting each EA and is based on the most recent census frame. Censuses are at best carried out once every ten years. Hence, there is a lag time between the census frame being created and when it is used for a survey. A census frame will typically have information about the locations of the EAs, and the approximate number of households and people in them along with EA official urban/rural designations. The number of clusters sampled in the DHS is most commonly set so the survey is powered to the Admin1 level. In the second stage of the sampling, usually 25 households are selected at random from within each EA, and information on everyone within the household is collected. Mothers are asked about full birth histories, with a number of questions regarding their children's birth dates and dates of death whenever applicable. Individuals are also asked about their education status.

Due to the complex sampling design of DHS and other surveys used for estimating health and development indicators relevant for the SDGs, it is important to be able to account for the sampling design when estimating such indicators. A number of classical survey statistics techniques can be applied in these settings. In particular, the weighted, Horvitz-Thompson (HT) estimator (Horvitz and Thompson, 1952) is able to account for the sampling design by taking a weighted average of the observations, with weights proportional to the inverse of the survey inclusion probabilities–the probability the observations are included in the survey–or possibly modified by multiplying survey inclusion probabilities with estimated probabilities of response in order to reduce nonresponse bias. HT estimators will be discussed further in Chapter 2. These are known as 'direct' estimates in the small area estimation (SAE) literature, which means they only use the responses within an area when the estimate for that area is produced. A major limitation is that no information from responses outside of

that area can be utilized. Additionally, when there is no or little data within an area, as is often the case for some Admin2 areas (a subdivision of administrative areas within Admin1 areas) these estimates are either unavailable or unreliable.

SAE models have been proposed (see, e.g., Rao and Molina (2015)) that are able to both account for the sampling design as well as for spatial smoothing. Many of these build on the model proposed in Fay and Herriot (1979), such as Mercer et al. (2015), and smooth the areal estimates to reduce their variance. These approaches use discrete spatial models at the areal level of interest, smoothing survey based direct estimates that incorporate survey weights via a two-stage modeling approach (Marhuenda et al., 2013; Chen et al., 2014; Mercer et al., 2015; Congdon and Lloyd, 2010; You and Zhou, 2011; Porter et al., 2014; Vandendijck et al., 2016; Watjou et al., 2017; Li et al., 2019). Because of this two-stage approach, small sample sizes within areas cause difficulties. Additionally, the ecological fallacy, assuming covariate associations at the areal and individual levels are the same, make inference on covariate associations difficult to interpret.

Studies using continuously indexed spatial models to produce estimates for health and development indicators at many different spatial aggregation levels using household surveys are numerous (e.g. Wardrop et al., 2018; Gething et al., 2016; Golding et al., 2017; Utazi et al., 2018b; Gething et al., 2015; Osgood-Zimmerman et al., 2018; Graetz et al., 2018; Diggle and Giorgi, 2016; Giorgi et al., 2018; Diggle and Giorgi, 2019; Tatem, 2017). This includes studies conducted at massive scales with 200 countries included, such as publications from WorldPop and the Institute for Health and Metrics Evaluation (IHME). The models used by WorldPop and IHME routinely ignore the survey design effects such as stratification by urbanicity, and naively account for cluster level and population numerator and denominator variation at the EA level. It is therefore important to further explore the effects of modeling choices when studying health and demogrpahic indicators using sparse household survey data.

It is also important to note that many commonly used models that are fitted to multi-stage household survey count data without using the above two stage approach estimate risk rather than prevalence. Here, by 'prevalence' we mean the empirical proportion of people in the population with a condition, whereas we will use 'risk' to mean the *average* empirical proportion, averaging over all possible populations drawn from a 'superpopulation,' a sampling distribution over possible populations. For instance, prevalence would include population numerator and denominator variation in the response, whereas risk averages out population numerator and denominator variation, and should instead be interpreted as a probability. This is significant because that additional variation in the response for models of prevalence may not completely average out for small enough areas, such as when producing fine scale estimates on a pixelated spatial grid. Figure 1.1 illustrates some of the differences between risk and prevalence, namely that risk is associated with the superpopulation, whereas prevalences are associated with a single population. Although either risk or prevalence can be the target of inference, prevalences are more concrete quantities for policymakers to consider. Prevalence models can be more directly compared to prevalence data since their levels of variation is more similar, and they can also be compared more directly to survey estimators, since survey estimators are not estimators of risk, but rather of prevalence. Since all the proposed continuous spatial models for population aggregates when EA locations are unknown are risk models, they significantly underestimate uncertainties for the prevalences within those pixels, potentially leading to misinterpretations about the precision and uncertainty in their predictions, incorrect conclusions drawn about the population of interest, and suboptimal interventions unless one is clear that risk is being estimated. In this dissertation we hope to shed light on this potential problem as well as provide potential solutions.

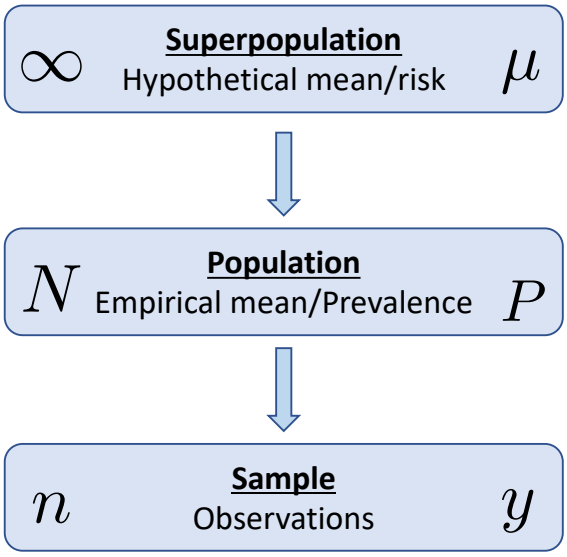


Figure 1.1: An illustration of the differences between superpopulations (risk), populations (prevalence), and samples (observations). The arrows between each of these concepts represents how sampling occurs: populations are sampled from superpopulations (hypothetically), and samples are sampled from populations.

1.2 Methodological Contribution of Dissertation

In this dissertation, we seek to assess how SAE models can better leverage information regarding the sampling design and spatial correlations present in the data to improve predictions of population aggregates. We mainly focus on accounting for stratification and cluster level variation for both areal and continuous spatial models, flexible covariance structures, aggregating spatial cluster level predictions, and EA location uncertainty. In Chapter 3, we conduct a thorough simulation study evaluating the performance of variants of several commonly used models when predicting neonatal mortality, identifying a method for accounting for stratification and cluster level variation in cluster level model based approaches, comparing them to design based approaches. In Chapter 4, we propose a framework for producing

areal predictions from continuous spatial models that can account for both stratification and cluster level variation. Additionally, this framework provides a method of naturally incorporating population numerator and denominator uncertainty for modeling prevalence in finite populations. Lastly, in Chapter 5, we propose a Bayesian extension to the LatticeKrig model (Nychka et al., 2015) that can both account for the non-Gaussian count observations of DHS data, as well as incorporating multiple scales of spatial correlation, making predictions more robust to spatial confounders and stratum misclassification.

1.3 Datasets

Throughout this dissertation, we will use both simulation studies and applications from Kenya, using health and demographic indicators, population density and census frame information in order to best produce estimates of population aggregates. All data used in this thesis come from the datasets described in this section.

1.3.1 2009 Kenya Population and Housing Census

The 2009 Kenya Population and Housing Census (Kenya National Bureau Of Statistics, 2014) provides information about the total number of people, households, and EAs per stratum. The census was taken over 96,251 total EAs, which constituted the sampling frame in the 2014 KDHS, although information about these EAs is not publicly available. However, a 10% sample from the census is available with geographical information providing population and demographic information. The EAs are a set of official areas constructed in the 2009 Kenya Population and Housing Census (Kenya National Bureau Of Statistics, 2014), dividing the population into similar sized units (in terms of population).

1.3.2 2014 Kenya Demographic Health Survey: Neonatal Mortality and Women's Secondary Education

The 2014 Kenya DHS (KDHS, 2014) is a DHS survey with design consistent with the standard described in the 2012 DHS Sampling and Household Listing Manual (ICF International, 2012, Sec. 5.2, p. 80–85). As mentioned in Section 1.1, the 2014 KDHS is a two-stage stratified cluster design. Strata consist of counties, of which there are 47, crossed with urban/rural designations, but 2 of the 47 counties only contain urban strata, those counties being Nairobi and Mombasa, making 92 strata in total. Of the 96,251 total EAs, 1,612 were selected from the strata using PPS sampling. From each of the selected EAs, 25 households were randomly sampled to include in the survey, forming the 1,612 clusters. Of the clusters, 995 are urban and 617 are rural, with most counties that contain urban and rural areas oversampling urban EAs. The survey design was chosen so that county (Admin1) level estimates could be sufficiently precise in estimating important outcomes.

1.3.3 2010–2015 WorldPop Population Density

Throughout this dissertation we utilize an estimate of population density based on 1km by 1km WorldPop spatially gridded population density estimates in 2010 and 2015 (Stevens et al., 2015; Tatem, 2017). We then use a constant rate of growth exponential interpolation of the 2010 and 2015 population density surfaces. In order to produce urbanicity classifications of spatial locations, we threshold the population density surface in each county so that the correct proportion of the population in each county is urban based on the 2009 Kenya Population and Housing Census, which designates the population of each county as either urban or rural. Figure 1.2 shows the population density surface and urbanicity classifications, although only the smoothed 5km by 5km estimates are used throughout this dissertation, unless otherwise stated.

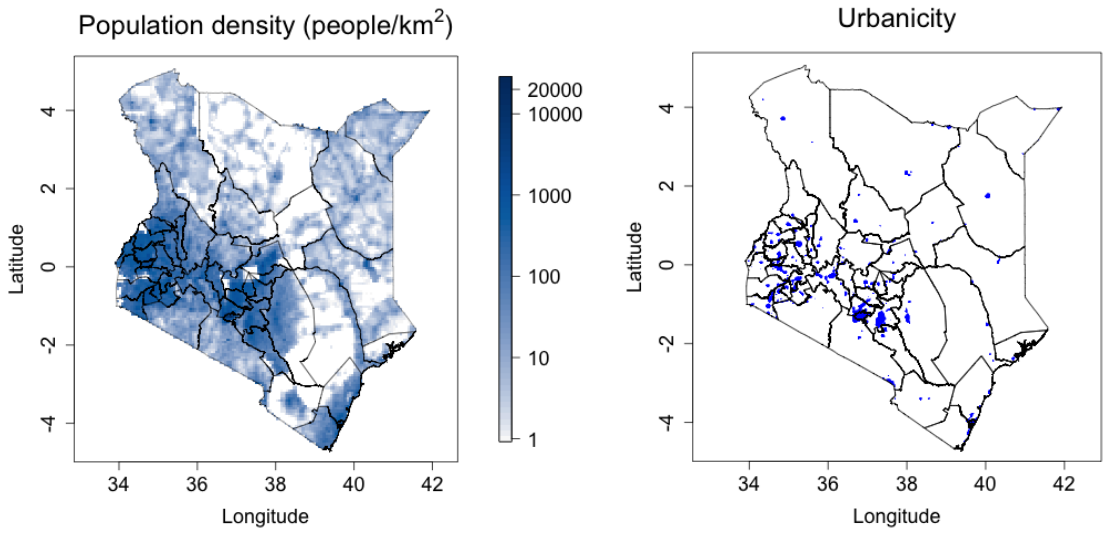


Figure 1.2: Left: population density interpolated from WorldPop estimates. Right: urban area classifications in Kenya used in the analysis are given in blue. Urban classifications are based on thresholding population density.

1.4 Organization of Dissertation

The rest of this dissertation is organized as follows. Chapter 2 provides methodological background that will be referenced throughout the dissertation. Chapter 3 presents a simulation study assessing a number of different direct and SAE models as well as variations with stratification fixed effects and cluster level random effects, and describes a new model for SAE. The models are used to estimate neonatal mortality in Kenya from 2010–2014, and the prevalence of secondary education completion in 2014 for women aged 20–29. In Chapter 4, we propose a framework for accounting for cluster level variation, uncertainty in EA locations, as well as population numerator and denominator variation for spatial models indexed at the cluster level and when aggregating to produce predictions at different spatial scales. We also explore how the uncertainty of predictions under this framework are affected

by the multiple different sources of uncertainty via simulation studies and validation. We analyze neonatal mortality data in Kenya to demonstrate the implications on real data. In Chapter 5, we propose a new model for accounting for flexible spatial covariance functions for non-Gaussian georeferenced data in a Bayesian framework, assessing performance relative to other commonly used models via simulation studies and validation. We also apply the model to the secondary education completion data. Lastly, in Chapter 6 we detail conclusions and future work.

Chapter 2

BACKGROUND

2.1 Notation

In this chapter, we will assume for concreteness that the goal is to estimate NMRs, although the notation and models can be used for many other applications. For EA level variables, we will use capital letters to denote population quantities, and lower letters to denote sampled quantities. Also, although the term 'area' could represent any areal unit, we will take it to be equivalent to county or Admin1 units unless otherwise stated. We use μ_{ic} to denote the risk of death in area i and EA c. The spatial location of EA c in area i is given by $\boldsymbol{s}_{ic}$. Unless otherwise stated we will also assume that the response at area i and EA c, y_{ic}, is binomial conditional on the risk of neonatal mortality, and the number of sampled neonatals in the cluster. However, the total number of neonatal deaths in area i and EA c will be denoted by Y_{ic}. We let n_{ic} and N_{ic} denote the number of neonatals in area i and EA c that are sampled and in the entire EA respectively.

When relevant, we will consider a fine 5 km $\times$ 5 km spatial grid indexed by the letter g. For instance, $\boldsymbol{s}_{ig}$ is the spatial location of grid cell g in area i. A population density surface is obtained via the WorldPop data as discussed in Section 1.3.3. We denote the general and target population (e.g. live births) density surfaces as $\tilde{q}(\cdot)$ and $q(\cdot)$ respectively, obtaining the target population density surface by scaling the overall population density in each stratum using census information so that $q(\boldsymbol{s})/\tilde{q}(\boldsymbol{s})$ matches the expected number of target population members per person in the stratum associated with location $\boldsymbol{s}$. Although we represent the population and target population density surfaces as continuous functions of space, numerically they are only evaluated at the centroids of each of the grid cells.

We use A_i, A_i^{URB}, and A_i^{RUR} to represent the geographical extent of area i and its partition into urban and rural strata. We will also use Q_i^{URB} and Q_i^{RUR} to denote the geographical extent of area i and the proportion of the target population in area i that is urban and rural respectively.

To make notation consistent, we will denote population prevalences using the letter p and risks using the greek letter μ. Following this rule, the goal is to estimate the following population prevalence:

$$p_i = \sum_{c \in B_i} \frac{N_{ic}}{N_i} \times \frac{Y_{ic}}{N_{ic}}, \tag{2.1}$$

where B_i is the set of EAs in area i and $N_i = \sum_{i \in B_i} N_{ic}$. We will use $C_i \subset B_i$ to represent the subset of EA indices in area i that are included in the sample. Hence, C_i represents the set of sampled clusters in area i.

2.2 Design-Based Approaches

2.2.1 Horvitz-Thompson Estimators

Classical Horvitz-Thompson estimators (Horvitz and Thompson, 1952) are obtained by weighting individual responses by the inverse of their probability of inclusion in the survey. We will assume neonatals within a sampled cluster all have equal probability of inclusion as will be the case, for instance, for household cluster surveys where either all households in each selected EA are included in the associated cluster or households within EAs are sampled via SRS. In this case, the cluster design weight, say w_{ic} for cluster c in area i, can be calculated as the sum of design weights of the neonatals in the cluster, and the design weight for the respondents in a cluster is determined by the inverse probability it was included in the sample. The HT estimator is then,

$$\hat{p}_i^{\text{HT}} = \frac{1}{\sum_{c \in C_i} w_{ic}} \sum_{c \in C_i} w_{ic} y_{ic}, \tag{2.2}$$

Although this estimator is unbiased for the prevalence in area i, and is reliable for large sample sizes, it will have large variance for small samples (Rao and Molina, 2015). See Appendix A.3 for information on calculating the weights for the 2014 KDHS. Estimates can be easily obtained via the `svyglm` package (Lumley, 2004, 2018) and also the `SUMMER` package (Martin et al., 2018) in `R`.

2.2.2 Smoothed Direct Estimators

There are several smoothed direct estimate approaches, all based on the seminal Fay-Herriot model (Fay and Herriot, 1979). Rao and Molina (2015) provides a review of many of these approaches, and we will focus on Mercer et al. (2015) in this dissertation.

First, direct design-based such as HT estimates are obtained for the prevalence in each area of interest. These estimates along are then transformed to the logit scale as, $Z_i = \text{logit}(\widehat{p}_i{}^{\text{HT}})$. Their variances are transformed using the delta method to obtain variances $\widehat{V}_i$ for each area i. A latent spatial signal may be modeled as:

$$\eta_i = \beta_0 + \frac{1}{\sqrt{\tau}}(\sqrt{\phi}S_i + \sqrt{1-\phi}\delta_i),$$

for intercept β_0, county level signal precision τ, proportion of spatial signal variance ϕ, and with random spatial and unstructured random effects S_i and δ_i for each area. The 'spatial' component S_i for each $i = 1, \ldots, 47$ is modeled using a mean zero intrinsic conditional autoregressive (ICAR) model (Besag et al., 1991), while the 'nonspatial' component δ_i for each area i are modeled as independent and identically distributed (iid) mean zero Gaussian random effects. Although the S_i and δ_i terms are usually considered spatial and nonspatial respectively, since they are indexed at the county level, this distinction is somewhat arbitrary as the δ_i may also be considered to be spatial. We ensure identifiability of the intercept term by applying a sum-to-zero constraint on the spatial component of the model. Under this constraint, and since the variance of $S_1, \ldots, S_{47}$ is scaled to have unit variance under the

constraint, ϕ can be interpreted as the proportion of county level variance, $1/\tau$, attributable to the spatial component even though it is intrinsic of order one under no constraints.

Like HT estimators, predictions from this model are *design consistent*, which means that under arbitrary sampling designs, as the proportion of the total population sampled approaches 1, the estimates will approach the true population prevalence.

Since both ICAR and nonspatial components of the model are included via the S_i and δ_i terms, this is considered a Besag-York-Mollié (BYM) model. The following section describes the BYM model in more detail along with the particular reparameterization used above that is sometimes known as the BYM2 model (Riebler et al., 2016).

2.3 Model-Based Approaches

We assume that the response is conditionally binomial: $y_{ic}|\mu_{ic} \sim \text{Bin}(n_{ic}, \mu_{ic})$. We will also write $\mu_{ic} = \mu(\boldsymbol{s}_{ic})$ for risk $\mu(\cdot)$ varying as a function of space. The logit risk is then modeled as:

$$\log\left(\frac{\mu(\boldsymbol{s}_{ic})}{1-\mu(\boldsymbol{s}_{ic})}\right) = \mathbf{X}_{ic}\boldsymbol{\beta} + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}, \tag{2.3}$$

where $\mathbf{X}_{ic}\boldsymbol{\beta}$ are the fixed effects for cluster c in area i and $u(\cdot)$ represents the spatial random effects. Cluster level random effects are represented as $\epsilon_{ic} \overset{iid}{\sim} \text{N}(0, \sigma_\epsilon^2)$.

2.3.1 Besag-York-Mollié Model and Reparameterization

The BYM2 model, introduced in Riebler et al. (2016), is a convenient reparameterization of the BYM model allowing for easily interpretable priors to be set on the variance parameters, the spatial random effect $u(\cdot)$ in (5.1) is represented as in the second stage of the smoothed direct model given above:

$$u(\boldsymbol{s}_{ic}) = \frac{1}{\sqrt{\tau}}(\sqrt{\phi}S_i + \sqrt{1-\phi}\delta_i).$$

Note that although the random effects used in the prediction, S_i and δ_i, are only indexed at the county level, the spatial effect $u(\cdot)$ is allowed to vary continuously throughout space. We will also consider variants of this model such as including iid Gaussian cluster level random effects into $u(\cdot)$, allowing predictions to be made separately for each cluster.

2.3.2 Gaussian Processes and the SPDE Approach

For continuous spatial models, where predictions are made continuously throughout space, Gaussian processes are a popular approach. We say that $\{u(\boldsymbol{s}) : \boldsymbol{s} \in \mathbb{R}^2\}$ is a mean zero, stationary Gaussian process if, for any set of locations $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_n$, we have $(u(\boldsymbol{s}_1),\ \ldots,\ u(\boldsymbol{s}_n))^T \sim \text{MVN}(\mathbf{0}, \sigma^2\mathbf{C})$ for spatial variance σ^2 and correlation matrix $\mathbf{C}$ with elements $\mathbf{C}_{ij} = \rho(\boldsymbol{s}_i, \boldsymbol{s}_j)$. Here, $\rho(\boldsymbol{v}, \boldsymbol{w})$ is a correlation function that depends on the input spatial locations $\boldsymbol{v}$ and $\boldsymbol{w}$. Typically, this correlation function is isotropic, in which case $\rho(\boldsymbol{v}, \boldsymbol{w}) = \rho(||\boldsymbol{v} - \boldsymbol{w})||)$. For a Gaussian process to have a valid distribution, the correlation function must be positive semidefinite.

One of the most common choices for ρ is the Matérn correlation function due to its flexibility, containing the exponential correlation function and also containing the Gaussian correlation function as a limit depending on the smoothness parameter, ν. In practice, however, this parameter is often fixed. While this is due in part to the computational benefit of having fewer parameters to estimate, the smoothness parameter is also often fixed because of the difficulty of identifying both the smoothness and ϕ, the spatial scale parameter, under infill asymptotics, the asymptotic regime where observations fill the spatial domain, but the spatial domain does not expand (Stein, 1999).

Classically, conditional on the set of spatial covariance parameters, σ^2, ϕ, and ν, *Kriging* is the method of using the optimal weights on the observations for making predictions at any given point when the data are continuous. The standard conditional normal formula is typically used when calculating these weights for Gaussian responses. For an introduction

to Kriging, see Gelfand et al. (2010, Chapter 3).

In practice, all relevant covariance parameters are fit using maximum likelihood or Bayesian estimation. Each evaluation of the (log) likelihood in a classical spatial model requires $\mathcal{O}(n^3)$ floating-point operations due to the necessity of computing a quadratic form involving the GP's precision matrix and $\mathcal{O}(n^2)$ memory for needing to store the GP precision matrix. As a result of this, a number of spatial models have been developed in order to reduce the computational burden of inference an prediction.

The stochastic partial differential equation (SPDE) approach developed in Lindgren et al. (2011) is a very popular method for overcoming these computational difficulties in classical spatial models. Gaussian processes with Matérn covariance and certain smoothness parameters can be represented using a SPDE. Lindgren et al. (2011) showed that the SPDE can be approximated numerically as a linear combination of a set of basis functions whose basis coefficients are given a specific covariance based on the Matérn covariance parameters. This process requires the smoothness to be $\nu = 1$ or $\nu = 2$, although smoothnesses in between those values can be approximated. Lindgren et al. (2011) chose a mesh of polygonal (triangular) basis functions. Since those basis functions have compact support, the precision matrix of the resulting GP is highly sparse, resulting in reduced memory costs. Additionally, only the inversion of a matrix of size $m \times m$ for m basis functions is required, which requires only $\mathcal{O}(m^{3/2})$ floating point operations in $\mathbb{R}^2$ space, since the basis function coefficients follow a discrete Gaussian Random Field (GRF), allowing for substantial computational speed improvements. Since the SPDE approach is implemented in Integrated Nested Laplace Approximation (INLA) (Rue et al., 2009b), it can also be applied to non-Gaussian responses such as exponential family data.

In order to make areal predictions using a continuous spatial model such as the SPDE model, the predictions are typically continuously integrated within the spatial domain of the area of interest. For classical spatial statistical models, this is known as *block kriging*, a good

introduction for which can be found in Gelfand et al. (2010, Chapter 29.3). Since there is no closed form in general for the areal integration, in practice the integral is calculated numerically over a spatial grid. While accounting for the target population density is sometimes ignored in these contexts, such as in Giorgi et al. (2018), we believe it is important to account for, since we are really interested in the population average rather than the spatial average. IThe predicted risk in an area takes the following form for area i and target population density surface $q(\cdot)$:

$$\mu_i = \int_{A_i} \mu(\boldsymbol{s}) \times q(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} \approx \sum_{j=1}^{G_i} \mu_{ij} \times q(\boldsymbol{s}_{ij}), \tag{2.4}$$

where G_i is the number of spatial pixels with centroids in the area used for approximating the integral, $\boldsymbol{s}_{ij}$ is the jth grid cell centroid location in area i, and A_i is the spatial extent of area i. In the above formulation, we assume that q is normalized to have unit integral (both analytical and numerical) within the area, so that $\int_{A_i} q(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s} = 1$, or, since in practice only the numerical integral is calculated, $\sum_{j=1}^{G_i} q(\boldsymbol{s}_{ij}) = 1$. Even more importantly, we are assuming that the risk, μ, can be represented as a smooth function of space.

Although (2.4) is typical for aggregating continuously indexed risk to the areal level, we will see in Section 4 that it can be problematic in the case where risk is not representable as a smooth spatial surface, μ. For instance, an iid Gaussian variate is sometimes assumed to be associated with each spatial location. This is known as a *nugget* in the spatial statistics literature, although we will consider it a cluster level random effect, or cluster effect, in the context of stratified cluster sampling where each cluster has a different spatial location. There are many possible interpretations as to what this cluster effect might be. First, it could be measurement error. In that case, we should set this random effect to zero when making predictions. Second, it could be overdispersion in the responses within a given EA. there are many different mechanisms by which overdispersion can occur, including the sampling of subgroups within the cluster to give dependence in the responses (McCullagh and Nelder, 1989). In that case, the cluster could be integrated out, or averaged over, when making

predictions, since we would like to know what the risk in a cluster is on average. Third, it might be a 'true' effect associated with that EA. Perhaps a health facility local to that EA is particularly well or poorly funded, or perhaps that EA has another highly localized trait associated with the outcome variable of interest. In this case, the cluster effects should be included when making predictions. Unfortunately, including these cluster effects is highly nontrivial, since the locations of the EAs in each area are not known. The difficulties of interpreting and accounting for the cluster effect will be explored in more detail in the coming chapters, particularly in Chapter 4.

2.3.3 LatticeKrig

Nychka et al. (2015) first proposed LatticeKrig (LK) in order to account for modeling Gaussian responses exhibiting multiscale spatial correlation, namely correlation occurring on both short and long spatial scales. Urbanicity, among other spatial confounders, could induce potentially both short and long scale dependence due to the small spatial extent of urban areas and the large extent of rural ones. If we do not know the urban/rural status throughout the spatial domain we may see short and large scale dependence. Nychka et al. (2015) begins with the stochastic process $Y = \{y(\boldsymbol{s}) : \boldsymbol{s} \in \mathcal{D}\}$ given spatial locations $\boldsymbol{s}$ in spatial domain $\mathcal{D}$. It is assumed that $y(\boldsymbol{s}_i)|\eta_i, \sigma^2_{\text{N}} \sim \mathcal{N}(\eta_i, \sigma^2_{\text{N}})$, $i = 1, \ldots, n$ for n observations, linear predictor η_i, and nugget variance σ^2_{N}. The linear predictor is assumed to follow the linear model, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$, for an $n \times p$ matrix of covariates $\mathbf{X}$ with an associated vector of coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)^T$, and $\boldsymbol{u}(\cdot)$ a smooth function of space.

In LK, u is decomposed in a series of lattices of increasing spatial resolutions, and the basis functions of each layer become increasingly fine:

$$u(\boldsymbol{s}) = \sum_{l=1}^{L} g_l(\boldsymbol{s}) = \sum_{l=1}^{L} \sum_{j=1}^{m(l)} c_j^l \phi_{l,j}(\boldsymbol{s}), \quad \boldsymbol{s} \in \mathcal{D} \subset \mathbb{R}^2.$$

The number of lattice layers is given by L, which is fixed before beginning the analysis, and

is typically set between 2 and 4 layers. Each layer l is a smooth function of space given by $g_1, ..., g_L$ with $m(1), ..., m(L)$ basis functions respectively, with $g_l = \sum_{j=1}^{m(l)} c_j^l \phi_{l,j}(\boldsymbol{s})$ for basis coefficients $c_1^l, ..., c_{m(l)}^l$ and associated basis functions $\phi_1^l, ..., \phi_{m(l)}^l$.

Nychka et al. (2015) proposed using rescaled Wendland radial basis functions (Wendland, 1995), with basis functions in layer l taking the form, $\phi_{l,j}(d) = \phi\left(\frac{d}{2.5\delta_l}\right)$, with

$$\phi(d) \;=\; (1-d)^6(35d^2 + 18d + 3)/3,$$

for $0 \leq d \leq 1$, and 0 otherwise. Each basis function is centered at a grid point on a lattice, with lattices of knot points illustrated on an example spatial domain in Figure 2.1. The lattice cell width of layer l is given by δ_l, in the factor of 2.5 in the formula for each $\phi_{l,j}$ ensures that the radius of each basis function is $2.5\delta_l$ so that there is sufficient overlap between the basis functions and therefore sufficient smoothness of the layer g_l. More overlap leads to smoother surfaces with fewer visible artifacts in the predictive spatial means and standard errors, while smaller overlap improves computational performance, as will be discussed in detail below.

The basis function coefficients for each layer follow independent mean zero spatial autoregressive (SAR) models with $\mathbf{c}_l \sim \text{MVN}(\mathbf{0}, \alpha_l \sigma_S^2 \mathbf{B}_l^{-1} \mathbf{B}_l^{-T})$. Given σ_S^2, the smooth spatial variance, α_l is the proportion of smooth spatial variance attributable to layer l, with $\sum_l \alpha_l = 1$. Note that the variance of the spatial process at any given point is not necessarily exactly σ_S^2 instead being somewhat affected by the heights of the basis functions at that point. The matrix $\mathbf{B}$ is an autoregressive matrix for layer l, with all values zero except taking value $4 + \kappa_l^2$ along the diagonal, and -1 in the ij-th element if grid cells i and j are neighbors in lattice layer l.

Lindgren et al. (2011) describes in detail how each layer is a Gaussian process with covariance that is approximately Matérn with smoothness parameter $\nu = 1$, and with effective range approximately $\rho_l \equiv \sqrt{8}\delta_l/\kappa_l$. Hence, each layer is able to account for a different scale, ρ_l, of spatial correlation. In LK, both δ_l and ρ_l are halved for each successive layer.

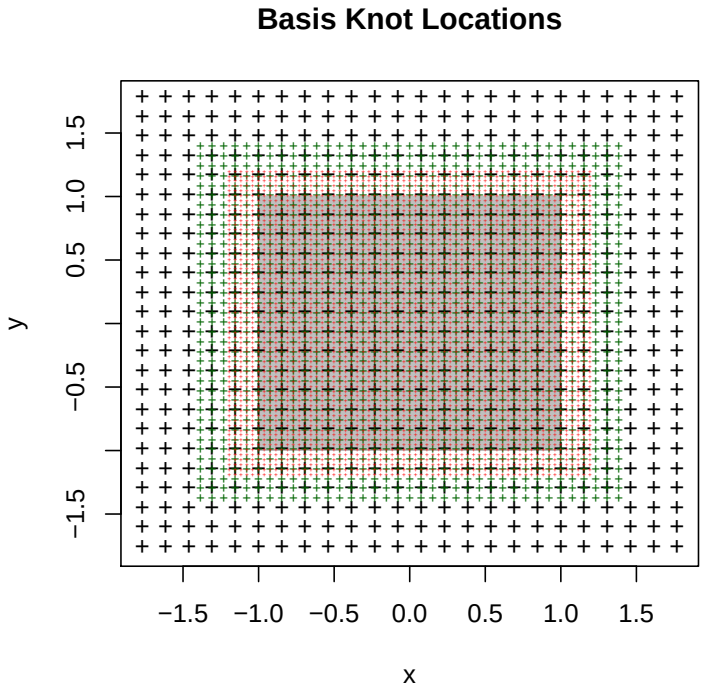


Figure 2.1: An example of a set of three lattice layers for the given $[-1, 1] \times [-1, 1]$ data domain represented by the shaded region. The '+' signs represent the knot points where the basis functions are centered: large black symbols representing the coarsest, first layer, medium green symbols representing the second layer, and small red symbols representing the last, finest layer.

Let $\mathbf{A}_l$ be the $n \times m(l)$ regression matrix from the basis coefficients for the l-th layer to the values of the basis function at the coordinates of the observations. Hence, $(\mathbf{A}_l)_{i,j} = c_j^l \phi_{l,j}(\boldsymbol{s}_i)$. The regression matrix from the basis coefficients to the values of the basis functions at the observation locations for all layers is then, $\mathbf{A} = (\mathbf{A}_1 \ ... \ \mathbf{A}_L)$ so that $\boldsymbol{u} = \mathbf{A}\boldsymbol{c}$, for $\boldsymbol{c} = (\boldsymbol{c}_1^T \ ... \ \boldsymbol{c}_L^T)^T$. The linear predictor can then be written as: $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{c}$.

In this formulation, p parameters are required for fixed effects, and $2L+1$ are required for the covariance. These $2L + 1$ parameters include 1 spatial variance parameter σ_S^2, an error (or nonspatial) variance parameter σ_N^2, $L - 1$ layer weight parameters, and L layer effective spatial range parameters. However, since in LK it is assumed that the effective ranges are halved for each successive layer, only one effective range parameter is required, yielding $L+2$ covariance parameters in total.

2.4 Comparison Measures

We compare model predictions using a number of measures of predictive performance including: the root mean square error (RMSE), the continuous rank probability score (CRPS) (Gneiting and Raftery, 2007), and the empirical coverage of 80% prediction intervals. These scoring rules allow us to evaluate model predictions with respect to data (in validation) and with respect to simulated populations (in simulation studies). Since each scoring rule evaluates a different aspect of model performance, we use several rules to not only rank model performance in terms of their central predictions and uncertainties, but to better understand why models may perform well or poorly.

We calculate the measures as an average of its values over each held-out observation. If observations are counts with denominator n_i for observation $i = 1, \ldots, n$, then we rescale the counts to be empirical proportions with $y_i = y_i^*/n_i$ for observed count y_i^* so that the scoring rules are calculated on a consistent scale.

Unlike RMSE, CRPS is a strictly proper scoring rule, and takes into account the accuracy

of the central predictions as well as the calibration of the uncertainty. Smaller values are preferable. Prediction intervals at the 80% level are derived from the 0.1 and 0.9 quantiles of the predictive distribution. They can be used to compute the prediction interval empirical coverage. For empirical proportions, and especially for small denominators, a fixed prediction interval will generally not provide the correct coverage even if the predictive distribution is correct due to the discreteness of the sample space (Geyer and Meeden, 2005). We therefore follow Geyer and Meeden (2005) by calculating *fuzzy* coverage instead, with details given in Appendix C.3. We have found that fuzzy coverage is much more precise than non-randomized coverage for discrete count data, allowing us to be sure that observed over- or under-coverages are due to the accuracy of the predictive uncertainty rather than the discreteness of the CIs.

Chapter 3

DESIGN- AND MODEL-BASED APPROACHES TO SMALL-AREA ESTIMATION IN A LOW AND MIDDLE INCOME COUNTRY CONTEXT: COMPARISONS AND RECOMMENDATIONS

3.1 Introduction

In light of SDG target 3.2 calling for reduced NMRs and target 4.1 calling for primary and secondary education completion for all, in this chapter we investigate NMRs and secondary education prevalence in Kenya. Since spatial disparities have been found in these quantities in Kenya, and since the 2014 KDHS has 1,612 sampled clusters (KDHS, 2014), the survey has enough observations in each county for a sufficient degree of precision to be achieved in county level estimates.

The 2014 KDHS, as a standard DHS survey, is a multistage cluster sample, and is stratified by county crossed with the official urban/rural designation. Since there are 47 counties, and each of them except for Nairobi and Mombasa have both urban and rural areas, with Nairobi and Mombasa having only urban areas, there are 92 strata in total. Previous studies have suggested that health an6d development indicators may vary by urbanicity designation, so not accounting for this stratification could lead to both bias and poor coverage in the predictions of relevant indicators (Van de Poel et al., 2007; Beatriz et al., 2018).

Despite the importance of accounting for survey effects, WorldPop and IHME, two organizations that produce estimates of demographic quantities on a large scale routinely ignore stratification as well as other design effects (see, e.g. Tatem, 2017; Osgood-Zimmerman et al., 2018). For instance, WorldPop frequently ignores cluster effects, and IHME includes exactly

one cluster effect per 5 km $\times$ 5 km pixel when generating pixel level predictions and when aggregating predictions to coarser areal units. Instead of accounting for stratification effects directly, they include a number of pixel level covariates in order to improve predictive performance. Some of these covariates may be correlated with official urbanicity classifications, thereby accounting for stratification to some extent purely by chance. However, these adjustments only indirectly account for stratification, and can therefore lead to biased predictions in the presence of significant associations between the response and urbanicity.

IHME and WorldPop fit the SPDE model as described in Section 2.3.2 to household survey data with and without cluster effects respectively. By assuming a single cluster effect per pixel, the IHME model implicitly assumes exactly a single EA per pixel. However, this is an assumption that is both false, and also problematic, since the added cluster level variation is disproportionately added to rural areas and areal uncertainties depend on pixel resolution. Although we attempt to mitigate these problems in this chapter, it is important to note that none of the proposed models will directly account for within stratum design effects due to variation in the design weights in each stratum as a result of nonresponse, PPS sampling, and EA target population size. For information on how these affect the design weights such as those used in HT estimators, see Section 2.2.1 and Appendix A.3.

In light of the gaps in these current modeling techniques, Section 3.2 proposes modifications of the BYM2 and SPDE models presented in Sections 2.3.1 and 2.3.2 that can better account for stratification and cluster level design effects. Section 3.3 introduces and presents a thorough simulation study, simulating several example populations of Kenya, with all EAs, households, and people simulated across the entire country along with hundreds of both representative and stratified surveys from each population. The predictions of each considered model will be compared using the scoring rules given in Section 2.4. Section 3.4 applies the models to secondary education prevalence for women aged 20–29 in 2014 in Kenya and also to NMRs in Kenya in 2010–2014. The considered models are also applied to two datasets,

and model validation scores examined. Lastly, Section 3.5 provides a discussion of the results as well as opportunities for future work. Note that this Chapter is heavily based on Paige et al. (2020).

3.2 Incorporating Design Effects into Spatial Models

In order to account for stratification in the model based approaches of Section 2.3, we begin by assuming $y_{ic}|\mu_{ic} \sim \text{Bin}(n_{ic}, \mu_{ic})$, and slightly modify the model for logit risk in (2.3):

$$\log\left(\frac{\mu(\boldsymbol{s}_{ic})}{1-\mu(\boldsymbol{s}_{ic})}\right) = \beta_0 + \beta^{\text{URB}} I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}, \tag{3.1}$$

where β^{URB} is an urban fixed effect, U is the set of all urban spatial locations, $I(\cdot)$ is an indicator function, and $\epsilon_{ic} \overset{iid}{\sim} \text{N}(0, \sigma_\epsilon^2)$ is a cluster level random effect. We then assume that $u(\boldsymbol{s}_{ic})$ either follows a BYM2 or SPDE model. Adding in the stratum level fixed effects has been shown to be sufficient for accounting for stratification in some settings (Lohr, 2010), and, for simplicity, we assume that the urban effect is the same for each county.

In this chapter, we will assume that the cluster level random effect can be interpreted as overdispersion in the responses within a given EA, from survey to survey. Under this interpretation, we would like to know what the risk in a cluster is on average, so we integrate out the cluster effect when making any predictions. Hence, the risk becomes:

$$\mu(\boldsymbol{s}_{ic}) \approx \int_{-\infty}^{\infty} \text{expit}(\beta_0 + \beta^{\text{URB}} I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic}) + \epsilon) f_\epsilon(\epsilon)\, \text{d}\epsilon \tag{3.2}$$

$$\approx \text{expit}\left(\frac{\beta_0 + \beta^{\text{URB}} I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic})}{\sqrt{1+k^2\sigma_\epsilon^2}}\right), \tag{3.3}$$

for $k = 16\sqrt{3}/(15\pi)$, using a logistic approximation to the Gaussian likelihood as given in Wakefield (2013), and where f_ϵ is the probability density of the cluster effect. Due to the long tail of the logistic relative to the Gaussian distribution, the logistic approximation used in (3.3) tends to bias μ slightly towards 0.5 due to the convexity of the logistic link being

positive for $\mu < 0.5$ and negative for $\mu > 0.5$. However, it is a commonly used approximation, and we have found that it works quite well.

Since we are ultimately interested in areal predictions rather than just predictions at the cluster level, we require a method of aggregation that accounts for stratification. Given an estimate of the proportion of the target population in county i that is urban and rural, Q_i^{URB} and Q_i^{RUR} respectively, aggregating predictions for the BYM2 model can be obtained using a weighted average. First note that, since the predictions in the urban and rural strata within area i do not vary within strata, we may write $\mu(\boldsymbol{s}_{ic})$ as either μ_i^{URB} or μ_i^{RUR} depending on whether $\boldsymbol{s}_{ic}$ is in an urban or rural area. The areal prediction is then just a weighted average of predictions for each stratum within the area:

$$\mu_i \quad = \quad Q_i^{\text{URB}}\mu_i^{\text{URB}} + Q_i^{\text{RUR}}\mu_i^{\text{RUR}}.$$

For the SPDE model, we will assume the target population density is normalized so that $Q_i^{\text{URB}} = \int_{A_i^{\text{URB}}} q(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}$ and $Q_i^{\text{RUR}} = \int_{A_i^{\text{RUR}}} q(\boldsymbol{s})\, \mathrm{d}\boldsymbol{s}$, with A_i^{URB} and A_i^{RUR} representing the urban and rural portions of A_i (the areal extent of county i) respectively. If q is normalized in this way, stratified prediction for county i does not change at all from in (2.4), expect that we will substitute in the logistic approximation from (3.3) when integrating pixel level risk.

We will consider 4 variations of both types of spatial models: BYM2 and SPDE. Models with and without urban effects are labeled 'U' and 'u', respectively. Similarly, models with and without cluster effects are labeled 'C' and 'c', respectively. For example, the BYM2 model with and urban effect but without cluster effects would be labeled 'Uc', or 'BYM2_{Uc}'. The 'uc', 'uC', 'Uc', and 'UC' variations of each model will be considered in this Chapter.

Following Fuglstad et al. (2019a), we set a joint PC prior on the continuous spatial standard deviation and effective range parameters σ_S and ρ, respectively. We use the joint PC prior described by Riebler et al. (2016) on the BYM2 standard deviation $1/\sqrt{\tau}$ and the proportion of variation that is spatial, ϕ. We also set a marginal PC prior on the cluster effect standard deviation, σ_ϵ. The priors in the simulation study and in the application are

set so that the median of the prior for ρ is one fifth the diameter of the spatial domain, and so that $P(\sigma_S > 1) = P(1/\sqrt{\tau} > 1) = P(\sigma_\epsilon > 1) = 0.01$. This results in the continuous spatial effects, BYM2 effects, and cluster effects for the spatial smoothing models each having a roughly 95% prior chance of lying between 0.5 and 2 on an odds scale. The PC prior for the spatial proportion in the BYM2 model, ϕ, is given a 2/3 prior probability of being less than 1/2, implying that we slightly favor the iid county level effects when apportioning residual variation. We choose this prior on ϕ in order to promote less complex models with a smaller spatial contribution.

3.3 Simulation Study

In this section, we conducted a simulation study assessing the predictive performance of a number of different SAE and survey-based techniques, including both design-based methods such as HT and smoothed direct estimators, and model-based methods such as several BYM2 and SPDE model variations. We also include a 'naive' logistic regression intercept model fitted at the county level. County level predictions of the methods are compared against true, simulated population averages at the county level.

3.3.1 Simulation Setup

To generate a true underlying population, we begin by spatially partitioning Kenya into urban and rural areas by thresholding population density. We choose the threshold of population density that determines the urbanicity designation so that the correct proportion of population within each county is urban using the population density surface from WorldPop and the 2009 Kenya Population and Housing Census. We then simulate every 96,251 EA location, ensuring that there is the correct number per stratum (these numbers are given by the 2009 census). Within each stratum, EA locations are drawn from pixels with probability proportional to the population density. All relevant information from the census is available

in the 2014 KDHS final report (KDHS, 2014).

After calculating empirical distributions from the census for the number of households per EA, the number of mothers per household, and the number of children born per mother, calculating separate empirical distributions for urban and rural areas, we use those empirical distributions to simulate the number of households and children in each EA.

We consider 3 spatial parameter scenarios, 4 populations per scenario, 2 survey designs per population, and 100 surveys per design (or 250 surveys for the HT estimators). The 4 populations per scenario will be referred to with the following names: constant risk (Pop_{suc}), spatially-varying risk (Pop_{Suc}), spatially-varying risk with an urban association (Pop_{SUc}), and spatially-varying risk with an urban association and a cluster effect (Pop_{SUC}). In the subscripts, we use U/u and C/c to indicate the presence of urban and cluster effects respectively, and we use S/s to indicate presence of a continuous spatial effect. For a given population, NMRs are stimulated at each of the 96,251 spatial locations using a SPDE model with the effects given in the population. For the constant risk population, only an intercept is used to simulate the NMRs. For the first scenario and the Pop_{SUC} population, the SPDE model parameters are $\rho = 150$ (in km), $\beta_0 = -1.75$, $\sigma_s^2 = 0.15^2$, $\sigma_\epsilon^2 = 0.1^2$, and $\beta^{\text{URB}} = -1$. Median EAs in rural and urban areas, with random effects of zero, have prevalences 17% and 6% respectively. The spatial ranges decreased to 50km in the second scenario, ended the third scenario, the spatial ranges 150km, but the spatial variance is increased to 0.3^2. Since these changes do not affect the Pop_{suc} population, only one of these populations is simulated. This means there are 10 different populations that we generate across all scenarios.

We carry out both "Unstratified" and "Stratified" surveys for each population, always taking 1,612 clusters per survey to match the 2014 KDHS. In both designs, the number of clusters per county are fixed to be the same as in the 2014 KDHS. In the Unstratified design, the average proportion of urban and rural clusters within each county is set to match the proportion of the urban and rural population for that county. The exact number of urban

and rural clusters in a stratum varies from survey to survey by at most 1, varying only to ensure that the average proportion matches the proportion of the underlying population to stay representative. In total, Unstratified surveys have an average of 430 urban and 1,182 rural clusters per survey. In the Stratified surveys, the number of clusters per stratum are fixed to be the same as in the 2014 KDHS, resulting in 995 and 617 urban and rural clusters respectively. Conditional on the total number of urban and rural clusters for each of the 92 strata, we use PPS sampling to determine which clusters are included in the surveys, sampling clusters with probability proportional to the number of households in each strata. SRS is used to select 25 households from each selected EA to include in the cluster, all children within the given households being included in the sample. We simulate 250 surveys for each population and design, with naïve and direct HT estimates fit to all 250 surveys (since their predictions are variable), and the other models being fit to only 100 of the surveys for each population and design due to computational limitations.

It is important to note that, although the SPDE_{UC} model is similar to what is used when simulating Pop_{SUC}, it is not identical. This is because the EA locations, population denominators, and cluster effects are all simulated and fixed when generating the population, whereas the SPDE_{UC} model uses (3.3) as an estimate for the areal risk.

3.3.2 Simulation Study Results

In the following section, we discuss the first simulation scenario, where populations including the spatial effect were simulated with 150km spatial range and 0.15^2 spatial variance, unless otherwise stated. A more detailed analysis of the other scenarios is given in Appendix A.3. The scoring rules summarizing the main results for the stratified design are plotted in Figure 3.2, and parameter summary statistics are given in Appendix A.3. Scoring rules for additional model variations and designs are given in Section 2.4. When interpreting these scoring rules, it is important to keep in mind that SDG 3 calls for a reduction in NMRs

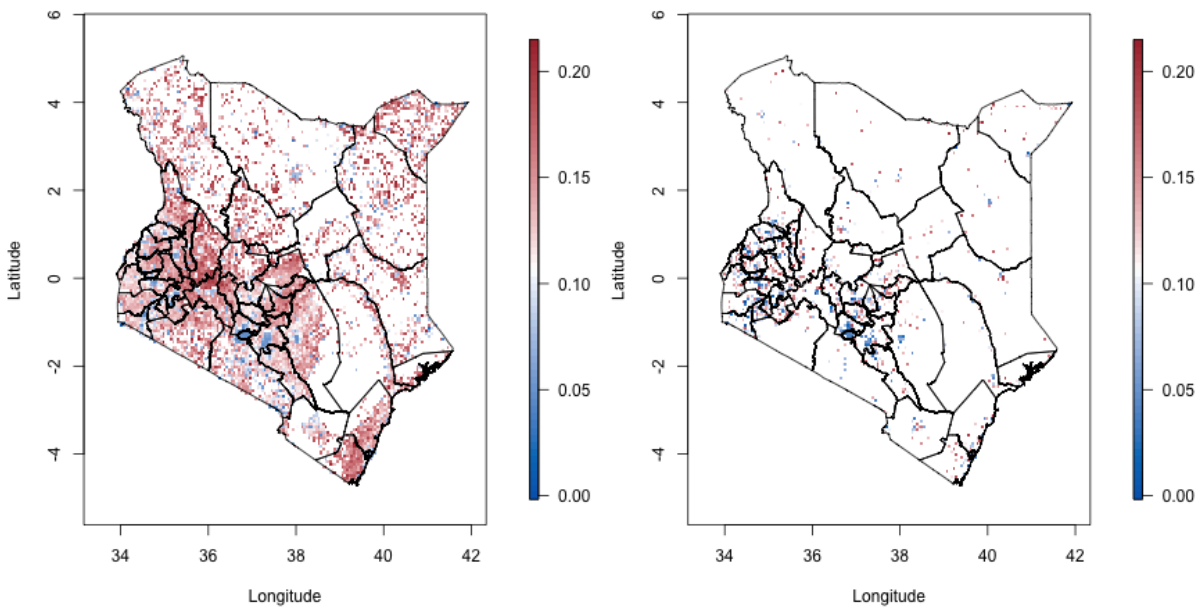


Figure 3.1: Simulated population of Kenya and associated NMRs at EAs (left), and at sampled clusters (right) for an example simulated dataset based on the "Stratified" design.

to 12 deaths per 1,000 children, which corresponds to 120 $\times 10^{-4}$ children. When absolute bias is large relative to this number, it is an indicator of poor model performance. Since we are especially interested in the performance of the models in a feasible scenario in which spatial, urban and cluster effects must be accounted for, we will be discussing Pop_{SUC} under a Stratified design unless we state otherwise.

Of the direct, smoothed direct, BYM2_{UC}, and SPDE_{UC} models, the BYM2_{UC} model performed the best or very close to the best in terms of CRPS, MSE, coverage, and CI width across all populations and scenarios. Although the BYM2_{UC} model was slightly positively biased, the precision of its central predictions and the well-calibrated predictive distribution and uncertainty intervals led to accurate coverages and good predictive performance.

Interestingly, although the SPDE_{UC} model was the most similar to the model used to

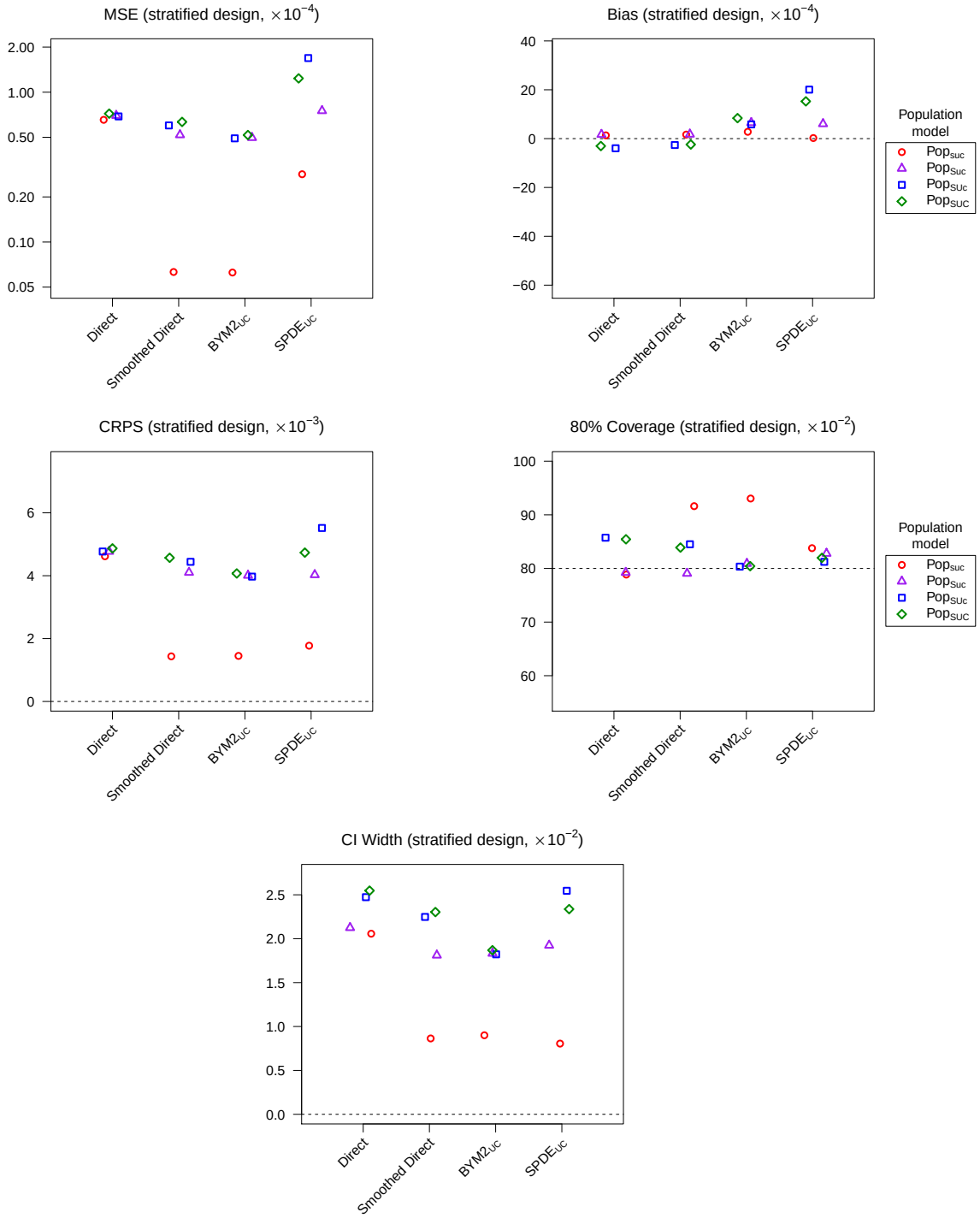


Figure 3.2: County level scoring rules plotted for each of the simulated populations and the main models considered for the Stratified design. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included in the models respectively. The "Population model" denotes the method by which the data were generated.

simulate Pop_{SUC}, it did not perform well in terms of MSE. Its MSE was 1.24×10^{-4} compared to 0.41×10^{-4} for the BYM2_{UC} model, 0.63×10^{-4} for the smoothed direct, and 0.72×10^{-4} for the estimates. Although SPDE_{UC} model estimates were somewhat positively biased, the high level of MSE was mainly due to lack of predictive precision. In spite of this, the SPDE_{UC} model had a CRPS of 4.7×10^{-3}, which was comparable to the value of 4.6×10^{-3} for the smoothed direct model, and was better than the direct estimates' value of 4.9×10^{-3}. Additionally, the coverage of the SPDE_{UC} model was 82%, second in accuracy only to the BYM2_{UC} model. Hence, although the predictions of the SPDE model were a little imprecise, the uncertainty of the posterior distribution was well-calibrated.

The direct, smoothed direct, and BYM2_{Uc} estimates had the smallest magnitude bias. An advantage of the smoothed direct model is that, regardless of the population and survey scheme considered, the model performed well from the standpoints of MSE, CRPS, bias, and coverage. Although the coverage for the Pop_{suc} (constant risk) population was over 90% for 80% nominal coverage intervals, that was also the case for the BYM2 models. Additionally, the constant risk setting is not realistic, and therefore should not be focused on too much. Overall, the smoothed direct model was robust in terms of its predictive accuracy and uncertainty.

Models that did not account for urbanicity either indirectly via sampling weights or directly as a covariate had relatively poor performance from MSE, bias, CRPS, and coverage standpoints. Even for populations without urban associations or under the unstratified design, there was little downside to including urban effects so long as the proportion of children in urban and rural areas was not poorly estimated (so that the area averaging was poorly performed). Including urban effects led to MSE, bias, CRPS, credible interval width, and coverage that was on average either better or nearly equal to the corresponding models without urban effects throughout all simulated populations and designs. The benefit of including urbanicity as a covariate was increased under the Stratified design relative to the

Unstratified design since urban and rural areas were not sampled proportionately in that case.

The scoring rules in Appendix A.3.2 shows the inclusion of a cluster effect led in general to better or equally good predictions for the BYM2 model in terms of MSE and CRPS. Although the MSE and bias of the BYM2_{uc} and BYM2_{uC} models were essentially the same, the inclusion of the cluster effect led to a dramatic improvement in coverage from 55% to 68%, indicating that cluster effects can lead to more accurate measures of uncertainty. Although the BYM2_{Uc} model arguably performs slightly better than the BYM2_{UC} model in terms of its MSE and CRPS, the coverage of the BYM2_{UC} model is better, and the uncertainty intervals are more conservative. The SPDE predictions were relatively more affected by the inclusion of the cluster effect, and its predictive performance was overall the most variable. Although there were some populations where the SPDE_{Uc} and SPDE_{UC} models had the best predictions in terms of MSE and CRPS, care must be taken when accounting for spatial and cluster level variation. To summarize, these simulations suggest that, amongst the BYM2 models, the BYM2_{UC} model is an effective and robust choice for the analysis of DHS household survey data, whereas more work must be done to ensure continuous spatial models share those same qualities.

Patterns in the model scores for the most part remained the same across the different simulation scenarios. The relative performances of the naive, direct, smoothed direct, and BYM2 models were mostly the same. The main differences were in the performance of the SPDE models. In spite of this, the SPDE models did not consistently perform better than the BYM2 models in any of the scenarios, again indicating the relative robustness of the BYM2 and smoothed direct models, and the caution necessary when making aggregated predictions with such flexible spatial models using sparse household surveys.

3.4 Mapping Health and Demographic Indicators in Kenya

In this Section, we use the 2014 Kenya DHS to estimate secondary education completion rates for women aged 20–29 in 2014, and to estimate NMRs for the five year period from 2010 to 2014. In the case of secondary education prevalence, we choose the 20–29 age group because most women that will complete their secondary education have already done so by that age, and also because the 2014 Kenya DHS indicates there are generational differences in secondary education levels. Although we find significant evidence of association with urbanicity in the case of secondary education completion, we did not find strong evidence of a marginal association between NMRs and urbanicity in Kenya. Since associations between urbanicity and the response lead to larger biases when stratification is not accounted for, we believe the secondary education completion dataset provides especially strong evidence for the importance of accounting for stratification in the design. Additional results for the women's secondary education and NMR applications are given in Appendix A.1 and A.2 respectively.

3.4.1 Prevalence of Women's Secondary Education

Risk Mapping

Central predictions as well as interval widths for the direct, naive, smoothed direct, and the full ('UC') spatial smoothing models are shown in Figure 3.3. The top row (point) estimates are quite similar, since there are a large number of clusters, but close examination shows there are differences. Prevalence tended to be higher in the central, southern, and western counties, and tended to be lower and with greater uncertainty in the more rural counties to the north and east. Appendix A.1 gives full numerical results and here we summarize. The odds (with associated 80% CIs) of young women in urban clusters completing their secondary education are larger, relative to rural clusters, by 210% (185%, 236%) or 170%

(148%, 193%) as respectively calculated from the BYM2_{UC} and SPDE_{UC} model parameter estimates given in Table 3.1.

Table A.1 in Appendix A.1 shows that the smoothed direct, SPDE_{UC}, and BYM2_{UC} models all estimate that the secondary education levels for young women in Kenya were highest in Nairobi, with point estimates (80% CIs) of 0.54 (0.49, 0.58), 0.55 (0.51, 0.58), and 0.53 (0.50, 0.57) respectively. On the other hand, Mandera was estimated to have the lowest secondary education levels for all models except for the SPDE_{UC} model (for which Turkana was estimated to have the lowest secondary education levels) with point estimates (80% CIs) of 0.088 (0.061, 0.13), 0.081 (0.058, 0.11), and 0.085 (0.060, 0.12) respectively. While Nairobi is designated as completely urban, approximately 18% of the population of Mandera is urban, which is very close to the median for counties in Kenya. This suggests there are other factors in Mandera that are reducing the secondary education prevalence for the women living there.

The credible interval widths were largest for the direct model and smallest for the SPDE_{UC} model. Of the displayed spatial smoothing models, the smoothed direct model had the largest predictive variances. Both the smoothed direct and the BYM2 models had relatively high uncertainties for counties with fewer neighbors, whereas the SPDE model variances tended to be high near the edges and where the clusters were spatially distant from each other.

Figure 3.4 shows the continuous, 5km$\times$5km pixel level predictions and credible interval widths for the SPDE_{uC} and SPDE_{UC} models. The urban effect is especially visible in the predictions of the SPDE_{uC}, since it oversmooths the effect of the urban areas into nearby rural regions. Interestingly, secondary education prevalences appear to be high not only in urban areas, but also in rural areas bordering urban centers. Figure 3.4 clearly shows that care that must be taken with stratification. In order to maintain confidentiality, the geographical locations of the DHS clusters are displaced (jittered): urban clusters by up to 2km, and rural clusters by up to 5km, with 1% of rural clusters jittered up to 10km. In general, we

are nervous about presenting and over-interpreting pixel level maps due to jittering, cluster level overdispersion, data sparsity, and confounding pixel level covariates such as urbanicity. Although continuous spatial models in theory allow for high resolution predictions, it is very important when modeling demographic indicators from sparse household survey data to account for these complicating factors.

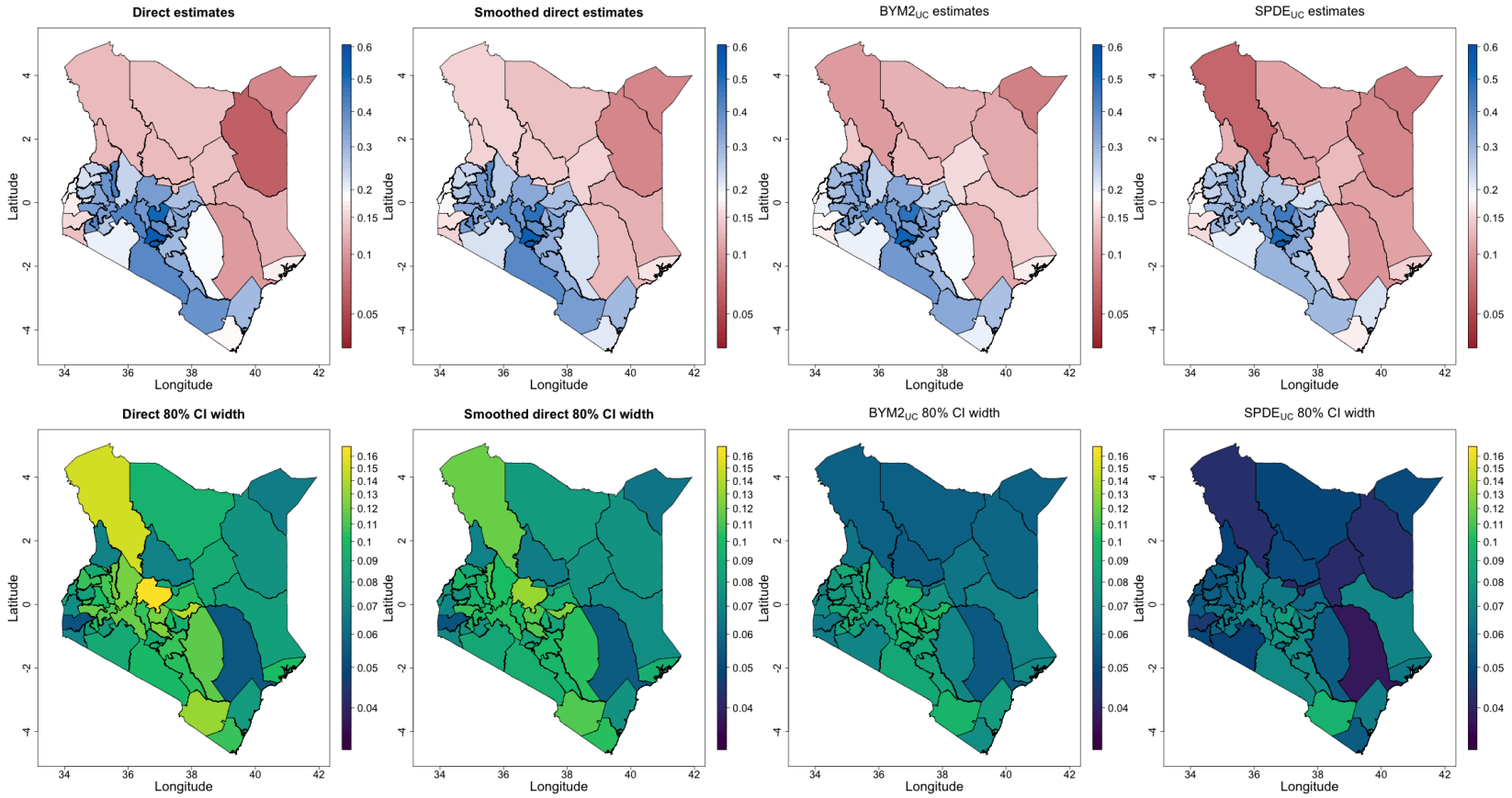


Figure 3.3: Kenya county level 2014 secondary education predictive estimates (top) and 80% uncertainty interval width (bottom) for women aged 20–29. The BYM2 and SPDE model predictions include both urban and cluster effects.

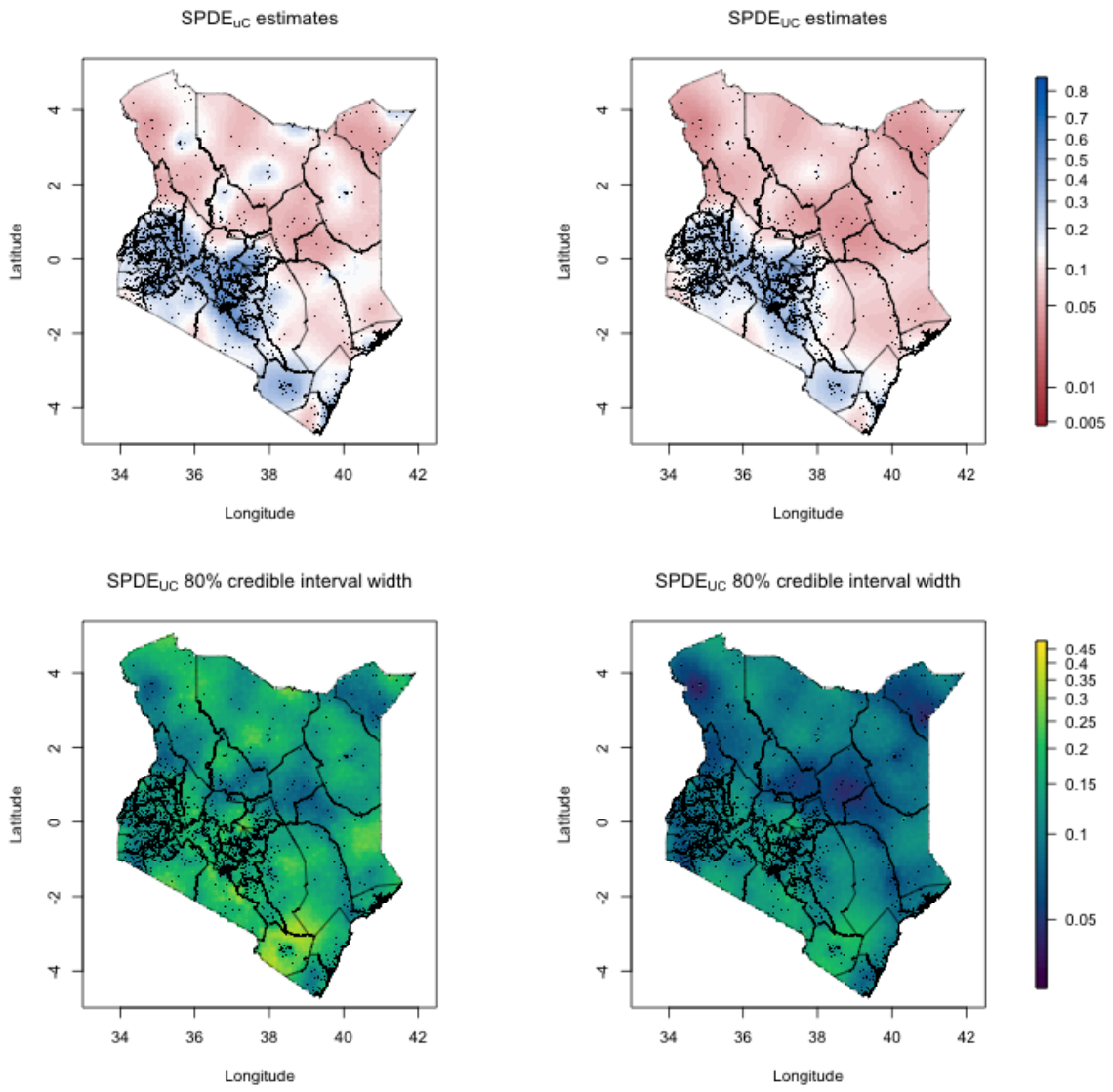


Figure 3.4: Kenya 5km × 5km pixel level 2014 secondary education predictive mean (top) and 80% uncertainty interval width (bottom) for women aged 20–29. Results are shown for the SPDE$_{uC}$ (left) and SPDE$_{UC}$ (right) models.

	Est	SD	Q10	Q50	Q90
Smoothed Direct					
Intercept	-1.04	0.05	-1.10	-1.04	-0.99
BYM2 Phi	0.80	0.16	0.57	0.84	0.97
BYM2 Tot. Var	0.32	0.08	0.22	0.31	0.42
BYM2 Spatial Var	0.26	0.09	0.15	0.25	0.37
BYM2 iid Var	0.06	0.05	0.01	0.05	0.13
BYM2 Tot. SD	0.56	0.07	0.47	0.55	0.65
BYM2 Spatial SD	0.50	0.09	0.38	0.50	0.61
BYM2 iid SD	0.22	0.10	0.10	0.22	0.36
$\boldsymbol{BYM2_{UC}}$					
Intercept	-1.64	0.06	-1.72	-1.65	-1.57
Urban	1.13	0.07	1.05	1.13	1.21
Cluster Var	0.51	0.05	0.44	0.51	0.58
BYM2 Phi	0.84	0.14	0.64	0.88	0.98
BYM2 Tot. Var	0.33	0.08	0.24	0.32	0.44
BYM2 Spatial Var	0.28	0.09	0.17	0.27	0.40
BYM2 iid Var	0.05	0.05	0.01	0.04	0.11
Cluster SD	0.72	0.04	0.67	0.71	0.76
BYM2 Tot. SD	0.57	0.07	0.49	0.57	0.66
BYM2 Spatial SD	0.52	0.09	0.42	0.52	0.63
BYM2 iid SD	0.21	0.10	0.09	0.20	0.34
$\boldsymbol{SPDE_{UC}}$					
Intercept	-2.53	0.21	-2.80	-2.53	-2.26
Urban	0.99	0.07	0.91	1.00	1.08
Range	212	44	161	206	271
Spatial Var	0.84	0.24	0.57	0.81	1.16
Spatial SD	0.91	0.13	0.76	0.90	1.08
Nugget Var	0.43	0.05	0.36	0.43	0.50
Nugget SD	0.65	0.04	0.60	0.65	0.70

Table 3.1: Parameter and hyperparameter estimate summary statistics from the BYM2 and SPDE models including both urban and cluster effects, fit to the 2014 secondary education prevalence data for young women from the 2014 Kenya DHS.

Validation

We calculate a number of scoring rules at the cluster level to evaluate the spatial smoothing models. We compute the scoring rules by leaving out data from one county at a time and averaging the scoring rules over all 47 such experiments. We carry out the validation at the county level, since this is generally the target of inference. In addition to calculating MSE (broken down into variance and bias and in urban as well as rural areas) we also compute CRPS, the deviance information criterion (DIC), and the conditional predictive ordinate (CPO). The naive, direct, and smoothed direct models are fit at the county level, so we did not include their validation results, since they are not comparable with the cluster level data models.

The SPDE_{Uc} and SPDE_{UC} models had the two best average MSEs, and the SPDE_{Uc} model has the best CPO and CRPS. The SPDE_{UC} had better MSE than the Uc model, but it had worse CPO and CRPS. In terms of MSE and CRPS, the BYM2_{Uc} model also performed well, and had the smallest magnitude bias. The good performance of the SPDE models may be due to their ability to model continuous changes in secondary education near the borders of each county, whereas the BYM2 models are unable to distinguish between clusters at the border of a county versus clusters in the center. Appendix A.1 shows that the spatial standard deviation (SD) of the SPDE_{UC} model is estimated to be 0.91, whereas the cluster effect is estimated to have a SD of 0.65. This is a higher proportion of variability going into the spatial term than in the BYM2_{UC} model, which estimates the total variance of county level random effects to be 0.57 and the cluster variance to be 0.71. This suggests the ability of the SPDE model to predict continuously through space gives it an advantage when making predictions at the cluster level.

	BYM2				***SPDE***			
	uc	uC	Uc	UC	uc	uC	Uc	UC
MSE ($\times 10^{-2}$)								
Average	5.9	*6.0*	5.1	5.2	5.6	5.4	5.0	**4.9**
Urban	*7.2*	*7.2*	6.2	6.2	7.1	6.8	6.2	**6.0**
Rural	5.0	*5.2*	4.5	4.5	4.6	4.5	**4.2**	**4.2**
Var ($\times 10^{-3}$)								
Average	59	*59*	51	52	56	53.8	50	**49**
Urban	61	62	62	62	*63*	**60**	62	60
Rural	45	45	45	45	*45*	42	42	**42**
Bias ($\times 10^{-3}$)								
Average	6.0	10.5	**1.0**	5.6	*-15.0*	-6.2	-3.6	-3.2
Urban	*-105*	-103	11.1	8.3	-93	-91	**1.0**	3.1
Rural	77	*83*	-5.4	**4.0**	35	48	-6.5	-7.2
CPO								
Average	0.22	*0.21*	0.25	0.24	0.26	0.25	**0.27**	0.26
CRPS								
Average	0.17	*0.21*	0.16	0.18	0.17	0.18	**0.15**	0.17

Table 3.2: Validation results calculated at the cluster level when leaving out one county at a time for the 2014 Kenya DHS secondary education data for women aged 20–29. The worst entries in each row are in *italics*, while the best entries in each row are in **bold**. In the table, the figures are rounded, but minimum and maximum were evaluated with more significant figures.

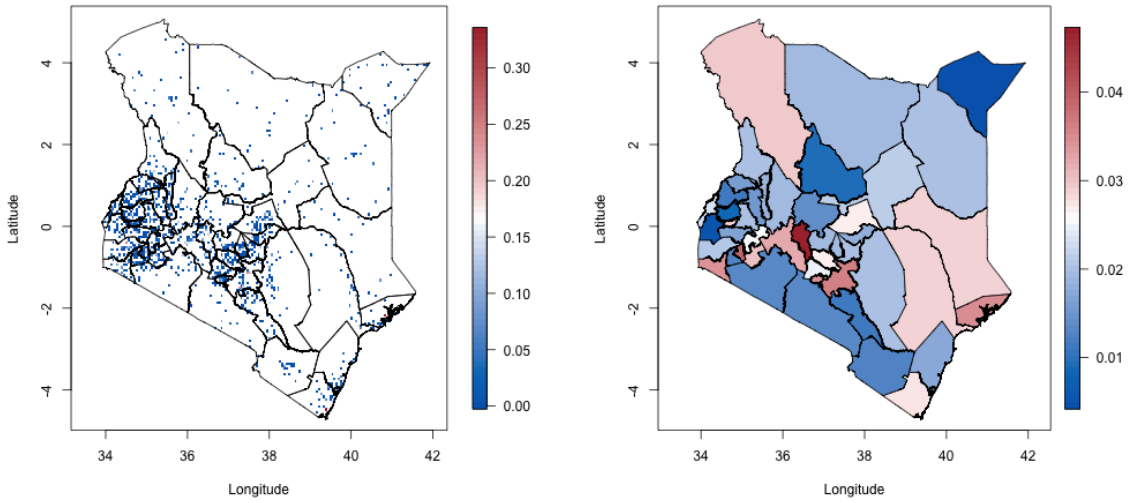


Figure 3.5: Empirical average of neonatal mortality rates in Kenya from 2010-2014 based on data from the 2014 Kenya DHS. Values are shown at both the cluster (left) and county levels (right).

3.4.2 Risk of Neonatal Mortality

Risk Mapping

We now estimate neonatal mortality rates (NMRs) for children in Kenya from 2010–2014 based on data from the 2014 Kenya DHS plotted in Figure 3.5. The figure shows that the vast majority of clusters have empirical NMRs very close to zero, though there are some clusters that have much higher NMRs with some even above 30%. Central NMR estimates, and 80% uncertainty interval widths in Figure 3.6, and individual county level predictions are given in Appendix A.2 along with upper and lower predictive quantiles at the county level. The largest NMRs were estimated to be in several counties just northwest of Nairobi as well as in central eastern Kenya, and the lowest NMRs were estimated to be in the counties near the central western and southwestern borders. Although we expected to find a

significant urban effect, we found little evidence suggesting any difference in NMRs between rural and urban areas. Additionally, there is much less spatial variation relative to cluster level variance. For instance, the BYM2_{UC} and SPDE_{UC} models estimate the spatial effect variance to respectively be 0.060 and 0.069, whereas the estimated cluster effect variances were respectively 0.18 and 0.23. We found little difference in the quality of predictions of the models when we validated them using the same method as above, indicating that the low signal-to-noise ratio in this application makes improvement in spatial estimation difficult.

The fitted smoothed direct, BYM2_{UC}, and SPDE_{UC} models had smaller spatial effect variances relative to the predictive uncertainties when compared to the analysis of secondary education prevalence. For the SPDE_{UC} model, for instance, this is evidenced by the fact that the median 80% credible interval width for the NMR data is 0.0074, whereas the point estimates have a range of 0.0098. The equivalent values in the secondary education application, which are respectively 0.062 and 0.476, indicate much greater variability across space. The estimated variances of the county level random effects are also relatively small, and are estimated to be 0.059 and 0.060 for the smoothed direct and BYM2_{UC} models, respectively. The variance of the spatial component of the SPDE_{UC} model was estimated to be 0.069. The cluster effect variance, however, was estimated to be comparatively larger, with BYM2 and SPDE_{UC} estimates of 0.183 and 0.225 respectively, further indicating the large amount of noise relative to any spatial signal in the data. In spite of the lack of spatial signal, the spatial smoothing models helped to reduce the predictive uncertainties relative to the naive and direct models as evidenced from the plotted credible interval widths of the predictions in Figure 3.6.

Considering the difficulty of including cluster level variation when aggregating SPDE_{UC} predictions to the county level, combined with the fact that a substantial majority of the variation in the data occurs at the cluster level as opposed to spatial level variation, we believe the smoothed direct and BYM2 models might be better suited for this particular application.

Not including variation due to cluster effects in the spatial aggregation, aside from integrating out cluster level variation, is probably what leads the SPDE_{UC} model predicted NMRs to have such narrow credible intervals relative to the other models.

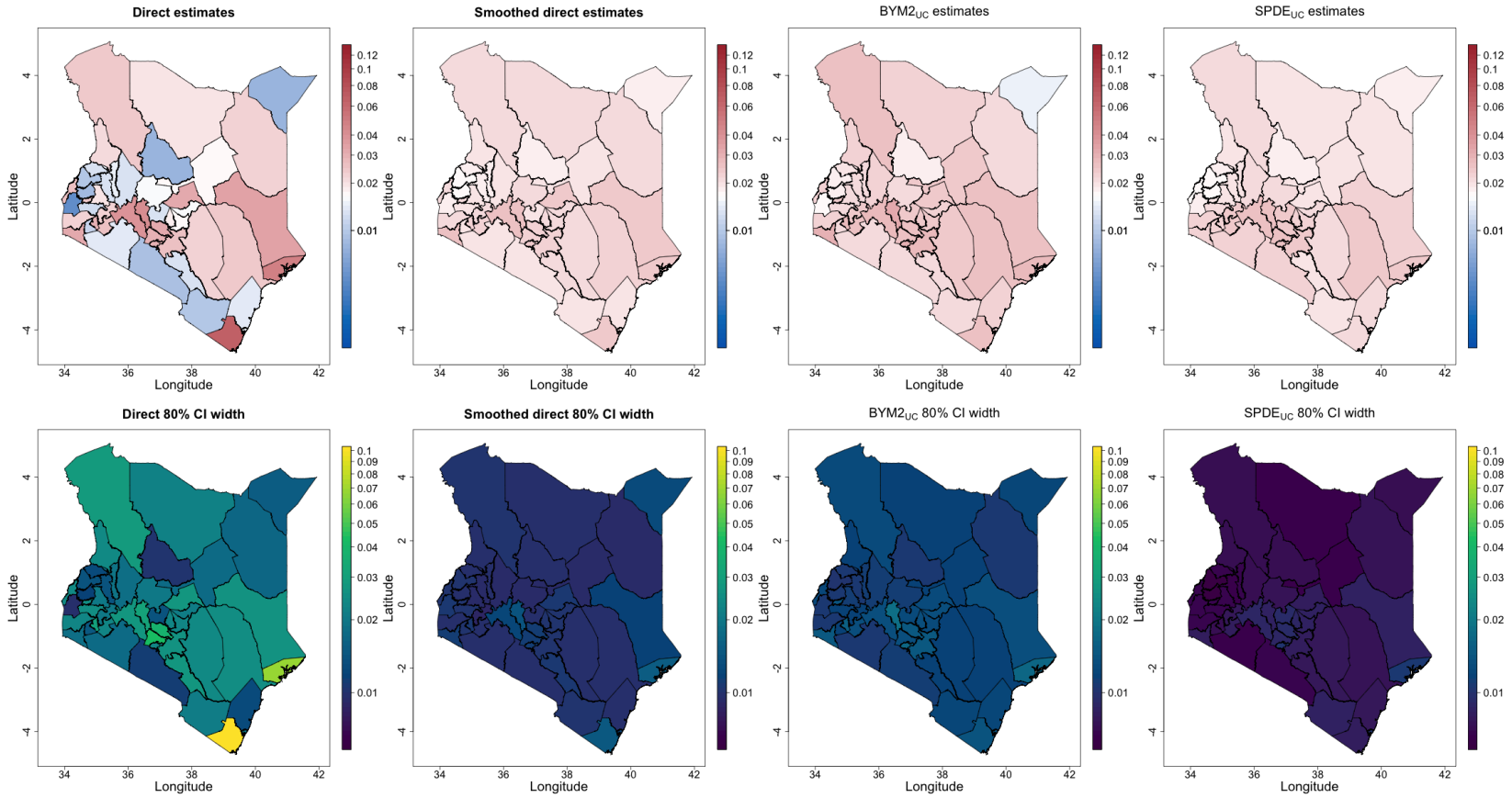


Figure 3.6: Kenya county level neonatal mortality rate predictive mean (top row) and 80% uncertainty interval widths (second row) from 2010-2014. The BYM2 and SPDE model predictions include both urban and cluster effects.

	Est	SD	Q10	Q50	Q90
Smoothed Direct					
Intercept	-3.85	0.07	-3.94	-3.85	-3.76
BYM2 Phi	0.30	0.25	0.04	0.23	0.70
BYM2 Tot. Var	0.06	0.04	0.02	0.05	0.11
BYM2 Spatial Var	0.02	0.02	0.001	0.01	0.05
BYM2 iid Var	0.04	0.03	0.009	0.03	0.08
BYM2 Tot. SD	0.23	0.08	0.13	0.22	0.33
BYM2 Spatial SD	0.11	0.07	0.04	0.10	0.22
BYM2 iid SD	0.19	0.08	0.09	0.18	0.29
$\boldsymbol{BYM2_{UC}}$					
Intercept	-3.99	0.09	-4.11	-3.99	-3.87
Urban	0.08	0.11	-0.07	0.08	0.22
Cluster Var	0.18	0.10	0.07	0.17	0.31
BYM2 Phi	0.42	0.31	0.044	0.36	0.88
BYM2 Tot. Var	0.06	0.04	0.02	0.05	0.11
BYM2 Spatial Var	0.02	0.02	0.002	0.02	0.05
BYM2 iid Var	0.04	0.03	0.004	0.03	0.08
Cluster SD	0.41	0.11	0.27	0.41	0.56
BYM2 Tot. SD	0.24	0.07	0.15	0.23	0.33
BYM2 Spatial SD	0.14	0.07	0.05	0.13	0.23
BYM2 iid SD	0.17	0.08	0.07	0.17	0.29
$\boldsymbol{SPDE_{UC}}$					
Intercept	-4.00	0.09	-4.12	-4.00	-3.89
Urban	0.08	0.11	-0.06	0.08	0.22
Range	241	195	79	178	451
Spatial Var	0.07	0.06	0.02	0.05	0.14
Spatial SD	0.24	0.10	0.13	0.23	0.38
Nugget Var	0.23	0.10	0.11	0.21	0.37
Nugget SD	0.46	0.11	0.33	0.46	0.61

Table 3.3: Parameter and hyperparameter estimate summary statistics from the BYM2 and SPDE models including both urban and cluster effects, fit to the 2010-2014 NMR data from the 2014 Kenya DHS.

Validation

We validate the spatial smoothing models that produce predictions at the cluster level by once again leaving out data from each county, one county at a time, and making predictions at the cluster level. Equivalent scoring rules to those used in Section 3.4.1 are calculated for the NMR dataset in Table 3.4, which shows that the SPDE_{UC} model performs just as well as any of the other spatial smoothing models when making cluster level predictions. In fact, all of the models perform very similarly, again suggesting that variation in the NMR data is primarily at either the cluster level or at spatial scales too fine to easily identify.

3.5 Discussion and Conclusions

Direct estimators remain the gold standard, provided there are sufficient data for an associated variance that is of acceptable size. The smoothed direct estimator can reduce the variance using the totality of data, albeit with an introduction of some bias due to the smoothing. This bias disappears, however, as sample size increases. When the direct estimates are unreliable, one is led to modeling at the cluster-level, and it is important to use a model that is consistent with the design. In this analysis, we introduced a binomial sampling model with discrete spatial random effects, and it performed well in the simulations and real applications. The BYM2 random effects capture different levels in the regional strata. Hence the random effects should be nested within the regional strata to allow for different levels for the different sampling strata. Our method is designed for situations in which the sampling probabilities depend on stratification. Extending the methods to account for more complex sampling plans, non-response, and calibration will be the subject of future research.

We have also been experimenting with a beta-binomial model that allows for overdispersion (within-cluster variation). Such approaches with discrete spatial models do not deal well with combining data at different geographical resolutions, and this is where the continuous spatial model is appealing. Unfortunately, aggregation with continuous models is the most

	BYM2				***SPDE***			
	uc	uC	Uc	UC	uc	uC	Uc	UC
MSE ($\times 10^{-4}$)								
Avg	24.5	24.6	24.6	24.6	24.6	24.6	24.6	24.6
Urban	29.6	29.6	29.6	29.7	29.6	29.6	29.7	29.7
Rural	21.3	21.4	21.3	21.4	21.4	21.4	21.4	21.4
Var ($\times 10^{-4}$)								
Avg	24.5	24.5	24.6	24.6	24.6	24.6	24.6	24.6
Urban	29.6	29.6	29.6	29.6	29.6	29.6	29.6	29.7
Rural	21.3	21.3	21.3	21.3	21.4	21.4	21.4	21.4
Bias ($\times 10^{-4}$)								
Avg	8.2	26.0	8.8	27.1	4.1	3.3	5.1	7.5
Urban	-0.1	17.7	11.0	29.0	-3.0	-4.0	8.0	10.9
Rural	13.4	31.3	7.4	25.9	8.6	7.9	3.2	5.3
CPO								
Avg	0.657	0.649	0.657	0.649	0.659	0.662	0.659	0.660
CRPS								
Avg	0.024	0.025	0.024	0.025	0.024	0.025	0.024	0.025

Table 3.4: Validation results calculated at the cluster level when leaving out one county at a time for the 2014 Kenya DHS NMR data from 2010-2014.

difficult, since a population surface is required, and the model and population surface must, in general, be stratified by urban/rural.

In the simulations and application, we found that accounting for the design nearly always improved predictions in the case that strata were associated with the response, and did not reduce predictive performance otherwise. This was true whether the design was accounted for using sampling weights or by including stratification indicators as covariates. Although not included in our results, we have found that when the proportions urban required for aggregating strata predictions to the county level were biased, including stratification effects in the BYM2 model sometimes made the predictions worse. This implies that not only must design stratification be accounted for, but in the case where it is included as a covariate, it is important to make an effort to obtain high quality estimates of the proportions of the studied population in each strata. In practice, this may be difficult.

Although we expected the SPDE models without urban effects to have better predictions than BYM2 models without urban effects since urbanicity is a spatial variable, we instead found that the spatial component of the SPDE models without urban effects had difficulty handling the sharp changes of urbanicity over short distances. As mentioned previously, WorldPop and IHME do not adjust for the urban/rural stratification, and WorldPop does not account for cluster level overdispersion, but they do include extensive covariates such as population density, which will, to some extent at least, adjust for urban/rural.

The question remains how best to include cluster effects in area-level aggregated predictions from spatial models. Since the SPDE model predictions are aggregated to the county level by numerically integrating predictions on a spatial grid, whereas cluster effects are modeled discretely at cluster and EA point locations, using a continuous integral to represent areal predictions of risk and prevalence may be problematic in the presence of large cluster effects. Simply leaving out cluster effects when aggregating predictions spatially may lead to undercoverage and also bias. Integrating out the cluster effect eliminates the bias, but not

the undercoverage. In order to account for cluster level variation, the true data-generating mechanism for the cluster effects should ideally be a consideration. Although we assume cluster effects are true latent differences in mortality rates, and fixed for each EA, it is possible they are due to measurement error, for instance, in which case they should be left out of predictions. We explore cluster level variation and how to more accurately account for the data generation process in Chapter 4.

The simulation study and prevalence application indicated that the smoothed direct model had the least dependence on the specific implementation, performing well in nearly all circumstances, whereas the SPDE and BYM2 models that included stratification and cluster effects performed particularly well when there was a stratification effect in the population. This was especially the case if the proportion of the population of interest (i.e., children, or women aged 20–29) that is urban in each county is accurately known. The BYM2 model with urban and cluster effects performed the best or nearly the best in all scenarios and with all populations in the simulation study, in terms of MSE, CRPS, and credible interval width. The SPDE model including urban and cluster effects performed better in the cluster level validation, but care must be taken in selecting a prior due to its flexibility, and in generating spatially aggregated predictions when the estimated cluster effect accounts for a large proportion of the total variation.

In the simulation study, the DHS we emulated was powered to the Admin2 level, which coincided with the level of inference. More commonly, DHS data are powered to the Admin1 level and it is an open question as to what the recommendations are in this case if inference is still required at Admin2. In other work (Li et al., 2019) we could only perform Admin1 level inference for countries in Africa using the majority of the DHS surveys, because there were insufficient samples to apply the direct and smoothed direct methods at the Admin2 level.

Chapter 4

ACCOUNTING FOR CLUSTER LEVEL VARIATION IN SPATIAL MODELS OF AGGREGATE POPULATION PREVALENCES

4.1 Introduction

Cluster level variation, i.e., variation attributable to a cluster level unstructured random effect, is present in DHS household surveys, and, as discussed in Chapter 3, it is a large source of variation in neonatal mortality outcomes in Kenya. Although in Chapter 3 cluster random effects were included in some of the models, and predictions were adjusted by integrating out those cluster level effects, predictive variances were not directly increased by the cluster level variation. This is due to the fact that no spatial statistical approach has been proposed to directly account for this extra variation when producing estimates for areal population averages in a rigorous way. Although Dong and Wakefield (2020) considered a model for a 'true signal' interpretation of the cluster effect, rigorous methods for aggregating true signal predictions were not proposed. IHME, WorldPop, and others frequently use approaches that either do not include cluster effects in the data model or do not adequately account for them, such as by incorrectly assuming there is exactly one cluster random effect per spatial grid cell or leaving it out when aggregating to the areal level even though it was included in the data model (see, e.g., Golding et al., 2017; Osgood-Zimmerman et al., 2018; Graetz et al., 2018; Utazi et al., 2018a; Neal et al., 2019; Giorgi et al., 2018). Ultimately this is due to the inherent difficulty of accounting for variation at discrete EA locations in a continuous integral for areal risk. The number of EAs per grid cell is not known, and the variation depends on the population density as well as the total number of EAs per area.

The difficulties of accounting for cluster level variation in a model based framework primarily arise from three sources. First, the number of EAs per spatial area is rarely known at scales finer than the stratum level (for DHS household surveys the stratum is typically Admin1 crossed with urban/rural). This becomes relevant when making predictions at the Admin2 level, for instance. Second, the EA spatial coordinates are not known. Even if the number of EAs per area of interest were known, their exact locations within that area are important if there is any spatial variation in the risk within the area. For continuous spatial models the second point is especially important, since risk is assigned continuously through space, even within areas. Third, the number of members of the target population, the population denominator, in each EA is not known. Although it is possible to give each EA in a given stratum equal weight when producing areal risk predictions, ignoring this additional variation in the population denominator among EAs, the population denominators are relevant for accounting for population numerator variation when estimating prevalence. No study has explored how to accurately account for cluster level variation and population numerator and denominator variation when estimating areal population prevalence using continuous spatial models when EA locations are unknown.

Aside from the difficulties of accounting for cluster level variation in small area models, the models considered in this chapter are also motivated by another question: should the lack of information on EA locations play a significant role in small area model predictions when fit from multistage cluster samples? While the exact EA locations within a stratum may have no effect on the predictions of areal spatial models, they do have some effect on continuously indexed spatial model predictions. If exact EA locations were known, then aggregating predictions over the 96,251 EAs in Kenya would be simpler, but since they are not, there is an additional source of error in the model predictions. Whether this effect is important enough for the associated uncertainty to be explicitly modeled has never been determined.

Lastly, and regarding population numerator and denominator variation, it is important to emphasize as we have done previously that prevalence and risk are two different concepts. We take 'prevalence' to mean the empirical proportion of the population with the given characteristic, whereas we take 'risk' to be the expected prevalence, averaging over hypothetical populations generated from a superpopulation sampling model. For example, for a single person, the prevalence is the binary outcome of whether they have a characteristic, whereas the risk is the probability that the individual has the characteristic under a Bernoulli superpopulation model for the individual's response. We can then also think of the population prevalence as the average of the prevalences for each individual in the population, and the population risk is the average of risks for each individual under part of our superpopulation sampling model.

Note that risk depends on the superpopulation sampling model of interest, and which part of the superpopulation model we take expectation over. If our superpopulation sampling model allows us to draw random EA locations as well as population numerators and denominators among other quantities of interest, we may take expectation over any or all of these quantities when calculating risk. However, since risk is typically thought of as a probability, it is important to at least take expectation over population numerator variation. The more other parts of the superpopulation sampling model we take expectation over, the more sources of variation get eliminated. Partly because of this, and partly due to the fact that we are viewing the target population as a possible draw from our superpopulation model, including all sources of variation in the superpopulation sampling model, as a general rule we believe that averaging out only population numerator variation is most consistent with the population itself and should therefore be preferred. It is because of this that we believe that accounting for lack of information on EA locations, cluster level variation, and population numerator and denominator variation is especially important when estimating prevalence. This will be discussed in more detail when we define the superpopulation model we propose

for neonatal mortality prevalence in Kenya.

What may add to the confusion in distinguishing between risk and prevalence is that, as more iid variation is averaged over among the EAs within an area, such as iid cluster effects and iid (conditional on the risk) EA population numerator and denominator variation, prevalence and risk become increasingly difficult to distinguish between. Hence, for sufficiently large areas with many EAs, it may not be unreasonable to use risk and prevalence interchangeably. In this chapter, however, we would like to determine whether there are substantial differences between prevalence and risk in practice, specifically in the context of estimating NMRs in Kenya using the 2014 KDHS. The models for risk used in Chapter 3 achieved near nominal rates of coverage at the county (Admin1) level. However, estimates of NMR at the Admin2 level, total deaths, and relative prevalences of urban and rural parts of an area are also of interest. The distinction between risk and prevalence may become more important when estimating these quantities.

All of the above difficulties and questions relate to the necessity of modeling, or at least understanding, individual EAs—either their locations or their population denominators. In this chapter we will explore how to directly account for these effects by modeling them directly in the context of estimating neonatal mortality prevalence in Kenya. In Chapter 3 we observed cluster level variation that was roughly three times larger than spatial variation when estimating NMRs. This is therefore a context where these modeling questions may be especially important.

In order to address the above gaps in the literature, we develop a general framework we call combined population aggregation models (CPAM) for building spatial models for prevalence that can account for these effects. CPAMs are composed of a point level risk model indexed at cluster point locations, and an aggregation model for calculating population prevalence in the areas of interest. CPAMs are flexible in that any fitted risk model indexed at the cluster level can be converted to a CPAM when combined with an aggregation model

that can be directly implied by available population density surfaces and census data. We compare prevalence predictions of neonatal mortality in Kenya from a CPAM with those of the risk model from which it is built, and find that, especially in less populous, geographically small areas, the CPAMs produce significantly wider credible intervals.

The rest of this chapter is organized as follows. In Section 4.2, we set up the problem by introducing the likelihood model for the response at each cluster, and then build the CPAM framework step-by-step from the SPDE_{UC} model introduced in Chapter 3, showing that the SPDE_{UC} model is an example of a CPAM. We conduct a simulation study in Section 4.3 using simulation parameters intended to simulate populations similar in several ways to the SPDE_{UC} model fit using the 2014 KDHS NMR data, evaluating the SPDE_{UC} model as well as other CPAMs. We then apply all considered CPAMs to the 2014 KDHS NMR data in Section 4.4, estimating neonatal mortality prevalence, total counts, and relative prevalences at the Admin2 (constituency, for Kenya) level, assessing the predictions via the scoring rules of Section 2.4 in a validation exercise. Lastly, we discuss conclusions in Section 4.5.

4.2 A Continuous Spatial Framework for Population-Level Aggregates

4.2.1 Problem Setup

We assume the following cluster level model for neonatal mortality in the sample for area (usually county) i and cluster c:

$$y_{ic}|n_{ic},\mu_{ic} \sim \text{Binomial}(n_{ic},\mu_{ic}) \tag{4.1}$$

$$\text{logit}(\mu_{ic}) = \eta_{ic} = \eta^S_{ic} + \epsilon_{ic} = \mathbf{X}_{ic}\boldsymbol{\beta} + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}, \tag{4.2}$$

where the response is the number of neonatals that died in a given cluster, $\epsilon_{ic} \sim N(0,\sigma^2_\epsilon)$ is a cluster level random effect that can allow for dependency between observations in the same cluster, and n_{ic} is the known number of neonatals in the sampled clusters. Spatial effects in the linear predictor are denoted by $\eta^S_{ic} = \eta^S(\boldsymbol{s}_{ic})$ representing a combination of spatial

random and spatially indexed covariate effects. We also assume the following model for our population responses at the EA level:

$$Y_{ic}|N_{ic}, \mu_{ic} \sim \text{Binomial}(N_{ic}, \mu_{ic}), \tag{4.3}$$

where Y_{ic}, N_{ic} are the number of neonatal deaths and the total number of neonatals born in 2010–2014 respectively in EA c in area i. Hence, within the area we have $|B_i|$ distinct prevalences, where B_i is the set of EA indices in area i, and $|\cdot|$ denotes the set cardinality. The count Y_{ic} is the number of neonatals in EA c and area i, where $\boldsymbol{s}_{ic}$ is the spatial location of EA c in the area. In our case, only the sampled EA locations are known, and the vast majority are unsampled.

Available data for fitting the model includes a DHS in the form of a multistage household survey stratified by county crossed with urban/rural classification. This dataset will be denoted $\mathcal{Y}$. In addition, another dataset such as census data, $\mathcal{W}$, provides information on the number of EAs per area and stratum, households per EA, neonatals per household, and neonatals per area and stratum. Lastly, we have available a high-resolution spatial surface, $\tilde{q}(\boldsymbol{s})$, that is proportional to an estimate of the overall population density throughout the considered country over a fine pixelated grid (5km $\times$ 5km) indexed by g. This could be provided by WorldPop, for instance, and can be adjusted at every pixel using census data, $\mathcal{W}$, to make $q(\boldsymbol{s})$, an estimate of the continuous spatial population density surface of neonatals (or the population density of another target population) that is correctly normalized to have unit integral in each area of interest. In other words, the census provides information about how many neonatals there are per person in each stratum, and the overall population density is then adjusted to estimate the density of neonatals per pixel. It is important to account for target population density when producing population aggregates from continuously indexed spatial predictions, since population aggregates not accounting for this will be biased in the presence of spatial variation within the area.

We have approximate 23,440 pixels used for numerical integration and over which overall

population and target population densities are evaluated. There are also 96,251 EAs in total in Kenya, yielding approximately 4.1 EAs per pixel on average. Of course, the number of EAs per pixel varies depending on the population within each pixel and its urbanicity designation since each EA within a given stratum has similar overall populations.

The main targets of inference are the proportion of neonatals that died at different aggregation levels. To begin, we can start by considering inference at the area level, and at the pixel level, where the pixels are small enough for the spatial component of the linear predictor, η^S, which includes an urban effect to account for stratification, to not vary significantly within its boundaries. The targets of inference in area i and pixel g are:

$$p_i = \sum_{c \in B_i} \frac{N_{ic}}{N_i} \times \frac{Y_{ic}}{N_{ic}} \tag{4.4}$$

$$p_{ig} = \sum_{c \in B_i^g} \frac{N_{ic}}{N_{ig}} \times \frac{Y_{ic}}{N_{ic}}, \tag{4.5}$$

where $N_i \equiv \sum_{c=1}^{B_i} N_{ic}$ and $N_{ig} \equiv \sum_{c \in B_i^g} N_{ic}$ are the number of neonatals in area i and pixel g respectively. Note that the set B_i^g gives the indices of the EAs that are in pixel g, and that p_i and p_{ig} are prevalences rather than risks.

Equations (4.4) and (4.5) represent how we would calculate the prevalence in an area or a pixel if we knew the exact number of EAs in that area or pixel along with the exact population denominators and responses, but in practice we do not always have access to this information, such as in many low and middle income countries at small enough areal levels. For instance, the number of EAs is known at the Admin1 level crossed with urban/rural designations, but not at the Admin2 level. In these cases, the goal will be to model each of these terms and to explore what happens when these unknown quantities are directly modeled and plugged into the above aggregation formulas.

For a given area, i, we consider five random vectors that lead to different sources of variation: $\boldsymbol{\eta}_{i:}^S, \boldsymbol{s}_{i:}, \boldsymbol{\epsilon}_{i:}, \boldsymbol{N}_{i:}$, and $\boldsymbol{Y}_{i:}$ where ':' denotes the EA index c varying over the values in B_i. For instance, $\boldsymbol{\eta}_{i:}^S$ denotes the vector $(\eta_{i1}^S, \ldots, \eta_{i|B_i|}^S)^T$. Under a given superpopulation

model for the above vectors, we could consider models for aggregate risk that, for instance, take expectations over all sources of variability but the spatial component, $\boldsymbol{\eta}_{i:}^{S}$, which we will label the 'S' aggregation model. We could also take expectation over population numerator and denominator variation only, including cluster level variation in the aggregate risk. We will label that the 'SC' aggregation model since it includes spatial and cluster level variation. Lastly, we could choose to model prevalence itself. Since this also includes population numerator and denominator variation, we call this the 'SCP' aggregation model. Table 4.1 shows the aggregation models considered as well as which sources of random variation in the superpopulation model are expected over when calculating risk in the suerpopulation model. Since the SCP aggregation model does not take expectation over the superpopulation model, it represents the prevalence for a given draw from the assumed superpopulation model.

Aggregation Model	**Aggregation Equations**
SCP	p_i
	p_{ig}
SC	$p_i \approx E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:}}[p_i]$
	$p_{ig} \approx E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:}}[p_{ig}]$
S	$p_i \approx E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:},\boldsymbol{\epsilon}_{i:},\boldsymbol{s}_{i:}}[p_i]$
	$p_{ig} \approx E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:},\boldsymbol{\epsilon}_{i:},\boldsymbol{s}_{i:}}[p_{ig}]$

Table 4.1: Areal and pixel level (unsimplified) aggregation equations under three considered combined population aggregation models, in order of decreasing predictive variance. Subscripts in the expectations denote the variables in the superopulation model over which expectations are taken when calculating risk.

The symbols S, C, and P denote whether only spatial variation is accounted for or whether

cluster level variation and population numerator and denominator variation are also accounted for. Expectations are calculated over the EA locations ($\boldsymbol{s}_{i:}$), cluster effects ($\boldsymbol{\epsilon}_{i:}$), population denominators ($\boldsymbol{N}_{i:}$), and population numerators ($\boldsymbol{Y}_{i:}$) in Table 4.1 Interpretations of these models as well as derivations are given in Section 4.2.2. However, since all of them can be related to each other via conditional expectations, the tower property of expectation implies that the central predictions of each of these models are identical, with differences between them being only in the amount of uncertainty in the predictions and their interpretations.

Our main goal will be to explore how best to aggregate from the cluster level to the pixel and area levels in a way that balances quality of the predictive distribution at multiple aggregation levels, and computational feasibility. An important side goal of this project will be to determine the most important sources of uncertainty when conducting inference at both area and pixel (5km resolution) levels. In the ideal scenario, we would validate the proposed model at the pixel level, and show that the data is well characterized by the predictive distribution even at such fine spatial scales.

We will now describe a modeling framework for aggregation spatial predictions that are continuously indexed to the areal level, since this is of particular interest for policymaking. We will consider continuous spatial models for population-level aggregates as being composed of two parts: a 'cluster level model,' a continuous spatial model for the response, in this case neonatal mortality, for a given cluster existing at a spatial point, and an 'aggregation model'. An aggregation model consists of a set of equations akin to equations (4.4) and (4.5) we will call 'aggregation equations' so that joint posterior draws from the point level model and any random variates in the aggregation model can be aggregated to different areal levels. We call the combination of point level and aggregation models a 'combined population aggregation model' (CPAM), or a 'combined model' for short. After introducing an existing CPAM, we then point out some problems it has when making predictions at different aggregation levels

from pixel to county and to even coarser aggregation levels. We then introduce and describe a solution to this problem by modifying the aggregation model.

4.2.2 A Geospatial Model for Neonatal Mortality: Cluster Level Model

For a cluster level model, we assume the linear predictor for area i and EA c takes the following form:

$$\eta_{ic} = \eta^S(\boldsymbol{s}_{ic}) + \epsilon_{ic} = \beta_0 + u(\boldsymbol{s}_{ic}) + \beta^{\text{URB}} I(\boldsymbol{s}_{ic} \in U) + \epsilon_{ic}, \tag{4.6}$$

where β_0 is the intercept, $u(\boldsymbol{s}_{ic})$ is a mean zero spatial random effect, β^{URB} is the association with the cluster being urban (as compared to rural), and $\epsilon_{ic} \overset{iid}{\sim} \text{N}(0, \sigma_\epsilon^2)$ is the cluster effect of cluster c in area i. The indicator $I(\boldsymbol{s}_{ic} \in U)$ is 1 if $\boldsymbol{s}_{ic}$ is in the set of urban locations, U, and 0 otherwise. The term $\beta^{\text{URB}} I(\boldsymbol{s}_{ic} \in U)$ is included in order to account for stratification by urban/rural, which is typical in this context (ICF International, 2012). We model u as a Gaussian process with Matérn spatial covariance, smoothness parameter $\nu = 1$, spatial variance σ_s^2, and spatial effective range ρ. The point level model can be fit using integrated nested Laplace approximations (INLA), setting PC priors (Simpson et al., 2017; Fuglstad et al., 2019a) on the covariance parameters so that the median of the prior for ρ is one fifth the diameter of the spatial domain, and so that $P(\sigma_s > 1) = P(\sigma_\epsilon > 1) = 0.01$ as in Chapter 3.

4.2.3 Continuous Aggregations of Neonatal Mortality: S Aggregation Model

We now introduce slight variation on the SPDE aggregation model introduced in Section 3.2 that we call the 'S' model, where S denotes the fact that spatial variation is accounted for. If the complete list of EA locations in the sampling frame were known, we could predict at the county level using the posterior distribution of a weighted sum of the predicted probabilities

(risks), say $\hat{\mu}_{ic}$, at the EA locations using (4.4) and (4.5). The vast majority of EA locations are unobserved, however. To deal with the difficulties of adding in cluster effects in the predictions, we can consider generating approximate central predictions by integrating out the cluster effect from the mortality rate at any location:

$$\mu(\boldsymbol{s}_{ic}) \approx E_{\epsilon_{ic}}[\mu(\boldsymbol{s}_{ic})] = \int_{-\infty}^{\infty} \text{expit}(\eta^S(\boldsymbol{s}_{ic}) + \epsilon) f_\epsilon(\epsilon)\, d\epsilon \tag{4.7}$$

$$\approx \text{expit}\left(\frac{\eta^S(\boldsymbol{s}_{ic})}{\sqrt{1 + k^2\sigma_\epsilon^2}}\right), \tag{4.8}$$

for $k = 16\sqrt{3}/(15\pi)$, using a logistic approximation to the Gaussian likelihood as given in Wakefield (2013). It is worth noting that under some interpretations of the cluster effect, such as when it represents the overdispersion in the NMR when resampling multiple times from an EA, the integral on the right of (4.7) is in fact the correct target of inference when estimating risk. However, we will instead use a 'true signal' interpretation of the cluster effect in this analysis, in which case the left hand side of (4.7) is the risk in a given EA rather than the right hand side. The true signal and overdispersion interpretations of the cluster effect are explored in Dong and Wakefield (2020). The right hand side of (4.7) has the advantage that it helps to simplify the handling of the 96,215 EAs in the predictions, enabling aggregate predictions to be made by integrating a continuous spatial surface, but it is only an approximation under our true signal interpretation of the cluster effect, approximating the cluster level risk with its expectation over the distribution of the cluster effect.

The results in (4.7) and (4.8) enables us to consider an approximation to $\mu(\boldsymbol{s}_{ic})$ that is a smooth function of spatial location rather than as a quantity that only exists at a given EA location, since cluster effects have been integrated out of the expression. Hence, rather than aggregating over all EAs individually, we can aggregate the approximate risk, $\tilde{\mu}(\boldsymbol{s}_{ic}) \equiv E_{\epsilon_{ic}}[\mu(\boldsymbol{s}_{ic})]$, which is smoothly varying in space and that agrees in expectation with $\mu(\boldsymbol{s}_{ic})$ due to (4.7). For more information about why the considered aggregation models agree in expectation for areal aggregates, see Appendix B.4. The approximate risk can be

aggregated with respect to population density to get an estimated average NMR over the population, we can plug in a continuous risk surface estimate making our estimate for area i's neonatal mortality:

$$\hat{p}_i = \int_{A_i} \tilde{\mu}(\boldsymbol{s}) \times q_{A_i}(\boldsymbol{s})\ d\boldsymbol{s} \tag{4.9}$$

$$\hat{p}_{ig} = \tilde{\mu}(\boldsymbol{s}_{ig}) \tag{4.10}$$

where A_i is the geographical extent of area i, $\boldsymbol{s}_{ig}$ is the spatial location (centroid) of pixel g in area i, and where for a given set of spatial locations, A, we define the following normalizations of the general population and the target population (neonatal) density functions:

$$q_A(\boldsymbol{s}) = \frac{q(\boldsymbol{s})}{\int_A q(\boldsymbol{s})\ d\boldsymbol{s}}$$
$$\tilde{q}_A(\boldsymbol{s}) = \frac{\tilde{q}(\boldsymbol{s})}{\int_A \tilde{q}(\boldsymbol{s})\ d\boldsymbol{s}},$$

for overall population density $\tilde{q}$ and target population density q, and where both are functions of spatial location. The integrals in the above definition and in (4.9) can be approximated as a numerical summation over the pixel grid. This aggregation model will be called the S aggregation model for reasons described in the next subsection.

in areas whose the urban and rural parts may be very fine scale, the integrals in $q_A(\boldsymbol{s})$ and $\tilde{q}_A(\boldsymbol{s})$ are estimated on a 1km $\times$ 1km grid rather than a 5km $\times$ 5km scale. This ensures that the proportion of population assumed by the model in each area that is urban and rural is more accurate, reducing bias in the areal predictions. We also combine Admin2 areas in Kenya that have spatial area less than 50 km^2 with neighboring ones in the same Admin1 area in order to reduce numerical approximation errors caused by representing small areas using the pixel grid, resulting in 273 (possibly combined) Admin2 areas in total out of the original 290. In particular, this ensures that every Admin2 area with any urban general population (as designated via our general population thresholding method) includes urban pixels and every Admin2 area with any rural general population includes rural pixels.

4.2.4 Challenges with the S Aggregation Model

While the aggregation model introduced by Equations (4.9) and (4.10) in Section 4.2.2 is an intuitive model in that the population estimate is represented as an integral with respect to population density, it has several problems. First, and perhaps most obviously, the variation in the population denominators N_i, N_{ig}, and N_{ic}, and also the population numerator variation of Y_{ic} appear explicitly in the targets of inference in equations (4.4) and (4.5), yet appear nowhere in the above estimators. A second problem is in the approximation of $\mu(\cdot)$ by its expectation (expecting over cluster variation) in (4.7). If the cluster variance is large, and if there are few clusters being averaged over, then this approximation becomes less accurate, and the uncertainty in the predictions will be underestimated. Intuitively, the variation being ignored by the model in those terms matters less for estimates at sufficiently coarse aggregation scales. At finer scales, however, these problems can be much more serious. In fact, pixel level estimates for continuous spatial models based on household survey data are problematic in part due to their anti-conservative predictive distributions, and also because they do not directly address the possibility that there could be no one living in a pixel, in which case it would not make sense to make predictions for that pixel from a finite population standpoint.

Both of the above problems arise from the EA locations not being known, and being represented as a continuous population density surface in the aggregation model, but as discrete point locations in the cluster level model. Since the S aggregation model can be thought of as a model for risk, averaging out many sources of variation in our assumed superpopulation model, at fine aggregation scales where the averaged out sources of variation may make more difference, the predictions of this model become problematic. In spite of this, model based estimates of risk are often compared directly with survey estimators for prevalence such as the Horvitz-Thompson estimator without distinguishing between their different targets of inference (Wilson and Wakefield, 2018). This relies on the assumption

that risk and prevalence tend to be similar for large populations (since risk is expected prevalence), which is rarely tested in practice.

Throughout this analysis we label aggregation models with the letters S, C, and P depending on whether they account for spatial variation in the risk, cluster level random effects and EA location uncertainty, and population denominator and numerator variation respectively. This is why we refer to the aggregation model given in Section 4.2.3 as the S aggregation model.

4.2.5 Modeling Cluster Level Variation: SC Aggregation Model

In order to address the aforementioned problems with the S aggregation model, we will start with modifying it by directly modeling the unknown EA locations, $\boldsymbol{s}_{i:}$, for each area i. This will enable us to model exactly how many cluster level random effects to add when making predictions for any given area. For simplicity, we assume that EA locations are independent of each other conditional on an underlying intensity surface representing the probability density of the EA locations being at any location, since relaxing this assumption would require a substantially more complex model. In addition to assuming conditional independence, we assume that the probability of an EA being at a certain location within area i and stratum k (being either urban or rural) is proportional to that location's population density. Under these two assumptions, the EA locations in area i and stratum k *must* follow a Poisson process with intensity proportional to the continuously indexed general population density surface, $\tilde{q}(\boldsymbol{s})$ for spatial location $\boldsymbol{s}$.

Since the totals in each stratum are typically known, the EA locations within each stratum follow what is known as a binomial process. Letting m_{ig}^k denote the number of EAs in area i, stratum k, and pixel g, and letting $m_{i:}^k$ denote the vector of the number of EAs in area i and stratum k for all the pixels in that area and stratum, our assumptions then dictate that $\boldsymbol{m}_{i:}^k \sim \text{Mult}_{|G_i^k|}(|B_i^k|, \boldsymbol{\pi}_{i:}^k)$ for $\pi_{ig}^k = \tilde{q}_{A_i^k}(s_{ig})$, and for G_i^k representing the indices of the pixels

in area i stratum k so that $|G_i^k|$ is the number of pixels in area i and stratum k. Hence, the $|B_i^k|$ EAs are distributed among the $|G_i^k|$ pixels in area i and stratum k with probability proportional to the general population density in each pixel. Since there are 96,215 EAs in Kenya and 273 (possibly combined) Admin2 areas, on average there are about 352 EAs per Admin2 area. The exact number of EAs per Admin2 area will vary based on variation in the multinomial distribution and also depending on the number of EAs in each Admin1 area, and the Admin2 area populations.

In practice, there may be some multinomial samples such that there are zero EAs in a constituency (an Admin2 area in Kenya) crossed with a stratum in spite of there being people in the constituency crossed with the stratum. Since this is unrealistic, and also causes problems when predicting prevalence in a constituency, we constrain the multinomial to have at least one EA per constituency, rejecting any samples for which the constraint is not satisfied.

Since there are many more unsampled EAs than sampled ones (1,612 clusters and 96,251 EAs in total), we will assume for simplicity the cluster effects at sampled cluster locations have the same posterior distribution as the cluster effects at the unsampled cluster locations. However, it is not too difficult to relax that assumption if necessary.

Since many pixels may have very low general population density, it may be rare to draw superpopulation samples with any EAs or population in those pixels. Prevalence is only defined for a pixel if it contains any population, so when a posterior distribution of prevalence within one such pixel is desired, it may be more efficient to draw from the marginal distribution for that pixel, conditioning on there being at least one EA in that pixel. In that case, $m_{ig} \sim \text{Binomial}(|B_i^k|, \boldsymbol{\pi}_{i:}^k)$, but is then conditioned on having at least 1 EA.

Including cluster effects yields the following aggregation equations:

$$p_i \approx \sum_k Q_i^k \frac{1}{|B_i^k|} \sum_{c \in B_i^k} \mu_{ic} \tag{4.11}$$

$$p_{ig} \approx \frac{1}{|B_i^g|} \sum_{c \in B_i^g} \mu_{ic}, \tag{4.12}$$

where $\mu_{ic} = \text{expit}(\eta_{ic})$ includes the cluster effect for cluster c in area i and where $Q_i^k \equiv \int_{A_i^k} \tilde{q}_{A_i}(s)\, d\boldsymbol{s}$ is the proportion of neonatals in stratum k out of all neonatals in area i. We will refer to this as the SC aggregation model.

4.2.6 Population Denominator and Numerator Variation: SCP Aggregation and Superpopulation Model

Although we average over the probabilities $\mu_{i:}$ within each area and pixel with equal weight, in truth, each EA will have a different population denominator, and so the weights for each cluster in the average should instead be proportional to the population denominators at each EA c in an area i, $\boldsymbol{N}_{i:}$. Since the population denominators are not known, we can add in a model for the population denominators based on the census data, $\mathcal{W}$. We assume a Poisson-Multinomial model for the number of households per EA and neonatals per household in a stratum, assuming each EA has at least 25 households, and, conditional on the total number of households and neonatals per stratum, distribute the rest of the households with equal probability for each EA and the children with equal probability for each household.

At this point, since we have already assumed a model for $Y_{ic}|N_{ic}, \mu_{ic}$ in (4.1), we have built a superpopulation model that we can sample entire populations from. For any finite population sampled from the superpopulation model, we then use,

$$p_i = \sum_{c \in B_i} \frac{N_{ic}}{N_i} \times \frac{Y_{ic}}{N_{ic}} \tag{4.13}$$

$$p_{ig} = \sum_{c \in B_i^g} \frac{N_{ic}}{N_{ig}} \times \frac{Y_{ic}}{N_{ic}}, \tag{4.14}$$

as the SCP aggregation equations to obtain the prevalence at the areal and pixel levels. Details on how to draw from the joint distribution of the SCP CPAM are given in Algorithm 1 in Appendix B.3. It is important to note that the SCP aggregation model is a model for prevalence rather than risk, unlike the other aggregation models.

We summarize the three models in Table 4.2, and derivations for the expectations are given in Appendix B.4.

Aggregation Model	**Aggregation Equation**	
SCP	p_i	$= \sum_{c \in B_i} \frac{N_{ic}}{N_i} \times \frac{Z_{ic}}{N_{ic}}$
	p_{ig}	$= \sum_{c \in B_i^g} \frac{N_{ic}}{N_{ig}} \times \frac{Y_{ic}}{N_{ic}}$
SC	$p_i \approx E_{\boldsymbol{Y}_{i:}, \boldsymbol{N}_{i:}}[p_i]$	$= \sum_k Q_i^k \frac{1}{\lvert B_i^k \rvert} \sum_{c \in C_i^k} \mu_{ic}$
	$p_{ig} \approx E_{\boldsymbol{Y}_{i:}, \boldsymbol{N}_{i:}}[p_{ig}]$	$= \frac{1}{\lvert B_i^g \rvert} \sum_{c \in B_i^g} \mu_{ic}$
S	$p_i \approx E_{\boldsymbol{Y}_{i:}, \boldsymbol{N}_{i:}, \boldsymbol{\epsilon}_{i:}, \boldsymbol{s}_{i:}}[p_i]$	$= \int_{A_i} \tilde{\mu}(\boldsymbol{s}) \times q_{A_i}(\boldsymbol{s}) \, \mathrm{d}\boldsymbol{s}$
	$p_{ig} \approx E_{\boldsymbol{Y}_{i:}, \boldsymbol{N}_{i:}, \boldsymbol{\epsilon}_{i:}, \boldsymbol{s}_{i:}}[p_{ig}]$	$= \tilde{\mu}(\boldsymbol{s}_{ig})$

Table 4.2: Areal and pixel level aggregation equations under three considered combined population aggregation models, in order of decreasing predictive variance. Subscripts in the expectations denote the variables in the superopulation model over which expectations are taken whenSix calculating risk.

4.3 Assessing CPAM Models with Simulation Study

4.3.1 Simulation Study Description

In order to evaluate the constituency level predictive performance of all three CPAMs (S, SC, and SCP), we conduct a simulation study involving several different populations simulated under the SCP aggregation model using the SPDE_{UC} model from Chapter 3 for the cluster level risk with parameters set as described below. Table 4.3 gives the 8 scenarios considered with the following set of parameters used to simulate the underlying population.

Scenario number	**Cluster level risk**	**Survey design**
1	$\eta_{ic} = -2.9 - 1 \cdot I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Stratified
2	$\eta_{ic} = -2.9 - 1 \cdot I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Representative
3	$\eta_{ic} = -3.9 + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Stratified
4	$\eta_{ic} = -3.9 + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Representative
5	$\eta_{ic} = -1 \cdot I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Stratified
6	$\eta_{ic} = -1 \cdot I(\boldsymbol{s}_{ic} \in U) + u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Representative
7	$\eta_{ic} = u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Stratified
8	$\eta_{ic} = u(\boldsymbol{s}_{ic}) + \epsilon_{ic}$	Representative

Table 4.3: The scenario numbers, linear predictors and sampling schemes of the 8 considered simulation study scenarios.

Under each of the simulation study scenarios, 10 different risk surfaces and populations are independently simulated, including EA locations, the number of households, and the number of neonatals in the time period. For each simulated population, a single survey is collected that is either 'Representative,' meaning the proportion of population surveyed in a county that is urban is on average correct, or is 'Stratified' and uses the equivalent sampling

design to the 2014 KDHS for the simulated population. These sampling designs are identical to those used in the simulation study in Chapter 3. For each simulated population and associated survey, we fit the SPDE_{UC} cluster level risk model, and use the posterior from the SPDE_{UC} cluster level model with the S, SC, and SCP aggregation models to produce aggregate predictions. All scenarios use effective spatial range $\rho = 150$km, spatial variance $\sigma_S^2 = 1/9$, and cluster effect variance $\sigma_\epsilon^2 = 0.4$. The third scenario has parameters most similar to the model fit to NMR data although with shorter spatial range. However, we consider the first scenario to have the most practical significance since it includes an urban effect and is therefore more illustrative of the effect of stratification.

For each population and survey, we generate constituency level predictions of neonatal mortality prevalence, total deaths, and the relative prevalence (the urban prevalence divided by the rural prevalence). Predictive distributions are compared to their true, simulated population values in terms of the bias, RMSE, CRPS, 80% coverage, and 80% credible interval width, with scoring rules being averaged over all 10 surveys and populations for a given simulation scenario.

For the SPDE_{UC} risk model we use a joint PC prior (Fuglstad et al., 2019b) on the effective range and spatial variance parameters satisfying $P(\rho > 289\text{km}) = 0.5$ and $P(\sigma_S > 1) = 0.5$. For the cluster effect variance we also apply a PC prior satisfying $P(\sigma_\epsilon > 1) = 0.5$. For the intercept and urban effect we set improper Gaussian priors with zero mean and infinite variance.

4.3.2 Simulation Study Results

This section will focus on interpretation of the scoring rules calculated from the simulation study based on the surveys simulated with $\beta_0 = -2.9$, $\beta^{\text{URB}} = -1$ under the stratified sampling scheme are given in Table 4.4 for prevalence, total deaths, and relative prevalence, all at the constituency level. All simulation study scoring rule tables for the other sampling

schemes and population parameters are given in Tables B.1–B.7 in Appendix B.1.

We find that the S aggregation model exhibits undercoverage for all quantities of interest: prevalence, total deaths, and relative prevalence. Coverage for the S aggregation model was especially poor for relative prevalence, with only 35% coverage for 80% nominal coverage credible intervals. The SC aggregation model shows a marked improvement in coverage for all quantities of interest compared to the S aggregation model, and the SCP aggregation model further improves the coverage, although only by 2% in the case of prevalence and total deaths. Even using the SCP aggregation model, the coverage of prevalence, total deaths, and relative prevalence were 71%, 73%, and 71% respectively.

Difficulties in identifying the cluster level risk model under spatial correlation likely contributes to under coverage in all of the aggregation models. Interestingly, results in Appendix B.1 suggest that for the populations with $\beta_0 = 0$ the predictions were less biased and had better coverage, possibly since the outcome is less rare so there is more information in each survey, whereas for the populations with $\beta_0 = -2.9$ and $\beta_0 = -3.9$ the predictions were slightly positively biased, likely leading in part to the undercoverage. The S aggregation model had predictions that were especially biased upwards for the $\beta_0 = -2.9$ and $\beta_0 = -3.9$ populations in part due to the logistic approximation used when integrating out cluster effects causing bias as predictions move away from 50% prevalence.

The sampling scheme did not appear to affect any of the scoring rules substantially in a consistent way, which indicates that the considered CPAMs are able to account for stratification well.

Overall, the SCP aggregation model consistently performed the best on average for all scoring rules and all populations and sampling schemes considered. The S aggregation model, however, performed the worst, and often by a substantial margin. The SC aggregation model often performed comparably to the SCP aggregation model, especially when predicting prevalence and total deaths, although the gap between the SC and SCP aggregation model was

largest when predicting the relative prevalence in urban versus rural parts of constituencies.

Since results from Chapter 3 suggest the S aggregation model performs well at the county level, the smaller populations and number of EAs must be leading to the poor performance of the S aggregation model at the constituency level. It is at the constituency (Admin2) level where the difference between the S and SCP aggregation models truly becomes noticeable, and especially in terms of the ability to account for uncertainty due to cluster level and population numerator and denominator variation.

These results also suggest that, while spatial models often are able to reduce uncertainty in prevalence predictions for small areas relative to survey weighted direct estimates, some of these shrunken uncertainty intervals may be smaller at the Admin2 level in large part due to their approximation of prevalence with risk, and their inability to account for cluster level and population numerator and denominator variation. Since aggregate predictions are difficult to directly validate from cluster level data, these problems have not yet been identified or adequately addressed.

4.4 Application to 2014 KDHS Neonatal Mortality

We apply the proposed CPAMs assuming a SPDE_{UC} model with urban and cluster level random effects to the 2014 KDHS NMR data, and using the same priors as in the simulation study. We predict neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) for the time period 2010–2014. Validation was performed in Table 3.4 at the cluster level, stratified by urban/rural. Additional results are given in Appendix B.2.

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	0.005	0.010	0.006	61	0.020
SC	0.005	0.010	0.005	69	0.022
SCP	0.005	0.010	0.005	71	0.022
Total Deaths					
S	97	247	128	55	350
SC	85	197	100	71	403
SCP	85	197	100	73	412
Relative Prevalence					
S	0.051	0.105	0.064	35	0.099
SC	0.036	0.098	0.053	57	0.155
SCP	0.038	0.098	0.051	71	0.209

Table 4.4: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = -2.9$, $\beta^{\text{URB}} = -1$ under the stratified sampling scheme.

4.4.1 Mapping Predictions

We predict neonatal mortality rates in Kenya from 2010-2014, both in terms of mortality risk and prevalence, using a SPDE_{UC} model for the cluster level model, and using S, SC, and SCP aggregation models. Predictions for the S aggregation model are generated without using the logistic approximation in order to eliminate the bias slight but consistent bias it induces in the S aggregation model predictions. Spatial central predictions for the SCP aggregation model along with 80% credible interval widths are given in Figures 4.1 and 4.2 respectively, each figure giving results at all considered spatial aggregation levels: 5km $\times$ 5km pixel, constituency, county, and province. Equivalent plots for total deaths and relative prevalence are shown in Appendix B.2. Pixel level central predictions are given from the S aggregation model, since all aggregation models agree in expectation by the tower property of expectation, but pixel level uncertainties for the SCP aggregation model are estimated via the sample variance from the 1000 posterior samples. Some pixels have a very low probability of having any EAs in them, and pixels that only had defined prevalences (or defined total deaths and relative prevalences for those figures) in at most 1 posterior sample were left out in the pixel level CI width plots. Similarly, if there were not at least 1 neonatal in the target population denominator per posterior sample on average, results were not shown for that pixel, and if a pixel was so unlikely to have an EA that the upper end of the 80% credible interval for total deaths was identically 0, the results in those pixels were not plotted. This resulted in 9.7% of the pixels being left out for results using prevalence CI widths and 38.6% of the pixel CI widths not being included for total deaths.

The discrepancy between the number of pixel level results being left off for prevalence and total deaths is due to the fact that prevalence is not defined when there is no population present in a pixel, whereas total deaths is. Hence, the posterior for prevalence within a pixel does not depend on posterior samples where that pixel contains no EAs, whereas the posterior for total deaths in the pixel level does. The upper limit of CI for total deaths is

therefore 0 when the pixel contains EAs in fewer than $100 \times (\alpha/2)$ percent of the posterior samples for $100 \times (1 - \alpha)$ percent CIs. Since we are using 80% CIs, pixels containing EAs in fewer than 10% of the posterior samples will have posterior 80% CI widths equal to 0 for their total deaths. For 95% CIs, pixels would have nonzero CI widths for total deaths if they contained EAs in at least 2.5% of the posterior samples. On average, there are about 54 neonatals per EA, so requiring at least 10% of posterior samples to contain EAs corresponds to requiring roughly 5.4 neonatals in the pixel per posterior sample on average, or 0.22 neonatals per km^2 per posterior sample (the exact expected number of neonatals per EA depends on the stratum, so these figures are approximate). Although leaving out 38.6% of the pixels seems at first like a large number, giving predictions for pixels with such low populations is somewhat problematic, and leaving them out avoids this potential issue.

While the highest prevalences are predicted in the center and southern portions of eastern Kenya, the areas with the highest predicted total deaths are in the central and southwestern parts of Kenya. Credible interval widths for prevalences tend to be lowest in the central and western parts of Kenya, which tends to be more densely populated and urban, whereas the uncertainties tend to be higher in the eastern and far northern parts of Kenya that tend to be more rural and sparsely populated. However, credible interval widths for total neonatal deaths are largest in areas with both high population and a low number of clusters surveyed. In general, uncertainties in total deaths are highest in central Kenya. Predicted prevalences tend to be slightly higher in urban areas, although the urban effect is not statistically significant.

We show the percent increase in 80% credible interval widths for each aggregation model relative to the S aggregation model at all considered areal aggregation levels in Figures 4.3–4.5, where the results for prevalence/risk, total deaths/expected total deaths, and relative prevalence/risk are shown in each figure respectively. For both prevalence and total deaths, the SCP aggregation model produced larger credible interval widths than the S aggregation

model at pixel and constituency levels (excluding pixels with 80% CI widths of 0 for the total death predictions due to containing 0 EAs at least 90% of the time), while producing similar sized credible interval widths at the county and province levels.

The relative prevalence credible interval widths tended to be much larger for the SC aggregation model than the S aggregation model, and the SCP aggregation model CI widths were in turn considerably larger than those of the SC aggregation model. We found this to be true at all aggregation levels where the relative prevalence was defined. This excludes the pixel level, since we have classified each pixel as either entirely urban or entirely rural.

The computation time of generating the S aggregation model predictions, including fitting the cluster level data model using INLA (which took 2.7 minutes), surprisingly took the most time (43.0 minutes), although this is due to the fact that numerically integrating out the cluster effect at each pixel grid cell took 39.3 minutes without using the logistic approximation. The numerical integrals were calculated using adaptive quadrature via the `integrate` function in `R` (Piessens et al., 2012). With the logistic approximation, the integration only took 4.6 minutes. The SC model took 8.9 minutes, and the SCP model took 11.1 minutes. Approximately 3.2 minutes of the SCP computation time were spent sampling EA population denominators, and another 2.8 minutes were taken generating pixel level predictions compared to 1.0 minutes making areal predictions at county and constituency levels. Most of the SC model computation time was spent generating pixel level predictions (4.3 minutes).

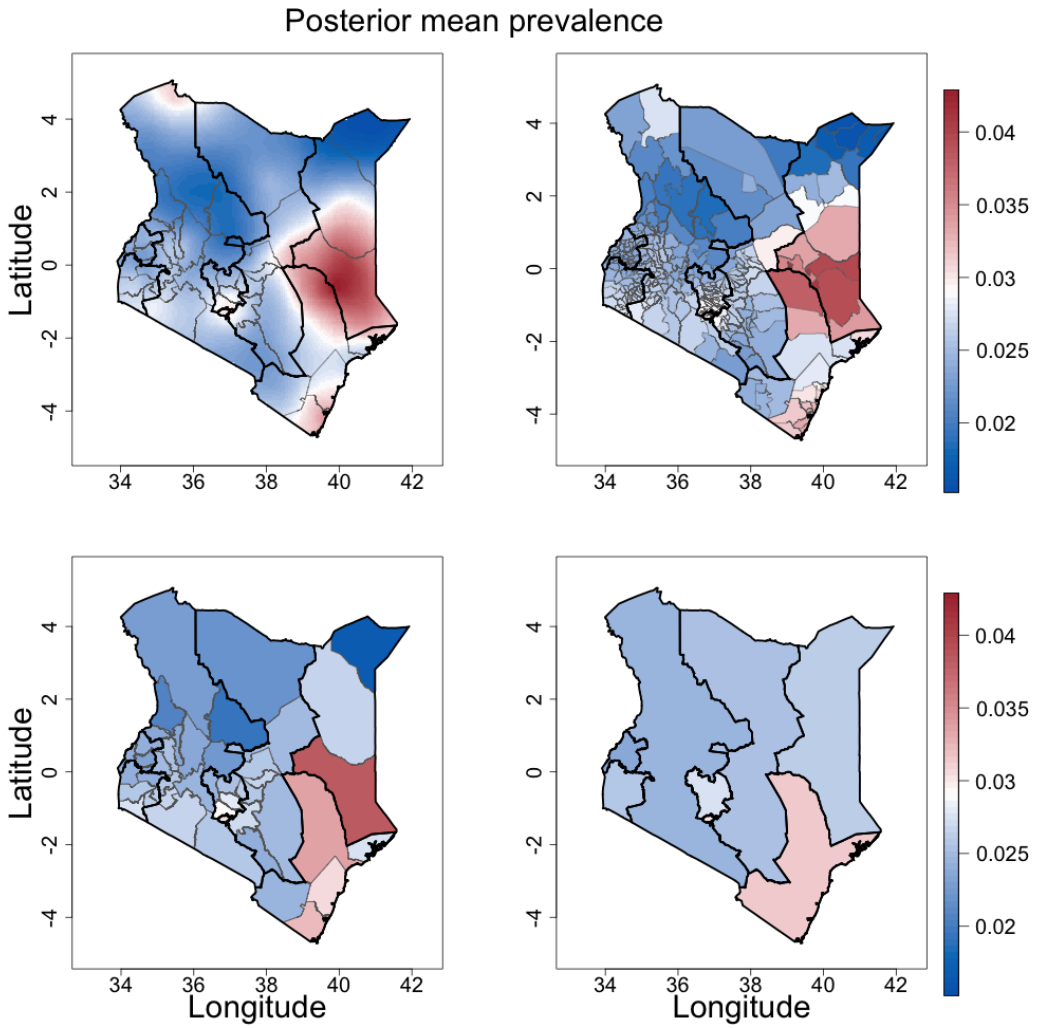


Figure 4.1: Predicted neonatal mortality prevalence for the SCP aggregation model at the 5km × 5km pixel (top left), constituency (top right), county (bottom left), and province (bottom right) levels. Province and county borders are shown as black and grey lines respectively.

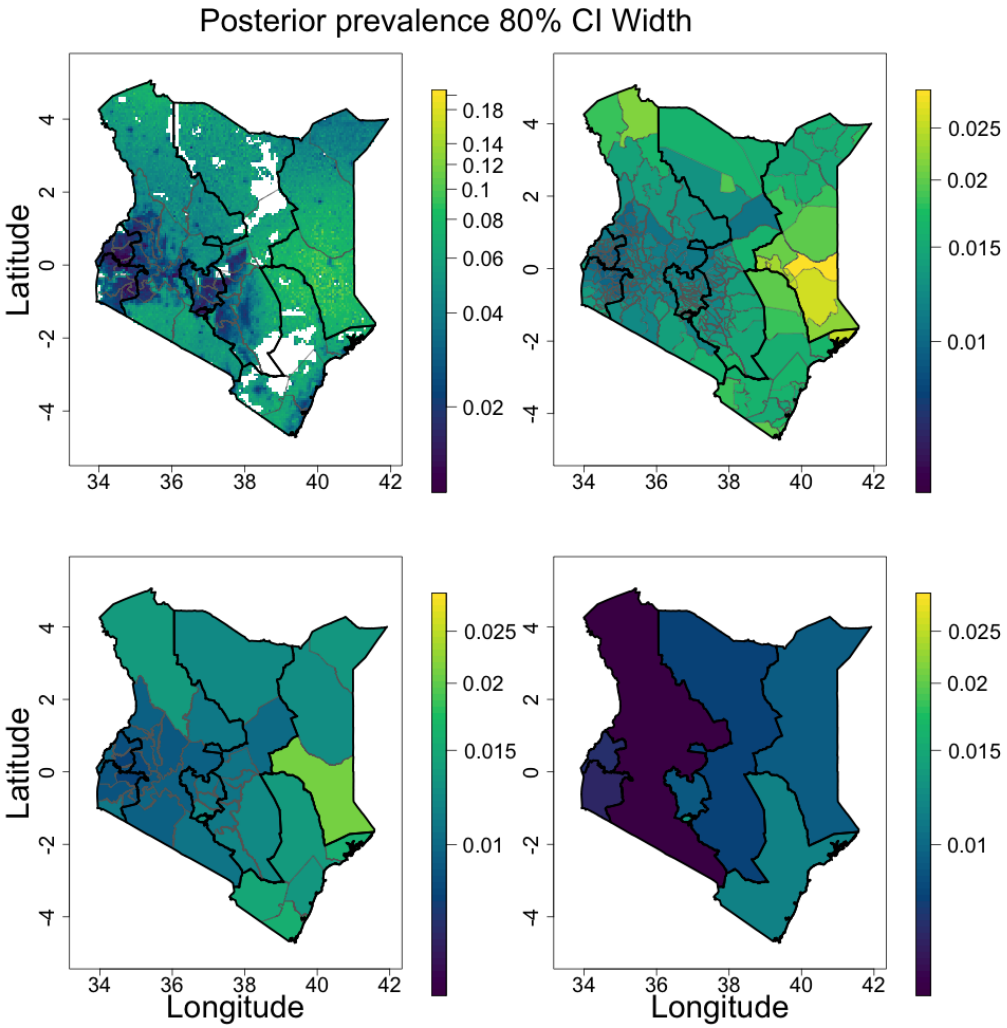


Figure 4.2: Predicted neonatal mortality prevalence 80% credible interval width for the SCP aggregation model at the 5km × 5km pixel (top left), constituency (top right), county (bottom left), and province (bottom right) levels. Province and county borders are shown as black and grey lines respectively.

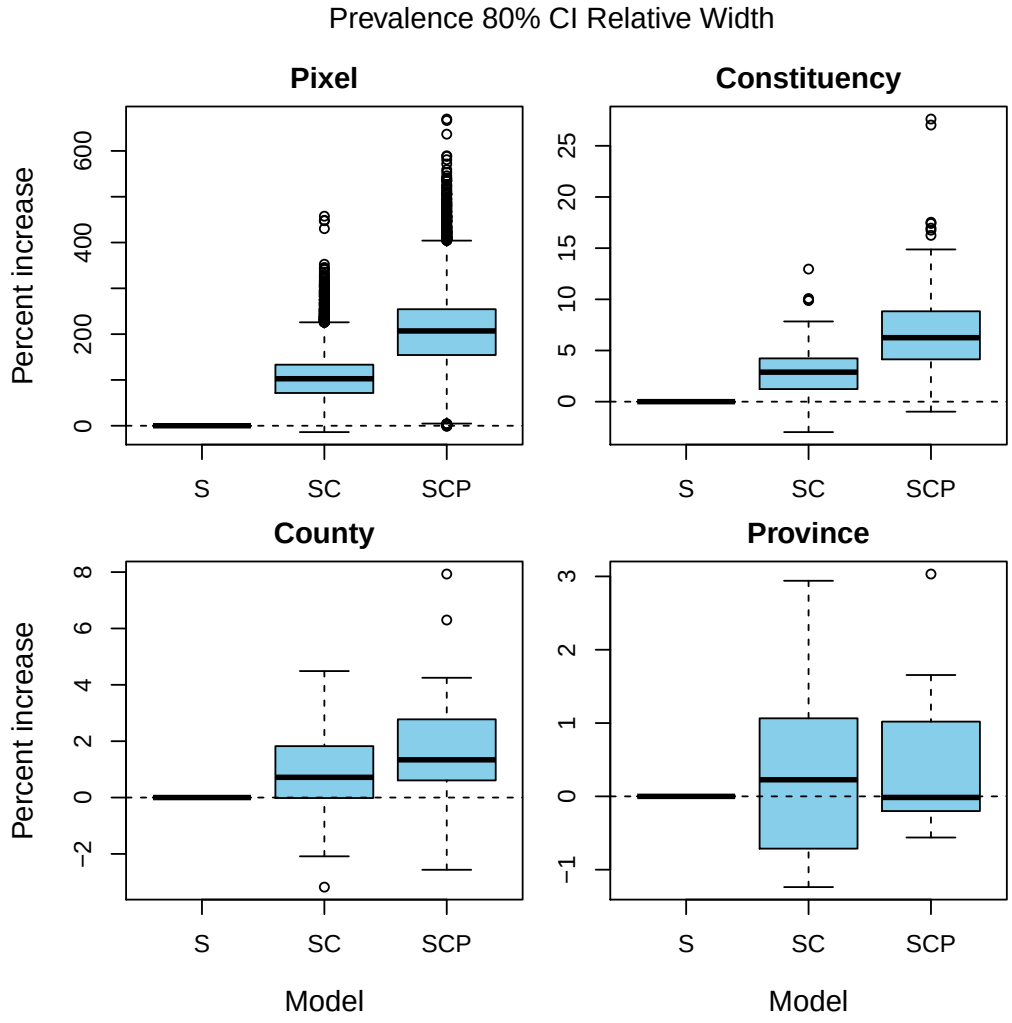


Figure 4.3: Neonatal mortality prevalence 80% credible interval width percent increase relative to the S aggregation model at the 5km × 5km pixel (top left), constituency (top right), county (bottom left), and province (bottom right) levels.

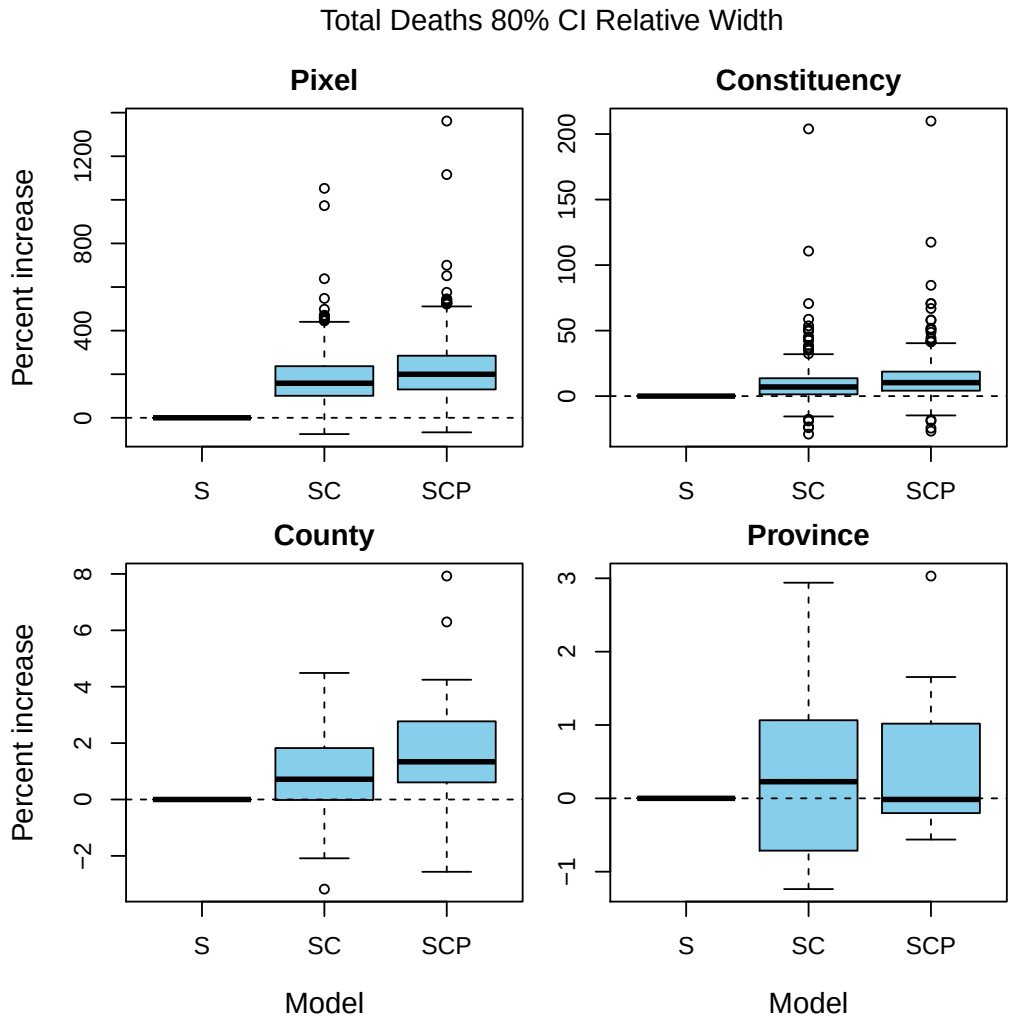


Figure 4.4: Total neonatal deaths 80% credible interval width percent increase relative to the S aggregation model at the 5km × 5km pixel (top left), constituency (top right), county (bottom left), and province (bottom right) levels.

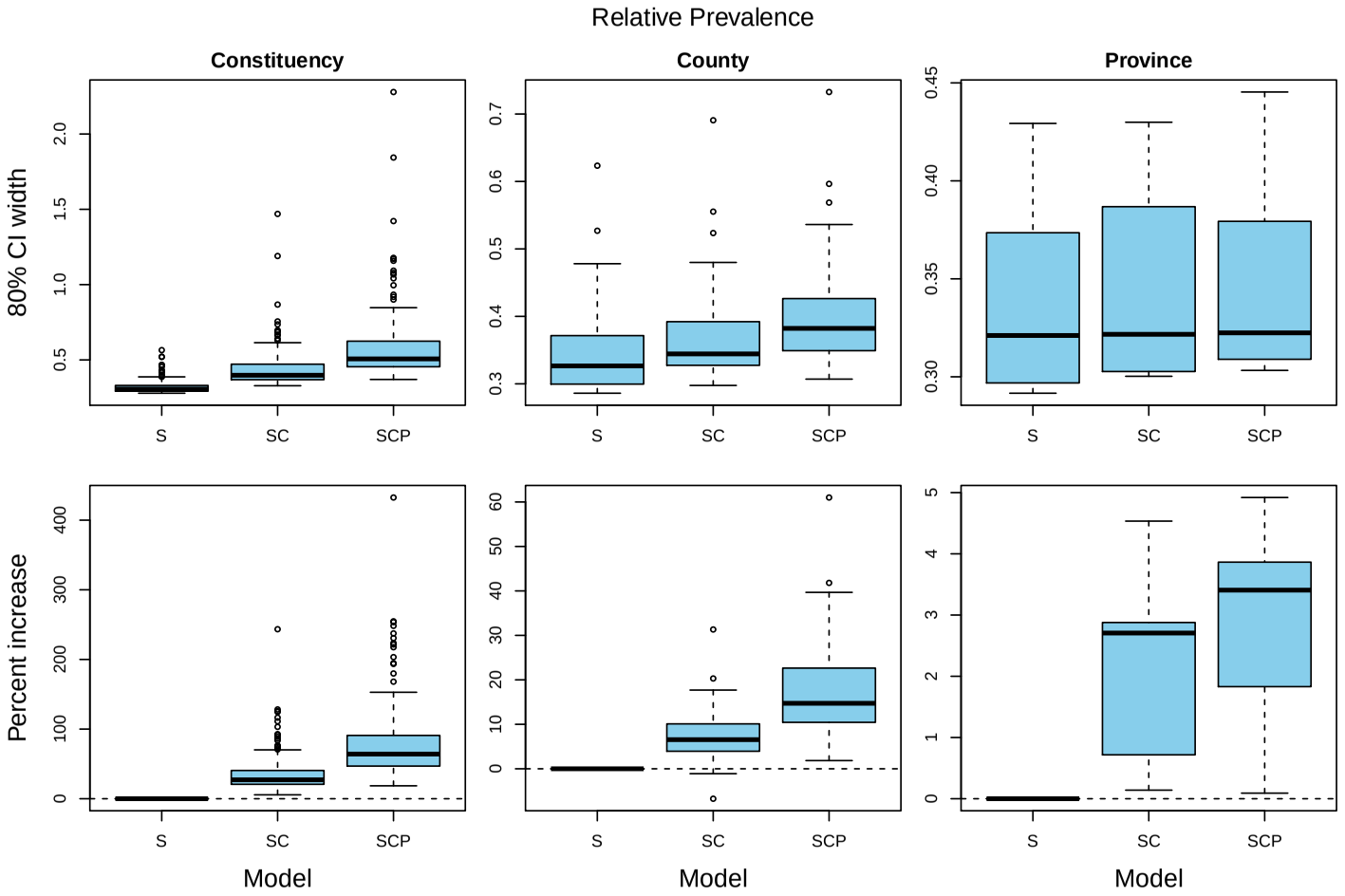


Figure 4.5: Neonatal mortality relative prevalence (urban relative to rural) 80% credible interval width (top row) and percent increase compared to the S aggregation model (bottom row) at the constituency (left column), county (middle column), and province (right column) levels.

4.5 Conclusions

In this chapter, we propose the first population prevalence model compatible with continuous spatial models indexed at the cluster level when EA locations are unknown. Rather than leaving cluster level random effects out of predictions of population aggregates and leaving out population numerator and denominator variation as is typical when information regarding EA locations are unknown, our proposed CPAM framework leverages census and population density information in order to incorporate these sources of uncertainty into the predictions. The proposed CPAM framework is able to account for spatial correlation, survey stratification, cluster level random effects, EA location uncertainty, and population numerator and denominator variation. Since both census and population density information are used separately from any data that is used to fit the spatial risk model before aggregation, no additional parameters are required aside from those fit when conducting inference in the standard way for the spatial risk model used in the CPAMs.

In our simulation studies, we consider the case where cluster variance is about 3 times larger than the smooth spatial variance, since this is the case for the KDHS 2014 NMR data. In this setting, accounting for cluster level variance, EA location uncertainty, and population numerator and denominator variation has a substantial effect on the constituency level posterior uncertainties for prevalence, total count, and relative prevalence between urban and rural areas. Target population denominator variation appears less important, conditional on the number of EAs per area $\times$ stratum, especially for estimates of prevalence. Differences between the CPAMs are significant at the constituency level when estimating prevalence, but are even larger when estimating total counts, and largest when estimating the relative prevalence of urban to rural portions of constituencies. This is likely because, while prevalences are normalized by the population totals, restricting them to $[0, 1]$, count totals are not normalized in this way, allowing for more variability. Relative prevalence depends on prevalences in both urban and rural parts of an area, each of those parts having a smaller

population total and smaller number of EAs than the constituency itself. Differences between prevalence and risk will therefore be more pronounced at urban and rural parts of a constituency then at the constituency itself, making the relative prevalence, a ratio of urban and rural prevalences within an area, substantially more variable than the relative risk. We therefore believe that relative risk estimates at the constituency level and finer have considerably less uncertainty than relative prevalence estimates, and using them interchangeably is problematic if more than just central estimates are required.

Although we focus on constituency level estimates, our results also have implications on predictions smaller than the constituency level, such as for urban or rural parts of constituencies or 5km $\times$ 5km pixels. For instance, since population denominators at this scale are much smaller than the constituency level, it is especially important to keep in mind the differences between pixel level risk and prevalence. In fact, both S and SC aggregation models are models for risk, and also see increasingly large differences at small aggregation levels. These differences increase due to the smaller number of EAs and people with which to average out unstructured random effects. Hence, even if one is interested in pixel level risk predictions rather than prevalence, it is still important to carefully consider what type of 'risk' to use, and to account for cluster level variation, EA location uncertainty, and possibly population denominator variation as well, when relevant.

An advantage of the CPAM framework is its generality. As long as a spatial model for health and demographic indicator risk is indexed at the cluster level, it is possible to apply the CPAM framework directly to posterior draws from the model. The framework actually generalizes the models used in Paige et al. (2020) and Dong and Wakefield (2020) and uses a similar cluster level data model to commonly used models by WorldPop, IHME, and others, building a framework enabling these models to account for cluster level variation, EA location uncertainty, and population numerator and denominator variation. The CPAM framework is also general in that the aggregation equations for each of the 3 CPAMs considered could

be applied to the continuous spatial model used in Chapter 5, and to the BYM2 model from Chapter 3. In fact, the aggregation equations simplify dramatically for the BYM2 model since the predictions are constant for each area $\times$ urbanicity combination. Distinguishing between risk and prevalence and incorporating cluster level variation would be even more important for the BYM2 model, since the cluster effect variance is typically larger than that of the SPDE model due to continuous spatial variation in risk within an area $\times$ urbanicity combination. The relationship between the different considered CPAMs, their aggregation equations, and some existing models in the literature is summarized in Table 4.2,

The CPAM framework uses a superpopulation sampling model to help build a mathematical consistency for understanding models for population aggregates when EA locations are uncertain. For instance, we showed that some existing models are obtained by taking expectations over different portions of the superpopulation sampling model, showing that those models do not account for uncertainty in the portions of the superpopulation model expected over. Additionally, we find that those models have the same central predictions as the proposed SC and SCP model. Although the SC and SCP models make seemingly strong assumptions about the distribution of the EA locations and population denominators in those EAs, the fact that they agree in expectation with accepted models lends additional credence to the proposed SC and SCP aggregation models.

Partly due to the proposed CPAMs not requiring any additional parameters and therefore not adding to the model fitting time, and also due to our implementation taking advantage of the computational speed of matrix operations enabled by assuming EA locations lie only at the centroids of each pixel, the proposed SC and SCP CPAMs computational performance are comparable to existing methods. The SC and SCP aggregation models were actually much faster than the S aggregation model when the S aggregation model did not use the logistic approximation in (4.8), with the SCP model taking about a quarter of the time and the SC model taking about a fifth of the time. However, when the logistic approximation was used

to speed up the process of integrating out the cluster effect at each pixel, the S aggregation model computation time was sped up dramatically, reducing the cluster effect integration time from 39.3 minutes (out of 42.0 minutes of S aggregation model total compuation time) to only 4.6 minutes, making the S aggregation model faster than the SCP aggregation model and approximately the same speed as the SC aggregation model, although inducing some bias in the predictions. If SCP model predictions are not affected substantially by population denominator variation, then computation times could be reduced even further by assuming population denominators are evenly spread among EAs within a stratum or using only a single sample of population denominators from the superpopulation model rather than using the multistage multinomial sampling approach implemented in this study. Computation times can be further reduced if prevalence predictions at the pixel level are not required.

The CPAM approach has several limitations. Any CPAM is limited by the accuracy of the cluster level risk model and the aggregation equations. If the cluster level risk model does not account for the survey design, then neither do the predictions of the resulting CPAM. Our example CPAMs can account for stratification and cluster level variation, but do not directly account for design weights and therefore do not directly account for other aspects of the design such as PPS sampling of the EAs in the survey used to fit the risk model. The aggregation equations we provided rely on accurate estimates of population density and population totals at the Admin1 level crossed with urbanicity. If the population density and population totals are not accurate, the performance of the aggregation equations will suffer accordingly. In spite of this limitation, the CPAM framework is general enough to be compatible with risk models that fix or mitigate these issues.

The misclassification of the spatial stratification variable, official urbanicity designations in the case of standard DHS surveys, as a function of spatial pixel may lead to inaccurate predictions in pixels that are either misclassified, or that have a mixture of urban and rural EAs in them. A potential solution to this is to, rather than modeling urbanicity as a binary

classification of a pixel, assume that the probability of any EA in a pixel as being urban or rural can be calculated from a simple logistic model that is linear in population density. This would likely double the numerical integration time, although pixel resolutions could be increased to offset this cost. Since model predictions at such fine scales are sometimes considered problematic anyway, increasing pixel resolution somewhat could be an appealing solution when combined with the ability to model probabilities of urbanicity in each pixel.

More broadly, the proposed CPAMs assume each EA can be represented as a single point location, and that every person in that EA live at that point. There are 96,251 EAs in Kenya, and only 23,440 spatial pixels, so the EAs tend to be very small in area, making any variation in the smooth spatial field over that area negligible as well. This means that the assumption that each EA is a point likely does not affect the predictions of risk and prevalence for any given EA. However, one would like a pixel level estimate of prevalence to be the prevalence over all people living in that pixel, and if an EA's area and population spans multiple pixels, the prediction of all of the proposed CPAMs except for the S aggregation model would instead represent the prevalence for the pixel as the prevalence over all people living in EAs whose centroids lie in that pixel. Although it may be reasonable to assume that each EA's boundaries are enclosed within a single constituency, this will not necessarily be the case for predictions for customized areas such as pixels, especially if the custom areas are small enough to the point where the difference of a few EAs in it will significantly impact predictions. It may be possible to fix this problem via additional smoothing of EA level predictions, although this could add an extra layer of complication to the model. Instead, it may be preferable to simply emphasize that predictions for sufficiently small areas with custom boundaries should be interpreted as aggregations over EAs with centroids in those areas.

The proposed CPAM framework for small area prevalence estimation may be useful in model validation, since it enables predictions with variability that is on the same scale as the

response. This is especially important because of the difficulty and inherent problems with validating areal predictions when only point level data are available. Although we validate our areal predictions using direct estimates, there are other potential avenues for validation. For instance, one could try to predict the sample prevalence among left out observations, calculating scoring rules against the left out sample prevalence. It would also be possible to validate at the pixel level, trying to predict the sample prevalences of left out pixels. While these methods may not directly inform population level predictions, it could be helpful for model assessment nonetheless.

One potential avenue of future research would be to extend the proposed CPAMs to include temporal components. However, since neonatal deaths are discrete events, continuous time estimates for the SCP aggregation model would be made more difficult by the need to develop a superpopulation model in space-time. In addition, since population denominators depend on the period of time considered, with short time periods sometimes having very small population denominators, it may make little sense to consider predictions for the SCP aggregation model for time periods that are sufficiently short, such as those shorter than one year for neonatal mortality, or for instantaneous time predictions. For sufficiently short time periods and for instantaneous time predictions it therefore may make more sense to use risk models such as the S and SC aggregation models could be used in the continuous time case, although all proposed CPAMs could be used if risk and prevalence estimates are generated for a time interval. Additionally, it will be necessary to aggregate over several census frames if the time domain is long enough. Aggregation over several census frames could be possible via generating separate estimates for each frame, and taking a weighted average with weights proportional to the posterior precisions (Geir-Arne Fuglstad and Wakefield, 2020).

It will also be important to be able to incorporate other relevant covariates into the model. To do so it will be important to either have explicit distributions of the value of that covariate for the population of interest, or for the covariate to not confound with the cluster

effects included in the model. For instance, smooth spatial covariates could be included very simply, but individual level covariates would require information on the distribution the variable takes among individuals in the target population, preferably within each stratum if there is sufficient data.

Chapter 5

EXTENDING LATTICEKRIG: BAYESIAN MULTIRESOLUTION-KRIGING USING GAUSSIAN MARKOV RANDOM FIELDS FOR GENERAL OBSERVATION MODELS

5.1 Introduction

The increasing size and complexity of spatial point datasets in fields such as climate sciences, public health, ecology, and social sciences have been concurrent with methodological developments in spatial statistics. While there are currently a host of methods available for handling inference with "big data" using traditional spatial models (Heaton et al., 2019), there has been less focus on accessible tools for more complex spatial dependence structures. In the context of multi-resolution spatial modeling, recent developments are the LatticeKrig (LK) model (Nychka et al., 2015) with the associated `R` package `LatticeKrig` (Nychka et al., 2016), and the multi-resolution approximation (M-RA) model (Katzfuss, 2017) with its implementation in the `R` package `GPvecchia` (Katzfuss and Guinness, 2020; Katzfuss et al., 2018; Zilber and Katzfuss, 2019). However, to the best of our knowledge there exist no Bayesian implementations of LK or M-RA allowing for non-Gaussian responses; `LatticeKrig` is limited to Gaussian responses as well, and `GPvecchia` allows general exponential families for the responses.

The most common approach to spatial modeling is to use parametric classes of spatial covariance functions with interpretable parameters such as the Matérn family. Depending on its smoothness parameter ν, the Matérn covariance class includes both exponential and Gaussian covariance functions. However, in practice, the smoothness parameter is commonly

fixed at a small number, in part due to the difficulty in estimating this parameter, and the computational benefit of having one fewer parameter (Stein, 1999). It is known that, under infill asymptotics, it is the behavior of the Matérn covariance function at short spatial scales that most determines the likelihood and pointwise predictions (Stein, 1999, Ch. 3). This means that while short scale behavior of the Matérm covariance may be fit accurately, long scale correlations in the data will often not be accurately reproduced by the fit model. However, as we later show in the simulation study, long range correlations become increasingly important when making predictions far from observations. Additionally, we show in Appendix C.1 that for areal predictions, errors in the covariances at spatial scales close to the average radius of the areas affect the uncertainty of those areal predictions the most, suggesting that long scale correlations are especially relevant when calculating the uncertainty of areal averages for large areas.

The difficulty in identifying spatial model parameters makes it especially important to integrate over uncertainty when calculating predictive uncertainty. In a frequentist setting the bootstrap can be applied, but it relies on asymptotics and is computationally expensive since it requires the model to be refit many times (Sjöstedt-de Luna and Young, 2003). Handcock and Stein (1993) and Gelfand et al. (2010, Ch. 3.7) recommend using Bayesian inference in spatial statistics due to the importance of accounting for uncertain covariance structure. However, Markov Chain Monte Carlo (MCMC) techniques are often difficult to implement with long running times and large memory requirements, especially with large numbers of observations (Filippone et al., 2013). Detailed output diagnostics are also necessary to assess convergence.

As such, the key limitation in providing Bayesian inference for multiresolution spatial models is the computational complexity involved. In this Chapter, we propose to take advantage of the deterministic algorithm for Bayesian inference based on INLA (Rue et al., 2009b). LK uses different layers of compact basis functions together with an associated

sparse precision matrix, and fits directly into the INLA framework of latent Gaussian models. We provide an implementation using `INLA`, which permits fast and accurate estimation of posterior marginal densities provided that the number of parameters is not too big (typically 2 to 5, but not exceeding 20 (Rue et al., 2017)). This extended version of LK is termed extended LatticeKrig (ELK). A key change from the original LK formulation is a reparametrization that improves interpretability and facilitates modeling and prior selection. Furthermore, the `INLA` implementation means that the ELK spatial model can be fit jointly with other random effects such as models for temporal trends or nonlinear covariate effects, handle non-Gaussian responses, integrate over parameter uncertainty, and incorporate prior knowledge through expert knowledge and/or for the purpose of robustness.

We will contrast ELK to the more traditional SPDE approach (Lindgren et al., 2011) as implemented in `INLA` to permit fast Bayesian approximate inference for latent Gaussian models where the traditional Matérn covariance function is used for spatial modeling (Lindgren and Rue, 2015). In this context, the SPDE approach is only one choice among many others for making the computations possible: employing low rank covariance matrices (Cressie and Johannesson, 2008; Banerjee et al., 2008; Finley et al., 2009), sparse covariance matrices (Knorr-Held and Raßer, 2000; Sang and Huang, 2012; Konomi et al., 2014; Neelon et al., 2014; Furrer et al., 2006; Hirano and Yajima, 2013), sparse precision matrices (Nychka et al., 2015; Katzfuss, 2017; Katzfuss and Hammerling, 2017; Lindgren et al., 2011; Datta et al., 2016a,b; Guinness, 2019; Guinness and Fuentes, 2017), or algorithmic approaches (Gerber et al., 2018; Guhaniyogi and Banerjee, 2018; Gramacy and Apley, 2015).

In the previous chapter, we observed that accounting for urbanicity could have a fairly dramatic effect on predictions of spatial models, both discrete and continuous, when using data from surveys that were stratified based on urbanicity. However, our classifications of pixels as either urban or rural was imperfect. While individual clusters in the 2014 KDHS had urbanicity classifications that were preassigned, these classifications were based on variables

that are not publicly known using the last census before the DHS. Each census takes place at best every 10 years, and even if the variables used for urban designations were known continuously through space, any classification would be imperfect due to time lags between the census and DHS survey. This forces modelers to either ignore urbanicity or assume that the classification is a static partition of space that remains accurate over large time spans, and to estimate urbanicity for unobserved locations based on proxy data such as population density (Paige et al., 2020; Wakefield et al., 2019). Designing spatial models that are robust to errors in spatial covariates such as via urban misclassification would therefore be very helpful when estimating NMRs and secondary education prevalence in Kenya and other countries using multistage household surveys such as DHS surveys in the developing world. We therefore propose the aforementioned ELK model at least in part in an effort to try to address problems with urban misclassification as well as other spatial confounders that may potentially operate on a multitude of spatial scales.

Because administrative areas are relevant for stratification in household surveys, and also since household surveys are often used to calculate population averages in administrative areas for policymakers, any spatial model used in this context must be able to simultaneously produce accurate averages in areas of varying size. Such models will therefore need to accurately estimate correlations at all spatial scales relevant for the sizes of the areas over which averages are calculated, and account for the uncertainty in those correlation estimates. Accounting for many different spatial scales is also made more important by the fact that population density and urbanicity can change over very short spatial scales. This is illlustrated in Figured 1.2, and Figure 3.4 shows that traditional spatial models using a Matérn covariance may lead to oversmoothing in the rural regions immediately surrounding urban areas when no urban effects are included in the model. We will show that the multiresolution approach of ELK helps to somewhat alleviate this effect.

The outline of this chapter is as follows. In Section 5.2 we introduce ELK. We evaluate

ELK, LK, and a SPDE model in a simulation scenario when fit to random fields with mixtures of short and long-range correlations in Section 5.3. In Section 5.4 the ELK and SPDE models are applied to the women's secondary education completion data, and their predictive performance is assessed. Section 5.5 concludes the chapter with a discussion.

5.2 A Bayesian Extension of LatticeKrig to Latent Gaussian Models

We make two major additions to the formulation of LK in Section 2.3.3: we allow for the model to be fit jointly with other structured random effects, and we allow for non-Gaussian responses. The model for the linear predictor is extended to $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{A}\mathbf{c} + \sum_{i=1} \mathbf{M}_i\boldsymbol{\gamma}_i$, where the matrices $\mathbf{M}_i$ are fixed and define a mapping to the observations from random effects collected in the vectors $\boldsymbol{\gamma}_i$ such as temporal trends, space-time interactions, and other modeled effects. The vector $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^T, \ldots, \boldsymbol{\gamma}_m^T)^T$ is assumed to follow a joint Gaussian distribution. Denote by $\boldsymbol{\theta}_{\mathrm{M}}$ and $\boldsymbol{\theta}_{\mathrm{L}}$ the vectors containing all model and family likelihood hyperparameters respectively. We can then formulate a latent Gaussian model in three stages. In stage 1, we have conditionally independent observations that may be non-Gaussian with likelihood $\pi(y(\boldsymbol{s}_i)|\eta_i, \boldsymbol{\theta}_{\mathrm{L}})$, $i = 1, 2, \ldots, n$. In stage 2, the latent model is a joint Gaussian distribution for $(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{c})|\boldsymbol{\theta}_{\mathrm{M}}$. Lastly, in stage 3, we assign a prior $\pi(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\theta}_{\mathrm{M}}, \boldsymbol{\theta}_{\mathrm{L}})$.

To better understand the $\sum_{i=1} \mathbf{M}_i\boldsymbol{\gamma}_i$ term, and to see why it adds so much generality to ELK, we could consider the relatively simple example of modeling a set of T repeated observations of n spatial locations through time points $t = 1, \ldots, T$. If our covariates aside from β_0, the intercept, can be split into one set of covariates changing only in space and one set of covariates changing only in time, we could then model the fixed effects in space and time as $\mathbf{X}_{\mathrm{S}}\boldsymbol{\beta}_{\mathrm{S}}$ and $\mathbf{X}_{\mathrm{T}}\boldsymbol{\beta}_{\mathrm{T}}$ respectively for $n \times p_{\mathrm{S}}$ matrix $\mathbf{X}_{\mathrm{S}}$ and $T \times p_{\mathrm{T}}$ matrix $\mathbf{X}_{\mathrm{T}}$. Similarly, we might assume that the spatial random effect varied only in space and the temporal random effect varied only in time. If the temporal trend is AR(1), then we can set $\boldsymbol{\gamma} \sim$ AR(1), for a

T dimensional vector $\boldsymbol{\gamma}$. We could then define the model as,

$$\boldsymbol{\eta} = \mathbf{1}_{nT}\beta_0 + (\mathbf{1}_T \otimes \mathbf{X}_{\mathrm{S}})\boldsymbol{\beta}_{\mathrm{S}} + (\mathbf{X}_{\mathrm{T}} \otimes \mathbf{1}_n)\boldsymbol{\beta}_{\mathrm{T}} + (\mathbf{1}_T \otimes \mathbf{A})\mathbf{c} + (\mathbf{I}_T \otimes \mathbf{1}_n)\boldsymbol{\gamma},$$

where '$\otimes$' represents the Kronecker product, and $\mathbf{I}_T$ is a $T \times T$ identity matrix so that $\mathbf{M} = \mathbf{I}_T \otimes \mathbf{1}_n$ adds the coefficients of $\boldsymbol{\gamma}$ identically to the coefficients of $\boldsymbol{\eta}$ associated with the corresponding time point. This model can be fit in the ELK framework. Although not included in this model, interactions between the spatial and temporal effects could be included as well.

The key computational contribution of Rue et al. (2009a) is the combination of this formulation with the INLA approach to make Bayesian inference for the multiresolution latent Gaussian model computationally feasible. The combination is practically achieved by the implementation of the new model within the `INLA` package. We term the extended version of LatticeKrig, with computationally feasible inference, as extended LatticeKrig (ELK). Our implementation exploits `GMRFLib`-library (Rue and Follestad, 2001) functions for sparse symmetric positive definite matrices based on methods described in Rue and Held (2005) when generating the latent coefficient precision and covariance matrices, and also precomputes relevant matrices and normalization factors whenever possible. Details on computations involved in our ELK implementation are given in Appendix C.2.

To ensure σ^2_{S} can be approximately interpreted as the spatial variance and $(\alpha_1, \ldots, \alpha_L)$ as the proportion of spatial variance attributed to the layers, we normalize separately the SAR processes associated with each layer so that the variance of each g_l in the center of the spatial domain is $\alpha_l \cdot \sigma^2_{\mathrm{S}}$. This requires the computation of normalization constants $\omega_1, ..., \omega_L$. Letting $\mathbf{A}^*_l$ be the $1 \times m(l)$ regression row vector that maps the layer l basis coefficients to the value of the basis functions at the center of the spatial domain, each ω_l can be calculated as: $\omega_l = (\mathbf{A}^*_l \mathbf{B}_l^{-1} \mathbf{B}_l^{-T} (\mathbf{A}^*_l)^T)^{-1}$. This is different from LK, since we only normalize the process to have variance σ^2_S in the center of the domain rather than at every point. This has the advantage that it is faster computationally, and we find that if the lattice resolutions and

buffers are chosen using the method discussed in the following paragraph, then the resulting process has spatial variance close to σ_S^2 across the whole spatial domain. In order to avoid matrix inversion and quadratic form computations each time $\mathbf{Q}$ is calculated, we precompute the mappings $f_l : \kappa_l \mapsto \omega_l$ using smoothing splines over a reasonable range of the values of κ_l.

In LK, the recommended setting for the layer resolutions are the relation $\delta_l = 2^{-(l-1)}\delta_1$, and when this relation is used in ELK under the assumption that $\kappa_1 = \ldots = \kappa_L$, we call this the 'fixed' model (ELK-F). We propose to also consider a 'tailored' ELK model (ELK-T) with resolutions chosen for capturing variation at different spatial scales and with κ_l parameters allowed to vary for each layer. Since ELK-T allows the κ_l parameters to vary for each layer, it requires $2L$ hyperparameters, whereas ELK-F requires $L+1$ hyperparameters, although more would be required if other latent effects were included in the $\mathbf{M}\boldsymbol{\gamma}$ term or for any likelihood family hyperparameters. For both models, a conservative guideline is for lattice resolutions to be at most a fifth of the effective range of the corresponding layer to avoid lattice artifacts and for accurate interpretation of the layer's effective range parameter. Since correlation lengths near the spatial domain diameter are very difficult to identify, we recommend choosing δ_1 to be finer than a fifth of the spatial domain diameter, and typically around a twenty fifth of the domain diameter, although the exact choice will depend on the context. Figure 2.1 in Appendix C.4 illustrates how the lattices might be arranged for a specific problem. In the figure and in Section 5.3 we use a buffer of 5 cell widths to avoid edge effects due to the zero boundary condition for the basis coefficients of each layer. The buffer size can be adjusted depending on the estimated effective correlation range for that layer.

Expert knowledge on spatial scales at which dependence is expected could be used to choose appropriate resolutions in ELK-T. Furthermore, the Bayesian formulation allows the inclusion of expert knowledge when setting priors for the interpretable parameters. For

simplicity, we suggest a Dirichlet distribution of order L for the proportion of variances assigned to each layer, that is $\boldsymbol{\alpha} \sim \text{Dirichlet}(a_1, \ldots, a_L)$ with $a_l = 1.5/L$ for $l = 1, \ldots, L$ in order to place equal weight in the prior on each layer, and to ensure the prior is slightly concave for the sake of identifiability. Since the chosen prior concentration parameter is 1.5, the Dirichlet prior is only slightly more concave than the flat Dirichlet distribution that would result if the concentration parameter were 1. On the spatial and nugget standard deviation we place penalized complexity (PC) priors satisfying $\text{P}(\sigma_{\text{S}} > 1) = 0.01$, although this will depend on the context and prior information. See Simpson et al. (2017) for details on PC priors.

We propose setting independent inverse exponential priors for the effective range in each layer, where the effective range for layer l is computed as $\rho_l = \sqrt{8}\delta_l/\kappa_l$. For ELK-T, we recommend beginning by placing a prior on one layer's effective range, scaling priors for other layer effective range parameters proportionally to the lattice grid cell width δ_l for ELK-T. For ELK-F, a single κ parameter is estimated so that $\kappa = \kappa_1 = ... = \kappa_L$, and only one effective range parameter requires a prior. When placing priors on ELK-F or ELK-T effective ranges in this way, a prior on the effective range for one of the layers would therefore determine all effective range priors. Throughout this work, we set the median effective range of the coarsest layer at a fifth of the spatial domain diameter, determining any other effective range priors accordingly. However, the effective range priors can be customized to better suit the context as well the expert knowledge of the modeler.

A fully functional proof-of-concept implementation of ELK is freely available on Github at `https://github.com/paigejo/ELK`. Since it is implemented in `R` and not natively in `C++`, it does not reach its full potential in terms of speed.

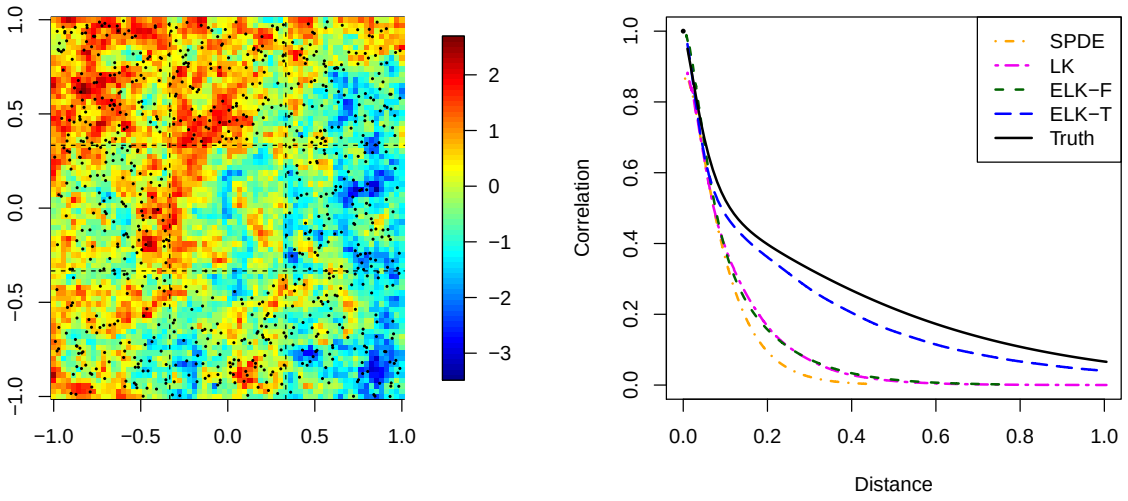
5.3 Assessing Performance Under Multiscale Dependence

5.3.1 Simulation Setting

We first simulate a spatial GRF u on the square $[-1,1]^2$. The GRF has the covariance function $C(d) = 0.5(C_1^*(d;0.08)+C_1^*(d;0.8))$, where $C_1^*(d;\sigma_S^2) = (\sqrt{8}d/\rho)K_1(\sqrt{8}d/\rho)$ denotes the Matérn correlation (Stein, 1999) at distance d with smoothness $\nu = 1$ and effective spatial range ρ, and where K_1 is the modified Bessel function of the first order and second kind. The correlation function is plotted in Figure 5.1 along with an example realization. The domain is then subdivided into a regular 3×3 grid, and we draw 800 observations at random locations, $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_{800}$, in the outer eight grid cells, but draw no observations within the central grid cell. We assume the unobserved latent process is $\eta_i = u(\boldsymbol{s}_i)$, and draw each observation $Y(\boldsymbol{s}_i)$ from $Y(\boldsymbol{s}_i)|\eta_i \sim \mathcal{N}(\eta_i, 0.1^2)$ for $i = 1, 2, \ldots, 800$. We fit several models, which we will describe in the next section, to the data, and generate predictions of the spatial process Y on a fine 70×70 grid and predictions of areal averages of the process Y for the nine subdivision areas approximated numerically as averages of the values of Y on the 70×70 fine grid over each of the 9 areas. The whole procedure is repeated 100 times, and, for each realization, the predictions are scored in comparison to the truth. We choose to use Y as the process for comparing predictions rather than u so that comparison metrics are more similar to cross-validation, where only Y, and not u, is directly observed at the observation locations.

5.3.2 Models Used in the Simulation Study

We use ELK-T with two layers: a grid of 14×14 basis knots and a grid of 126×126 knots over the spatial domain (not including the five knot buffer for each layer), which results in lattice resolutions of 0.154 and 0.016 respectively. In this case, the coarse and fine scale layers have at least five basis functions per 0.8 and 0.08 spatial units respectively. Further



(a) One realization of the spatial field

(b) True and estimated correlation functions

Figure 5.1: (a) One of the 100 spatial field realizations. Black dots indicate the 800 observation locations and dashed lines indicate the 3×3 grid used for areal predictions (b) True and estimated correlation functions averaged over 100 realizations.

we use LK and ELK-F with three layers composed of 14×14, 37×37, and 53×53 lattice grids over the spatial domain with 0.154, 0.077, and 0.038 resolutions respectively. LK is fit using `LatticeKrig` in `R`, and for both LK and ELK-F, we use a single layer-independent parameter κ. Additionally, we fit an approximation to the Gaussian process with Matérn covariance and smoothness $\nu = 1$ using the SPDE approach with `INLA` (Lindgren et al., 2011; Lindgren and Rue, 2015). The mean triangular mesh segment length is approximately 0.0064 within the spatial domain.

This gives in total four models: ELK-T, ELK-F, LK, and SPDE. In all cases we use a PC prior for the nugget variance satisfying the tail probability $\mathrm{P}(\sigma_\epsilon > 1) = 0.01$, and for

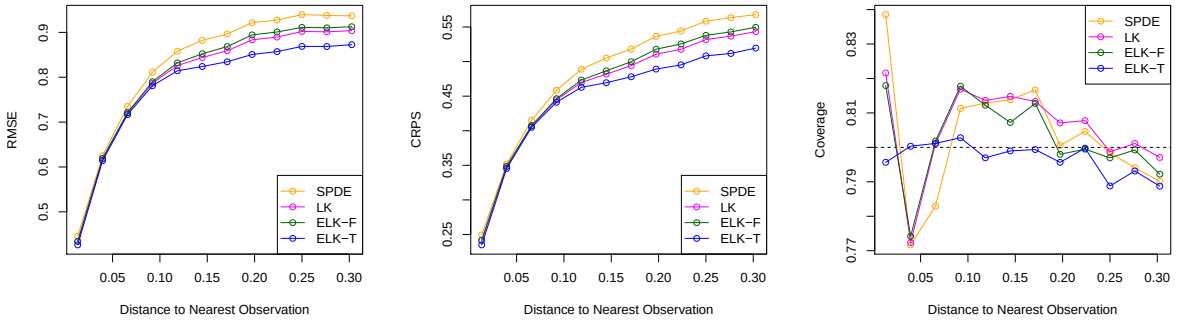
the SPDE model we use the prior derived in Fuglstad et al. (2019b) on the effective range and spatial variance. The median effective range for the prior is a fifth of the spatial domain diameter, and the spatial standard deviation again satisfies $P(\sigma_S > 1) = 0.01$. We use PC priors for the ELK-T and ELK-F spatial standard deviation also satisfying $P(\sigma_S > 1) = 0.01$, and use the effective range priors recommended in Section 5.2.

5.3.3 Results

For each realization and each of the Bayesian models, we generate 1,000 independent samples from the posterior distribution of Y (or conditional distribution in the case of LatticeKrig), estimate uncertainty in the parameters, and calculate covariance functions for each of the 100 independent parameter samples. We used only 100 parameter samples when generating covariance function draws since for each draw the corresponding precision matrix for u must be inverted, which is especially computationally intensive for LatticeKrig since it does not take advantage of `GMRFLib` library functions for factoring sparse symmetric positive definite matrices, and since ELK uses a simplified normalization scheme that precomputes normalization factors. In the case of LatticeKrig, we use the Hessian of the negative log likelihood to draw covariance parameter samples.

Figure 5.1b) shows the central correlation function estimate for each of the models together with the true correlation function. The ELK-T model approximates the true correlation function over all distances well, while the other models strongly underestimate the spatial correlation after distance of 0.1, and have negligible correlation after distances of approximately 0.5.

Pointwise predictive scoring rules calculated by distance from prediction location to nearest observation are shown in Figure 5.2. The RMSE and CRPS of ELK-T are the best in all of the distance bins. Differences in RMSE and CRPS among the models tend to increase as the distance to the nearest observation increases, but interestingly the differentiation is



(a) RMSE vs distance to nearest observation

(b) CRPS vs distance to nearest observation

(c) 80% Coverage vs distance to nearest observation

Figure 5.2: Scoring rules calculated in bins depending on distance to nearest observation. The scores are averaged over 100 simulations, and include (a) RMSE, (b) CRPS, and (c) 80% uncertainty interval coverage.

larger in the first bin than in the second bin. We believe this is due to the fact that ELK-T is able to capture the short range spatial correlation better than the three other models. The differences in RMSE and CRPS values for each model become increasingly large with longer distance to closest observation, indicating increasingly differing ability to accurately predict with longer distances.

Table 5.1 shows the summarized point and areal prediction scores. In terms of both pointwise and areal scores, the SPDE predictions have the worst RMSE, CRPS, and coverage in all cases, although the coverage of all the models matches the nominal level of 80% in the pointwise case. The coverage of the SPDE model is especially poor near the observations, indicating its inability to simultaneously capture short and long scale spatial correlations. Figure 5.2 clearly demonstrates that even though SPDE achieves the correct nominal cov-

	RMSE	CRPS	80% Cvg	Runtime (min.)
Pointwise				
SPDE	*0.605*	*0.342*	**80**	**2.0**
LK	0.594	0.334	**80**	*51.1*
ELK-F	0.594	0.335	**80**	9.4
ELK-T	**0.587**	**0.329**	**80**	12.1
Areal				
SPDE	*0.137*	*0.056*	*75*	**2.0**
LK	0.121	0.051	77	*51.1*
ELK-F	0.125	0.052	77	9.4
ELK-T	**0.108**	**0.048**	**79**	12.1

Table 5.1: Scoring rules averaged over 100 simulated realizations and over a regular 70×70 grid of prediction locations across the entire spatial domain and areally integrated over all nine cells in the 3×3 regular grid partitioning the domain. Averages are calculated for each of the considered models. *Italics* indicate worse performance, **boldface** indicates better performance.

erage overall, this is in spite of considerable over- and undercoverage depending on how far prediction locations are from the observations. There is also far more variability in coverage between bins for the SPDE model than ELK-T.

The runtime for the SPDE model is clearly the best. This is in part due to having an implementation that is pre-existing and optimized in the `INLA` package, whereas ELK-T and ELK-F were implemented manually using the comparatively slow `rgeneric` framework intended for prototyping new models and special cases in `INLA`. However, the fact that the

SPDE model requires only two hyperparameters excluding any family likelihood hyperparameters, compared to the four required in this case for ELK-F and ELK-T, further improves its computational performance. LK had the longest runtimes in large part due to the implementation of the predictive distribution sampling when calculating SEs. Drawing the 1,000 samples took over 33 minutes on average for LK, whereas drawing the same number of samples for the ELK-F model took under 2 minutes on average, and also included sampling over uncertainty in the hyperparameters.

The areal scores in Table 5.1 indicate a strong improvement from the SPDE model to ELK-F, and from ELK-F to ELK-T in terms of RMSE and CRPS. From the SPDE model to ELK-T, pointwise RMSE and CRPS scores improved respectively from 0.605 to 0.587 (3.0%) and from 0.342 to 0.329 (3.8%). However, in the integral prediction case, RMSE and CRPS scores improved respectively from 0.137 to 0.108 (21%) and from 0.056 to 0.048 (14%).

Table C.1 in Appendix C.4 shows that the improvements in areal predictions are even larger when considering only the central grid cell, but Table C.2 in Section C.4 shows that there are improvements even when only the eight outer grid cells are considered. In summary, the results of this application show that multi-scale covariance models are essential both for accurate estimation of the covariance structure and for making predictions when the true covariance function is a mixture of short range and long range behavior.

5.4 Prevalence of Secondary Education Completion

5.4.1 Analysis

We return to the counts of secondary education completion for young women aged 20-29 in Kenya in 2014 using the 2014 KDHS. The 2014 KDHS household survey contains responses from individuals sampled from 1,612 clusters in 47 counties, each of which except Nairobi and Mombasa (which are both entirely urban) contain both urban and rural strata, making 92 strata in total. These 47 counties subdivide the 8 geographical provinces in Kenya. The

response at cluster c, conditional on the probability of secondary education completion, $\mu(\boldsymbol{s}_c)$ at cluster spatial location $\boldsymbol{s}_c$, $c = 1, \ldots, 1612$, is modeled as, $Y(\boldsymbol{s}_c)|\mu(\boldsymbol{s}_c) \sim \text{Bin}(n_c, \mu(\boldsymbol{s}_c))$, where n_c is the total number of women aged 20-29 sampled in the cluster. The probability $\mu(\boldsymbol{s})$ is modeled on logit scale as,

$$\eta_c = \log\left(\frac{\mu(\boldsymbol{s}_c)}{1-\mu(\boldsymbol{s}_c)}\right) = \beta_0 + u(\boldsymbol{s}_c) + \beta^{\text{URB}}\mathbf{1}\{\boldsymbol{s}_c \in U\} + \epsilon_c, \quad c = 1, 2, \ldots, 1612, \tag{5.1}$$

with intercept β_0, spatial random effect $u(\boldsymbol{s}_c)$ with spatial variance σ^2_{S}, fixed effect for urban areas β^{URB}, and mean zero iid Gaussian cluster random effect ϵ_c with variance σ^2_ϵ. The indicator $\mathbf{1}\{\boldsymbol{s}_c \in U\}$ is 1 if $\boldsymbol{s}_c$ is in U, the set of urban areas in Kenya, and 0 otherwise. LK is not applicable due to the binomial likelihood. We consider four alternatives for u: SPDE_u/SPDE_U and ELK-T_u/ELK-T_U models, where 'U' and 'u' respectively denote that urban effects are or are not included.

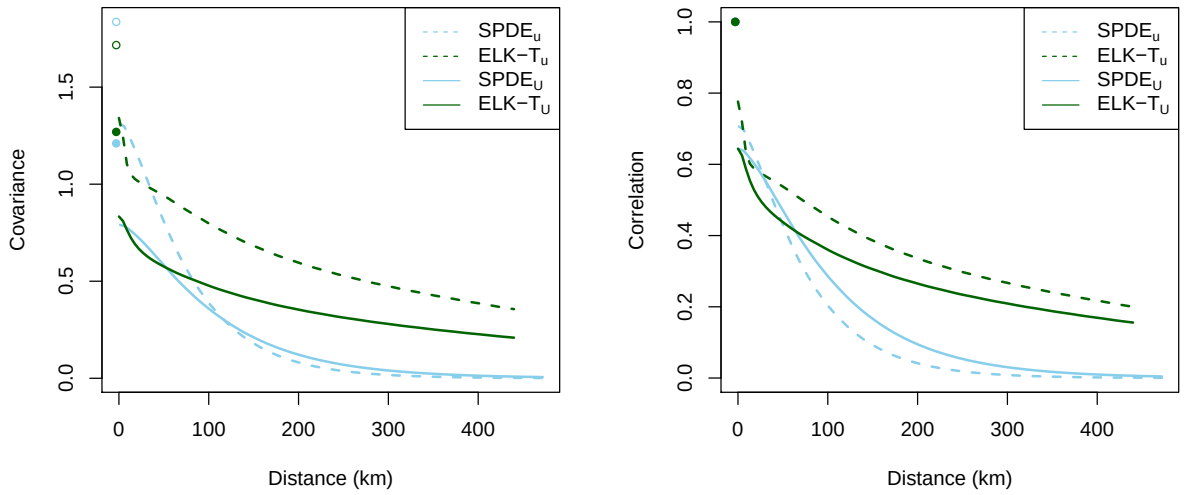
For ELK-T_u and ELK-T_U, the coarse lattice layer has 37km resolution, while the fine layer resolution was set to be 5km resolution in order to be able to capture sharp changes from urban localities to their rural surroundings. The SPDE model has an average triangular mesh segment length of approximately 15km across the spatial domain. The spatial domain diameter is approximately $1,445$km, so the prior median effective range was set to be one fifth of that, or 289km, for the SPDE model and for the coarsest layer of the ELK models. We again place PC priors on the spatial and cluster variance parameters such that $\text{P}(\sigma_\epsilon > 1) = 0.01$ and $\text{P}(\sigma_\text{S} > 1) = 0.01$, except now the parameters should be interpreted on logit scale. All covariates except for the intercept are given noninformative Gaussian priors with zero mean and 0.001 precision, and the intercept is given an improper $\text{Unif}(-\infty, \infty)$ prior.

Central estimates for the correlation and covariance functions of the fitted models are shown in Figure 5.3. Compared to the SPDE models, the ELK-T models incorporate more long scale spatial correlation while also modeling short scale correlations with more subtlety as shown by their long tailed covariance and correlation functions with sharp downward trends at small spatial distances. Including urbanicity as a covariate substantially reduces

the spatial variance for all models, and also reduces the variance of the spatial nugget. We find that including an urban effect explains spatial variation at both short and long scales, because sharp changes due to urban/rural boundaries are accounted for, as well as long scale correlations across rural regions. We see this effect in the estimated correlation function of the ELK-T models, where the magnitude of the relatively sharp downward trend in correlation at small distances decreases when the urban effect is included, and where the long tail shortens slightly as well. Since the likelihood under Matérn correlation (or Matérn approximations like the SPDE model) is primarily affected by short correlation scales, the sharp changes in education due to changes in urbanicity rather than the long scale correlations induced by large areas being rural drive the correlation function estimate. Hence, including the urban effect in the SPDE model removes some of the otherwise unmodeled spatial correlation at short spatial scales, increasing the estimated effective range. It is worth noting, however, that even with an urban effect, the ELK-T_U model covariance estimates are still different to those in the $SPDE_U$ model at both short and long scales.

In Figure 5.4 we give pixel level predictions at the 5km×5km resolution of secondary education prevalence as well as relative credible widths, which we define as credible widths divided by the corresponding central estimates. Areal predictions are created based on aggregation of pixel estimates weighted by population density as described in Equations (5-6) of Paige et al. (2020), except leaving out cluster effects by setting them to 0 rather than integrating over them as done in Equation (7) of Paige et al. (2020). Predictions and relative credible widths aggregated to the county and province levels are shown in Section S.3 in Figures 9 and 10. Tables of the county level and province predictions for the models with urban effects as well summary statistics for the model parameters are given in S.3 in Tables 7-9.

The pixel level predictions show nearly indistinguishable differences in predictions and uncertainties between the $SPDE_U$ and ELK-T_U models, but much more significant differences



(a) Covariogram Estimates

(b) Correlogram Estimates

Figure 5.3: (a) Spatial covariance, and (b) correlation estimates. The spatial nugget is plotted as the dots at zero distance with the color corresponding to the model given in the legends. Filled dots are plotted for models including urban effects, and unfilled dots are plotted for models without urban effects.

in the predictions between the SPDE$_u$ and ELK-T$_u$ models. In particular, the ELK-T$_u$ model shows reduced spatial oversmoothing near urban areas, and higher uncertainties overall. These uncertainties reflect that an important confounder in urbanicity is not included as a covariate. The reduction in oversmoothing is especially noticeable in the north and east counties with large rural areas and spatially concentrated urban areas, although there are reductions in oversmoothing in other areas as well. The differences between the models without urban effects, and the similarities between the models with urban effects are further highlighted in the pair plots in Figure 5.5, which shows the predictions of the SPDE$_u$, ELK-

T_u, and $SPDE_U$ models sequentially move towards the predictions of the ELK-T_U.

That the $SPDE_U$ and ELK-T_U predictions are essentially indistinguishable lends credence to our predictions by showing they are robust to modeling assumptions. It also suggests that there is little identifiable spatial covariance at very short scales that is not already accounted for by urbanicity, and that the overall effect of remaining spatial confounders probably varies smoothly over medium to long spatial scales.

5.4.2 Validation

We use two different schemes to validate our models: leave one province out, and stratified, eight-fold cross validation (CV). In the leave one province out scheme, we calculate scoring rules based on the predicted distributions of the left out clusters in each of the 8 provinces consecutively, averaging the scores within each province, and then averaging the province scores to get the final reported scores. In the stratified, eight-fold CV, we randomly partition the set of clusters in each of the 92 strata (47 counties with each except of Nairobi and Mombasa begin urban and rural) into eight roughly equal sized folds. We make sure that for a given stratum, the difference between the number of clusters in each fold is different by at most one, and that which folds get more clusters than others is random. We choose eight folds since the smallest stratum has only eight clusters. The two different validation schemes give an idea of both short and long scale predictive errors due to the distribution of how far away left out clusters are from in sample observations. The leave one province out scheme better identifies long scale errors, and the stratified CV better identifies short and medium scale errors. The boundaries of the 8 provinces are plotted in Figure 5.4 along with county boundaries.

The results from the leave one province out and the stratified CV are given in Table 5.2. The ability of the ELK-T model to account for more flexible spatial covariance structures than the SPDE model leads to as good or better predictions as shown by RMSE, CRPS, and

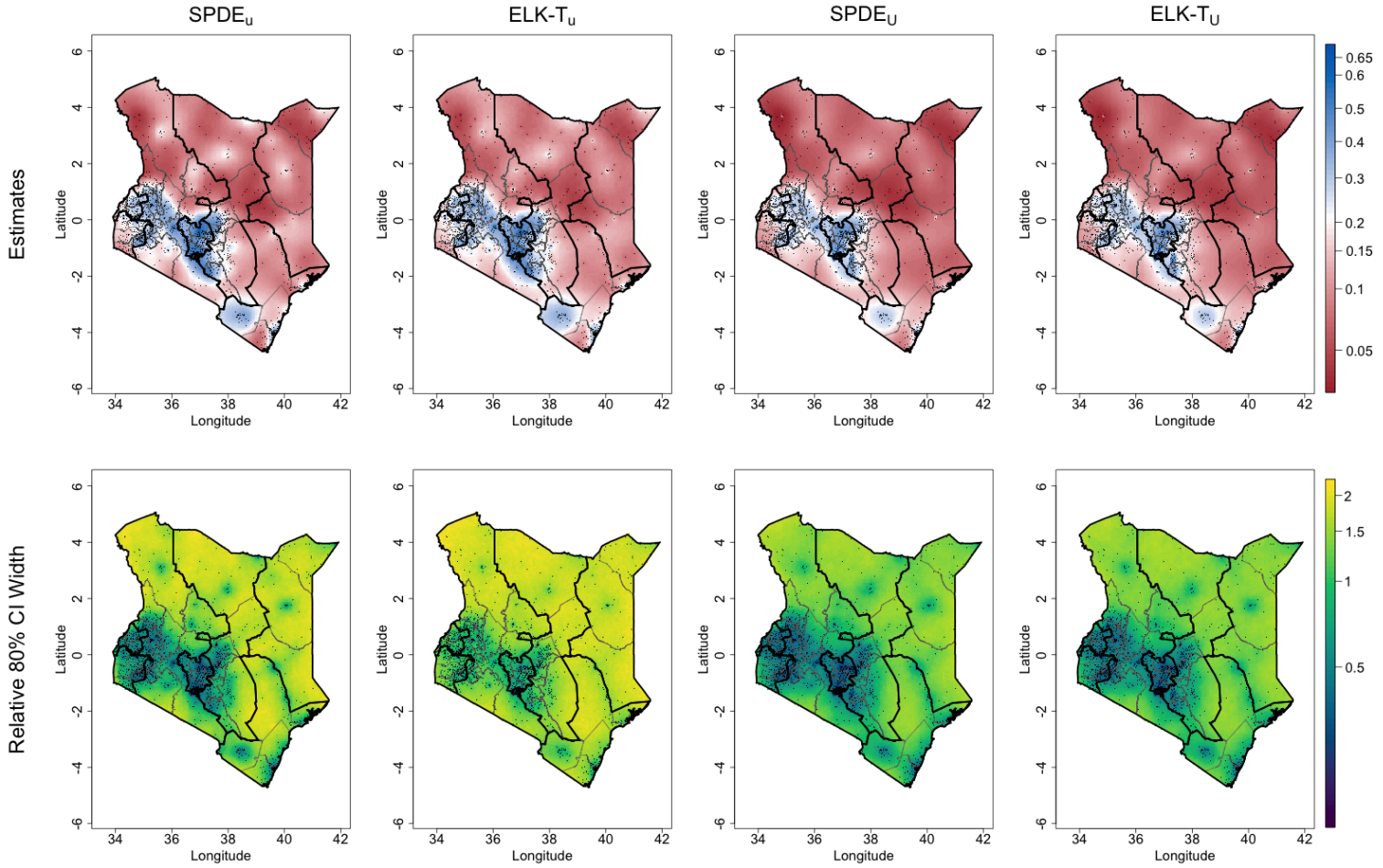


Figure 5.4: Central 5km×5km pixel level predictions (top row) and relative 80% credible interval widths (bottom row) of secondary education prevalence for young women in Kenya in 2014. Models with subscript 'U' and 'u' respectively do and do not include urban effects. Observation locations are plotted as black dots, provinces as thick black lines, and counties as thin gray lines.

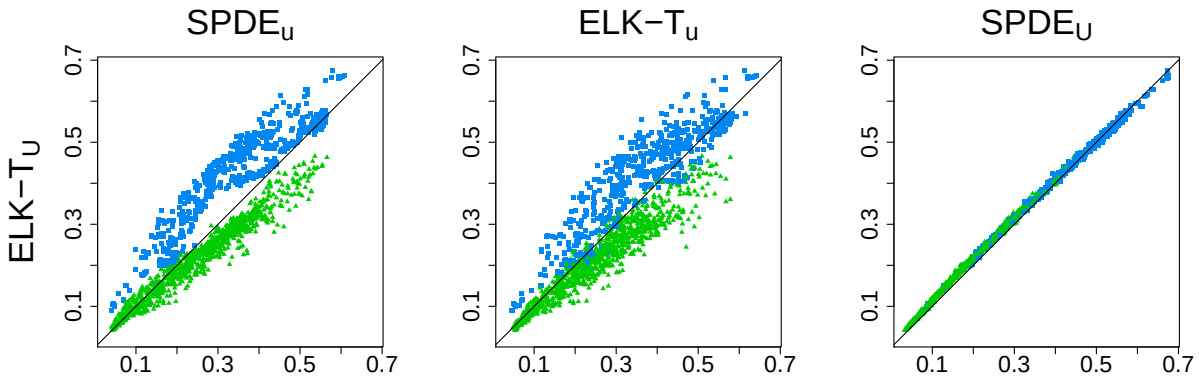


Figure 5.5: Pair plot of the cluster level estimates comparing the considered models' estimates of secondary education prevalence to the ELK-T_U. The '▲' symbols are rural clusters, while '■' symbols are urban clusters.

	RMSE	CRPS	80% Cvg	Width
Leave One Province Out				
$SPDE_u$	*0.238*	*0.129*	76	0.52
$SPDE_U$	0.224	0.119	*74*	**0.47**
ELK-T_u	0.234	0.125	**77**	*0.53*
ELK-T_U	**0.223**	**0.117**	**77**	0.49
Stratified 8-Fold				
$SPDE_u$	*0.226*	*0.119*	73	0.46
$SPDE_U$	**0.218**	0.114	*72*	**0.42**
ELK-T_u	0.223	0.117	**77**	*0.49*
ELK-T_U	**0.218**	**0.113**	75	0.45

Table 5.2: Scoring rules calculated for each model using leave one province out and stratified 8-fold cross validation. Scores are averaged for each province, over urban areas, and over rural areas. *Italics* indicate worse performance, **boldface** indicates better performance.

coverage standpoints, although the improvement is clearly greater when the urban effect is absent in the model. Improvements were especially obvious in the leave one province out CV, where long range correlations mattered more, and relative improvements were greater for CRPS than for RMSE. For leave one province out CV, RMSE improved by 1.7% when urban effects were not included in the SPDE and ELK-T models respectively, and by 0.4%, while CRPS improved by 3.1% when urban effects were not included, and by 1.7% otherwise. The SPDE_{U} model had the worst coverage with 74%, and both ELK-T models tied for the best coverage with 77%.

5.5 Discussion

The LK approach introduced by Nychka et al. (2015) attempts to address the question of how to flexibly model spatial covariance at different spatial scales in a computationally feasible way. However, in a spatial context where identifiability is already difficult, spatial confounders and the flexibility of LK when layer correlation ranges are allowed to independently vary further reduces identifiability. In this case, it may be necessary to account for prior information such as expert knowledge or to penalize model complexity, and it will certainly be important to integrate over parameter and hyperparameter uncertainty. By allowing for this without significant reductions in computational performance reductions, ELK's Bayesian framework is a valuable extension over standard LK. It is not only more robust, but better accounts for multiple levels of uncertainty. Because of this, modelers might be less wary of fitting models with more complex covariance structure.

In ELK-T, due to the flexibility in choosing layer resolutions and the fact that its effective range parameters are fit independently, ELK's Bayesian framework is particularly important. We found ELK-T performed much better than ELK-F for the simulations we considered and the application since it was better able to efficiently model variation at contextually relevant spatial scales. In light of this, ELK's use of Bayesian inference is all the more important.

Another advantage of ELK is that it eliminates the assumption of Gaussian responses by extending the LK framework to latent Gaussian models. This allows modeling responses with a diverse set of distributions, such as distributions in the exponential family, and even some others such as the betabinomial distribution, as long as priors on the latent model components are Gaussian. The implementation of ELK in `INLA`, avoids the computational expense of MCMC when integrating over parameter uncertainty. Moreover, we show that, computationally, ELK performs approximately better than LK when uncertainty in the predictions and covariance parameters is desired. ELK also has access to the suite of models that can be fit in `INLA` such as nonlinear random effects models for time series or covariates.

It is important to note that ELK-F requires $L + 1$ covariance parameters for L layers excluding variance parameters of the likelihood family, and ELK-T requires $2L$ covariance parameters. Due to the exponential growth in the computation time requirements of optimization and integration over hyperparameter uncertainty as the number of hyperparameters grow, there is a limit to the number of layers for which computation is feasible. It is recommended for the number of hyperparameters in `INLA` models to be between 2 and 5, but certainly not exceeding 20 (Rue et al., 2017). Hence, computationally this method should typically use at most 4 layers for ELK-F and 2 or 3 layers for ELK-T for likelihoods without extra hyperparameters. It is certainly limited to 19 layers for ELK-F and 9 layers for ELK-T, which are far more than is necessary for both models. In general, for most practical purposes we see little reason to include more than 3 layers for ELK-T and 5 layers for ELK-F even if computation is feasible due to difficulty in model identification and lack of difference in predictive performance, although there may be some exceptions to this rule for ELK-F in particular since it can only model effective correlation ranges 2^{L-1} times larger than the range modeled by the finest layer.

In the simulation study, we show that the ability of LK and ELK to model spatial covariance flexibly can substantially improve predictive performance at both short and long scales.

We find that, while short scale dependence is most important for point level predictions near observations, long scale dependence can matter more when making predictions in data sparse regions, and when making areal predictions.

When we apply the ELK model to a 2014 Kenya DHS dataset with information on the prevalence of secondary education for women aged 20-29 in 2014, we find substantial reductions in spatial oversmoothing relative to a SPDE model, especially when urbanicity was included as a covariate in the models. Evidence of short scale spatial confounding was present in the estimate of the spatial correlation function in the ELK model with no urban effect, indicating that ELK can make predictions more robust to spatial confounding as well as be indicative of the spatial scales at which spatial confounding is occurring. This in turn can suggest what variables should be included as covariates, and as an informal check for spatial confounding. In general, it is very difficult to tell whether an unmeasured covariate is confounding results, but ELK provides at least a modicum of insurance against this. Since DHS household surveys tend to consist of clusters that are spatially concentrated in urban areas and sparsely distributed in rural areas, this is an application that ELK is well suited for.

Depending on the context, one may choose to select lattice resolutions that are independent of each other rather than changing by a factor of two from one layer to the next as in standard LK. In both the illustrative example and the application, we found that forcing each consecutive layer to have double the resolution along each dimension made modeling the fine and long scale changes simultaneously difficult from a computational perspective due to the number of hyperparameters and basis functions required. In such situations, we advocate for tailoring the resolutions of each lattice to enable them to model a set of effective ranges of interest.

Chapter 6

CONCLUSIONS

In this dissertation, we describe a number of methods for estimating subnational health and demographic indicators using multistage stratified cluster surveys such as DHS household surveys. Many models for population prevalence that are used today, including models from IHME, WorldPop, and others, ignore stratification by urbanicity and do not adequately account for nonspatial random variation at the cluster level and population numerator and denominator variation when EA locations are unknown (Golding et al., 2017; Osgood-Zimmerman et al., 2018; Graetz et al., 2018; Utazi et al., 2018a; Neal et al., 2019; Giorgi et al., 2018), and no model-based SAE approach has been proposed to account for multiscale spatial correlations, uncertainty in EA locations, and population numerator and denominator variation at the EA level. We address these modeling concerns in this dissertation in simulation studies as well as applications to NMR and women's secondary education prevalence data from the 2014 Kenya Demographic Health Survey (KDHS) (KDHS, 2014).

In Chapter 3, we investigated the extent to which accounting for the sample design affects the predictive performance at the aggregate level of interest for health policy decisions. We considered various commonly-used models and introduced a new Bayesian cluster level model including an effect for stratification and with a discrete spatial smoothing prior. The investigation was performed via a simulation study in which realistic sampling frames were created for Kenya, based on population and demographic information, with a survey design that mimicked the 2014 KDHS. We found that including stratification and cluster level random effects can improve predictive performance. Spatially smoothed direct (weighted) estimates and area level models accounting for stratification were robust to the underlying population

and survey design. Continuous spatial models showed some promise in the presence of fine scale variation; however, these models required the most "hand holding". Subsequently, we examined how the models performed on real data, estimating the prevalence of secondary education for women aged 20–29 and neonatal mortality rates, using data from the 2014 KDHS. Validating at the cluster level, we again found that including a fixed effect for stratification improved predictive performance for the model-based estimates, while not including an effect for stratification led to dramatic oversmoothing for the stochastic partial differential equation (SPDE) model in rural areas surrounding urban locations. Although we found evidence of cluster level variation, accounting for this variation in areal predictions proved nontrivial. This work has been published (Paige et al., 2020).

In Chapter 4, we build on the modeling approach of Chapter 3, extending the aggregation framework as well as other approaches such as Dong and Wakefield (2020) with our proposed framework for generating population aggregates of health and development indices from spatial models for risk indexed at the cluster level called Combined Population Aggregation Models (CPAM). In addition to spatial correlations and stratification, the proposed CPAMs are able to account for cluster level variation, EA location uncertainty, and target population numerator and denominator variation simultaneously by using a superpopulation model for sampling over variation in these quantities. We show that not accounting for these additional sources of variation can lead to anticonservative uncertainty estimates, especially at the Admin2 level and smaller for prevalence and total deaths, and also at the Admin1 level for the relative prevalence of urban versus rural parts of an area. The proposed superpopulation model is directly implied (without any need for parameter fitting) by publicly available population density services and census data. The SC and SCP CPAMs we provide for risk and prevalence give central predictions for aggregate quantities that agree with those of commonly used aggregation models such as used in Paige et al. (2020) even though uncertainty in the aggregate predictions increase due to the additional variation from the superpopulation

model. Because CPAMs can account for population numerator and denominator variation, they can be used to model prevalence rather than risk. We provide a number of examples of CPAMs of increasing complexity until all of the aforementioned sources of variability are accounted for, but that all use the same SPDE model for superpopulation risk (although it is simple to plug in other spatial risk models indexed at the cluster level), and sharing the same central predictions. We assessed the predictions of each of the considered aggregation models via both simulation study and validation, and also compared the predictive uncertainties of the models for 2014 KDHS NMR data, finding that cluster level and population numerator variation in particular can make a considerable impact on predictive uncertainties at the constituency level.

In Chapter 5, we provide method to make continuously indexed spatial models for demographic and health indicators from survey data more robust to misclassification in spatial stratification variables. We propose a multiresolution approach for spatial analysis of non-Gaussian responses using latent Gaussian models and Bayesian inference via integrated nested Laplace approximation (INLA). The approach extends 'LatticeKrig', but uses a reparameterization of the model parameters that is intuitive and interpretable so that modeling and prior selection can be guided by expert knowledge about the different spatial scales at which dependence acts. The priors can be used to make inference robust and integration over model parameters allows for more accurate posterior estimates of uncertainty. The extended LatticeKrig (ELK) model is compared to a standard implementation of LatticeKrig (LK), and a Matérn covariance model, and we find modest improvement in spatial oversmoothing and prediction for the ELK model for counts of secondary education completion for women in Kenya collected in the 2014 KDHS. Through a simulation study with Gaussian responses and a realistic mix of short and long scale dependencies, we demonstrate that the improvements in RMSE and coverage achieved by ELK relative to the SPDE approach increases with distance to nearest observation and are larger for areal than pointwise predictions. This is

especially true when the ELK basis layer resolutions are tuned so that it is possible to identify the multiscale spatial correlation, and when layer correlation scale parameters are allowed to vary independently.

In the future, we hope to propose validation metrics that can be used with point level observations and areal level predictions. In particular, we would like any validation target statistics to incorporate joint spatial correlations within the area of interest, and account for the survey design of the data. Validating against direct estimates, while an important step in the right direction, adds variation in the direct estimate to variation in the population level prediction estimates, and the direct estimate variance may dominate, leading to coverages that are near 80% simply because the direct estimates themselves have 80% coverage. Instead, directly predicting the sample prevalence may allow for higher power assessments of predictive coverage.

Regarding validation, Thomson et al. (2020) proposes using gridded population survey sampling, a novel sampling approach where the sampling frame is enumerated at the pixel level rather than at the EA level. For pixel level predictions, gridded population surveys could provide an opportunity for validation. In fact, even without gridded population surveys, validating pixel level prevalences may help to illuminate how problematic pixel level predictions really are. We believe that urban misclassification, for instance, may be a problem when generating pixel level predictions, and that pixel level validation may help to assess whether or not this is really the case. To make continuous spatial models more robust to urban misclassification, it would be possible to model the probability of any given EA as being urban or rural, conditional on a pixel's population density and the stratum. This could be fit using the population density and census information, and could therefore only come into play after the spatial risk model has already been fit.

More work must be done to ensure that spatial risk models can account for survey effects. For instance, Rabe-Hesketh and Skrondal (2006) and Skrondal and Rabe-Hesketh

(2009) provide a method to directly incorporate survey weights when fitting multilevel models. Additionally, it is possible to account for survey weights in linear smoothers, and it may also be possible in the kriging prediction equations, indicating that there may be many possible avenues for incorporating survey weights into continuous spatial risk models. Both the survey design and the population frame must be accounted for when making areal predictions from survey data, and incorporating survey-weighted risk models into a CPAM aggregation framework would allow for this. When addressing this problem, the design must be accounted for in both the model fitting as well as in the continuous spatial predictions, and the population frame in the aggregation to the areal level.

BIBLIOGRAPHY

Banerjee, S., Gelfand, A. E., Finley, A. O., and Sang, H. (2008). Gaussian predictive process models for large spatial data sets. *Journal of the Royal Statistical Society, Series B*, 70:825–848.

Beatriz, E. D., Molnar, B. E., Griffith, J. L., and Salhi, C. (2018). Urban-rural disparity and urban population growth: A multilevel analysis of under-5 mortality in 30 sub-Saharan African countries. *Health & Place*, 52:196–204.

Besag, J., York, J., and Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistics and Mathematics*, 43:1–59.

Chen, C., Wakefield, J., and Lumley, T. (2014). The use of sample weights in Bayesian hierarchical models for small area estimation. *Spatial and Spatio-Temporal Epidemiology*, 11:33–43.

Congdon, P. and Lloyd, P. (2010). Estimating small area diabetes prevalence in the US using the behavioral risk factor surveillance system. *Journal of Data Science*, 8:235–252.

Cressie, N. and Johannesson, G. (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B*, 70:209–226.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016a). Hierarchical nearest-neighbor Gaussian process models for large geostatistical datasets. *Journal of the American Statistical Association*, 111:800–812.

Datta, A., Banerjee, S., Finley, A. O., and Gelfand, A. E. (2016b). On nearest-neighbor Gaussian process models for massive spatial data. *Wiley Interdisciplinary Reviews: Computational Statistics*, 8:162–171.

Devarajan, S. (2013). Africa's statistical tragedy. *Review of Income and Wealth*, 59:S9–S15.

Deville, J.-C. and Tille, Y. (1998). Unequal probability sampling without replacement through a splitting method. *Biometrika*, 85:89–101.

DHS Program (2019). The DHS program – AIDS indicator surveys (AIS). https://dhsprogram.com/What-We-Do/Survey-Types/AIS.cfm.

Diggle, P. and Giorgi, E. (2016). Model-based geostatistics for prevalence mapping in low-resource settings. *Journal of the American Statistical Association*, 111:1096–1120.

Diggle, P. J. and Giorgi, E. (2019). *Model-based Geostatistics for Global Public Health: Methods and Applications*. Chapman and Hall/CRC, Boca-Raton.

Dong, T. Q. and Wakefield, J. (2020). Modeling and presentation of vaccination coverage estimates using data from household surveys. *arXiv preprint arXiv:2004.03127*.

Fay, R. and Herriot, R. (1979). Estimates of income for small places: an application of James–Stein procedure to census data. *Journal of the American Statistical Association*, 74:269–277.

Filippone, M., Zhong, M., and Girolami, M. (2013). A comparative evaluation of stochastic-based inference methods for Gaussian process models. *Machine Learning*, 93:93–114.

Finley, A. O., Sang, H., Banerjee, S., and Gelfand, A. E. (2009). Improving the performance of predictive process modeling for large datasets. *Computational Statistics and Data Analysis*, 53:2873–2884.

Fuglstad, G.-A., Hem, I. G., Knight, A., Rue, H., and Riebler, A. (2019a). Intuitive principle-based priors for attributing variance in additive model structures. *Bayesian Analysis*, advance publication.

Fuglstad, G.-A., Simpson, D., Lindgren, F., and Rue, H. (2019b). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*, 114:445–452.

Furrer, R., Genton, M. G., and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, 15:502–523.

Geir-Arne Fuglstad, Z. R. L. and Wakefield, J. (2020). Spatio-temporal estimation of demographic and health indicators from household surveys. In preparation.

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

Gerber, F., de Jong, R., Schaepman, M. E., Schaepman-Strub, G., and Furrer, R. (2018). Predicting missing values in spatio-temporal remote sensing data. *IEEE Transactions on Geoscience and Remote Sensing*, 56:2841–2853.

Gething, P., Tatem, A., Bird, T., and Burgert-Brucker, C. (2015). Creating spatial interpolation surfaces with DHS data. Technical report, ICF International. DHS Spatial Analysis Reports No. 11.

Gething, P. W., Casey, D. C., Weiss, D. J., Bisanzio, D., Bhatt, S., Cameron, E., Battle, K. E., Dalrymple, U., Rozier, J., Rao, P. C., Kutz, M., Barber, R., Huynh, C., Shackleford, K., Coates, M., Nguyen, G., Fraser, M., Kulikoff, R., Wang, H., Naghavi, M., Smith, D., Murray, C., Hay, S., and Lim, S. (2016). Mapping plasmodium falciparum mortality in Africa between 1990 and 2015. *New England Journal of Medicine*, 375:2435–2445.

Geyer, C. J. and Meeden, G. D. (2005). Fuzzy and randomized confidence intervals and p-values. *Statistical Science*, 20:358–366.

Giorgi, E., Diggle, P. J., Snow, R. W., and Noor, A. M. (2018). Geostatistical methods for disease mapping and visualization using data from spatio-temporally referenced prevalence surveys. *International Statistical Review*.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Golding, N., Burstein, R., Longbottom, J., Browne, A., Fullman, N., Osgood-Zimmerman, A., Earl, L., Bhatt, S., Cameron, E., Casey, D., Dwyer-Lindgren, L., Farag, T., Flaxman, A., Fraser, M., Gething, P., Gibson, H., Graetz, N., Krause, L., Kulikoff, X., Lim, S., Mappin, B., Morozoff, C., Reiner, R., Sligar, A., Smith, D., Wang, H., Weiss, D., Murray, C., Moyes, C., and Hay, S. (2017). Mapping under-5 and neontal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *The Lancet*, 390:2171–2182.

Graetz, N., Friedman, J., Osgood-Zimmerman, A., Burstein, R., Biehl, M. H., Shields, C., Mosser, J. F., Casey, D. C., Deshpande, A., Earl, L., Reiner, R., Ray, S., Fullman, N., Levine, A., Stubbs, R., Mayala, B., Longbottom, J., Browne, A., Bhatt, S., Weiss, D., Gething, P., Mokdad, A., Lim, S., Murray, C., Gakidou, E., and Hay, S. (2018). Mapping local variation in educational attainment across Africa. *Nature*, 555:48.

Gramacy, R. B. and Apley, D. W. (2015). Local Gaussian process approximation for large computer experiments. *Journal of Computational and Graphical Statistics*, 24:561–578.

Guhaniyogi, R. and Banerjee, S. (2018). Meta-kriging: Scalable Bayesian modeling and inference for massive spatial datasets. *Technometrics*, 60:430–444.

Guinness, J. (2019). Spectral density estimation for random fields via periodic embeddings. *Biometrika*, 106:267–286.

Guinness, J. and Fuentes, M. (2017). Circulant embedding of approximate covariances for inference from Gaussian data on large lattices. *Journal of computational and Graphical Statistics*, 26:88–97.

Handcock, M. and Stein, M. (1993). A Bayesian analysis of kriging. *Technometrics*, 35:403–410.

Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M., et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24:398–425.

Hirano, T. and Yajima, Y. (2013). Covariance tapering for prediction of large spatial data sets in transformed random fields. *Annals of the Institute of Statistical Mathematics*, 65:913–939.

Horvitz, D. and Thompson, D. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47:663–685.

Hyman, J. M. (1983). Accurate monotonicity preserving cubic interpolation. *SIAM Journal on Scientific and Statistical Computing*, 4:645–654.

ICF International (2012). *Demographic and Health Survey Sampling and Household Listing Manual*. Calverton, Maryland, USA: ICF International.

Jerven, M. (2013). *Poor numbers: how we are misled by African development statistics and what to do about it*. Cornell University Press.

Katzfuss, M. (2017). A multi-resolution approximation for massive spatial datasets. *Journal of the American Statistical Association*, 112:201–214.

Katzfuss, M. and Guinness, J. (2020). A general framework for Vecchia approximations of Gaussian processes. *Statistical Science*. To appear.

Katzfuss, M., Guinness, J., Gong, W., and Zilber, D. (2018). Vecchia approximations of Gaussian-process predictions. *arXiv preprint arXiv:1805.03309*.

Katzfuss, M. and Hammerling, D. (2017). Parallel inference for massive distributed spatial data using low-rank models. *Statistics and Computing*, 27:363–375.

Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, and National Council For Population And Development/Kenya (2009). *The 2009 Kenya Population and Housing Census Volume IC: Population Distribution by Age, Sex, and Administrative Units*. Nairobi: Kenya National Bureau of Statistics.

Kenya National Bureau of Statistics, Ministry of Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, and National Council For Population And Development/Kenya (2015). *Kenya Demographic and Health Survey 2014*. Rockville, Maryland, USA.

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

Konomi, B. A., Sang, H., and Mallick, B. K. (2014). Adaptive Bayesian nonstationary modeling for large spatial datasets using covariance approximations. *Journal of Computational and Graphical Statistics*, 23:802–829.

Lehmann, E. L. and Romano, J. P. (2005). *Testing Statistical Hypotheses*. Springer Texts in Statistics, 3rd. ed. edition.

Li, Z. R., Hsiao, Y., Godwin, J., Martin, B. D., Wakefield, J., and Clark, S. J. (2019). Changes in the spatial distribution of the under five mortality rate: small-area analysis of 122 DHS surveys in 262 subregions of 35 countries in Africa. *PLoS one*, 14. Published January 22, 2019.

Lindgren, F. and Rue, H. (2015). Bayesian spatial modelling with R-INLA. *Journal of Statistical Software*, 63.

Lindgren, F., Rue, H., and Lindström, J. (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic differential equation approach (with discussion). *Journal of the Royal Statistical Society, Series B*, 73:423–498.

Lohr, S. (2010). *Sampling: Design and Analysis, Second Edition*. Brooks/Cole Cengage Learning, Boston.

Lumley, T. (2004). Analysis of complex survey samples. *Journal of Statistical Software*, 9:1–19.

Lumley, T. (2018). survey: analysis of complex survey samples. R package version 3.35.

Marhuenda, Y., Molina, I., and Morales, D. (2013). Small area estimation with spatio-temporal fay–herriot models. *Computational Statistics & Data Analysis*, 58:308–325.

Martin, B. D., Li, Z. R., Hsiao, Y., Godwin, J., Wakefield, J., and Clark, S. J. (2018). *SUMMER: Spatio-Temporal Under-Five Mortality Methods for Estimation*. R package version 0.2.1.

McCullagh, P. and Nelder, J. (1989). *Generalized Linear Models, Second Edition*. Chapman and Hall, London.

Mercer, L., Wakefield, J., Pantazis, A., Lutambi, A., Mosanja, H., and Clark, S. (2015). Small area estimation of childhood mortality in the absence of vital registration. *Annals of Applied Statistics*, 9:1889–1905.

Midzuno, H. (1951). On the sampling system with probability proportionate to sum of sizes. *Annals of the Institute of Statistical Mathematics*, 3:99–107.

Neal, S., Ruktanonchai, C. W., Chandra-Mouli, V., Harvey, C., Matthews, Z., Raina, N., and Tatem, A. (2019). Using geospatial modelling to estimate the prevalence of adolescent first births in Nepal. *BMJ Global Health*, 4.

Neelon, B., Gelfand, A. E., and Miranda, M. L. (2014). A multivariate spatial mixture model for areal data: examining regional differences in standardized test scores. *Journal of the Royal Statistical Society: Series C*, 63:737–761.

Nychka, D., Bandyopadhyay, S., Hammerling, D., Lindgren, F., and Sain, S. (2015). A multiresolution Gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24:579–599.

Nychka, D., Hammerling, D., Sain, S., and Lenssen, N. (2016). LatticeKrig: Multiresolution kriging based on Markov random fields. R package version 6.4.

Osgood-Zimmerman, A., Millear, A. I., Stubbs, R. W., Shields, C., Pickering, B. V., Earl, L., Graetz, N., Kinyoki, D. K., Ray, S. E., Bhatt, S., Browne, A., Burstein, R., Cameron, E., Casey, D., Deshpande, A., Fullman, N., Gething, P., Gibson, H., Henry, N., Herrero, M., Krause, L., Letourneau, I., Levine, A., Liu, P., Longbottom, J., Mayala, B., Mosser, J., Noor, A., Pigott, D., Piwoz, E., Rao, P., Rawat, R., Reiner, R., Smith, D., Weiss, D., Wiens, K., Mokdad, A., S.S., L., Murray, C., Kassebaum, N., and Hay, S. (2018). Mapping child growth failure in Africa between 2000 and 2015. *Nature*, 555:41.

Paige, J., Fuglstad, G.-A., Riebler, A., and Wakefield, J. (2020). Design- and model-based approaches to small-area estimation in a low and middle income country context: Comparisons and recommendations. *Journal of Survey Statistics and Methodology*.

Piessens, R., de Doncker-Kapenga, E., Überhuber, C. W., and Kahaner, D. K. (2012). *Quadpack: a subroutine package for automatic integration*, volume 1. Springer Science & Business Media.

Porter, A. T., Holan, S. H., Wikle, C. K., and Cressie, N. (2014). Spatial Fay–Herriot models for small area estimation with functional covariates. *Spatial Statistics*, 10:27–42.

Rabe-Hesketh, S. and Skrondal, A. (2006). Multilevel modelling of complex survey data. *Journal of the Royal Statistical Society, Series A*, 169:805–827.

Rao, J. and Molina, I. (2015). *Small Area Estimation, Second Edition*. John Wiley, New York.

Riebler, A., Sørbye, S., Simpson, D., and Rue, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research*, 25:1145–1165.

Rue, H. and Follestad, T. (2001). GMRFLib: a C-library for fast and exact simulation of Gaussian Markov random fields. Technical report, SIS-2002-236.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL.

Rue, H., Martino, S., and Chopin, N. (2009a). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B*, 71:319–392.

Rue, H., Martino, S., and Chopin, N. (2009b). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71:319–392.

Rue, H., Riebler, A., Sørbye, S. H., Illian, J. B., Simpson, D. P., and Lindgren, F. K. (2017). Bayesian computing with INLA: a review. *Annual Review of Statistics and Its Application*, 4:395–421.

Sandefur, J. and Glassman, A. (2015). The political economy of bad data: Evidence from African survey and administrative statistics. *The Journal of Development Studies*, 51:116–132.

Sang, H. and Huang, J. Z. (2012). A full scale approximation of covariance functions for large spatial data sets. *Journal of the Royal Statistical Society: Series B*, 74:111–132.

Simpson, D., Rue, H., Riebler, A., Martins, T., and Sørbye, S. (2017). Penalising model component complexity: A principled, practical approach to constructing priors (with discussion). *Statistical Science*, 32:1–28.

Sjöstedt-de Luna, S. and Young, A. (2003). The bootstrap and kriging prediction intervals. *Scandinavian Journal of Statistics*, 30:175–192.

Skrondal, A. and Rabe-Hesketh, S. (2009). Prediction in multilevel generalized linear models. *Journal of the Royal Statistical Society, Series A*, 172:659–687.

Stein, M. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. Springer.

Stevens, F. R., Gaughan, A. E., Linard, C., and Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PloS One*, 10:e0107042.

Tatem, A. J. (2017). WorldPop, open data for spatial demography. *Scientific Data*, 4.

The World Bank (2019). Living standards measurement study (LSMS) — surveyunit. http://surveys.worldbank.org/lsms.

Thomson, D. R., Rhoda, D. A., Tatem, A. J., and Castro, M. C. (2020). Gridded population survey sampling: A systematic scoping review of the field and strategic research agenda. *International Journal of Health Geographics*, 19:1–16.

Tuckwell, H. C. (2018). *Elementary Applications of Probability Theory*. Routledge.

UNICEF - Statistics and Monitoring (2012). Multiple Indicator Cluster Surveys (MICS). http://www.unicef.org/statistics/index_24302.html.

United Nations (2020). *Sustainable Development Goals*. http://sustainabledevelopment.un.org/owg.html.

USAID (2019). *Demographic and Health Surveys*. United States Agency for International Development, http://www.dhsprogram.com.

Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Nilsen, K., Takahashi, S., Metcalf, C. J. E., Lessler, J., and Tatem, A. J. (2018a). A spatial regression model for the disaggregation of areal unit based data to high resolution grids with application to vaccination coverage mapping. *Statistical Methods in Medical Research*, pages 1–16.

Utazi, C. E., Thorley, J., Alegana, V. A., Ferrari, M. J., Takahashi, S., Metcalf, C. J. E., Lessler, J., and Tatem, A. J. (2018b). High resolution age-structured mapping of childhood vaccination coverage in low and middle income countries. *Vaccine*, 36:1583–1591.

Van de Poel, E., O'Donnell, O., and Van Doorslaer, E. (2007). Are urban children really healthier? evidence from 47 developing countries. *Social Science & Medicine*, 65:1986–2003.

Vandendijck, Y., Faes, C., Kirby, R. S., Lawson, A., and Hens, N. (2016). Model-based inference for small area estimation with sampling weights. *Spatial Statistics*, 18:455–473.

Wagner, Z., Heft-Neal, S., Bhutta, Z. A., Black, R. E., Burke, M., and Bendavid, E. (2018). Armed conflict and child mortality in Africa: a geospatial analysis. *The Lancet*.

Wakefield, J. (2013). *Bayesian and Frequentist Regression Methods*. Springer, New York.

Wakefield, J., Fuglstad, G.-A., Riebler, A., Godwin, J., Wilson, K., and Clark, S. (2019). Estimating under five mortality in space and time in a developing world context. *Statistical Methods in Medical Research*, 28:2614–2634.

Wardrop, N., Jochem, W., Bird, T., Chamberlain, H., Clarke, D., Kerr, D., Bengtsson, L., Juran, S., Seaman, V., and Tatem, A. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences*, 115:3529–3537.

Watjou, K., Faes, C., Lawson, A., Kirby, R., Aregay, M., Carroll, R., and Vandendijck, Y. (2017). Spatial small area smoothing models for handling survey data with nonresponse. *Statistics in Medicine*, 36:3708–3745.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, 4:389–396.

Wilson, K. and Wakefield, J. (2018). Child mortality estimation incorporating summary birth history data. *Under revision, available at arXiv*.

You, Y. and Zhou, Q. M. (2011). Hierarchical Bayes small area estimation under a spatial model with application to health survey data. *Survey Methodology*, 37:25–37.

Zilber, D. and Katzfuss, M. (2019). Vecchia-Laplace approximations of generalized Gaussian processes for big non-Gaussian spatial data. *arXiv preprint arXiv:1906.07828*.

Appendix A

APPENDIX OF CHAPTER 3

A.1 Additional Results: Application to Secondary Education in Kenya

Table A.1: Predicted secondary education prevalences and 80% credible intervals for each of the 47 counties in Kenya for women aged 20–29 in 2014.

	Smoothed Direct			***BYM2$_{UC}$***			***SPDE$_{UC}$***		
County	Est	Q10	Q90	Est	Q10	Q90	Est	Q10	Q90
Baringo	0.250	0.202	0.304	0.240	0.203	0.282	0.250	0.218	0.283
Bomet	0.276	0.240	0.316	0.284	0.245	0.326	0.267	0.236	0.300
Bungoma	0.231	0.186	0.283	0.287	0.249	0.330	0.259	0.232	0.286
Busia	0.193	0.156	0.237	0.182	0.150	0.218	0.176	0.150	0.203
Elgeyo Marakwet	0.403	0.348	0.460	0.357	0.313	0.406	0.282	0.251	0.313
Embu	0.333	0.292	0.376	0.346	0.298	0.395	0.295	0.262	0.330
Garissa	0.125	0.093	0.166	0.146	0.114	0.183	0.110	0.077	0.147
Homa Bay	0.167	0.142	0.196	0.184	0.152	0.217	0.168	0.146	0.192
Isiolo	0.147	0.113	0.189	0.155	0.126	0.189	0.126	0.105	0.147
Kajiado	0.403	0.358	0.449	0.364	0.322	0.408	0.304	0.269	0.342
Kakamega	0.264	0.227	0.303	0.277	0.238	0.317	0.244	0.216	0.274
Kericho	0.341	0.294	0.392	0.352	0.309	0.398	0.322	0.290	0.354
Kiambu	0.538	0.495	0.580	0.529	0.482	0.578	0.489	0.450	0.525
Kilifi	0.284	0.247	0.324	0.281	0.244	0.325	0.224	0.192	0.258
Kirinyaga	0.304	0.259	0.354	0.341	0.294	0.390	0.361	0.324	0.399
Kisii	0.353	0.316	0.392	0.346	0.305	0.388	0.283	0.254	0.312
Kisumu	0.332	0.281	0.387	0.311	0.270	0.357	0.286	0.255	0.317
Kitui	0.224	0.176	0.280	0.195	0.163	0.233	0.156	0.129	0.185
Kwale	0.211	0.163	0.268	0.201	0.166	0.240	0.178	0.150	0.205
Laikipia	0.323	0.261	0.391	0.324	0.277	0.374	0.258	0.226	0.292

Table A.1: Predicted secondary education prevalences and 80% credible intervals for each of the 47 counties in Kenya for women aged 20–29 in 2014. *(continued)*

	Smoothed Direct			***BYM2$_{UC}$***			***SPDE$_{UC}$***		
County	Est	Q10	Q90	Est	Q10	Q90	Est	Q10	Q90
Lamu	0.169	0.128	0.219	0.172	0.139	0.211	0.151	0.119	0.185
Machakos	0.309	0.262	0.360	0.338	0.296	0.382	0.321	0.285	0.358
Makueni	0.309	0.263	0.360	0.294	0.253	0.338	0.259	0.224	0.296
Mandera	0.088	0.061	0.125	0.085	0.060	0.117	0.081	0.058	0.107
Marsabit	0.134	0.099	0.180	0.125	0.099	0.157	0.112	0.088	0.136
Meru	0.273	0.229	0.320	0.258	0.223	0.298	0.220	0.191	0.249
Migori	0.152	0.119	0.191	0.173	0.143	0.209	0.155	0.131	0.180
Mombasa	0.394	0.356	0.433	0.416	0.367	0.466	0.428	0.387	0.468
Murang'a	0.391	0.336	0.449	0.364	0.320	0.413	0.373	0.333	0.412
Nairobi	0.536	0.490	0.582	0.533	0.496	0.571	0.545	0.513	0.577
Nakuru	0.405	0.353	0.458	0.408	0.366	0.451	0.382	0.346	0.418
Nandi	0.324	0.279	0.372	0.323	0.284	0.364	0.271	0.243	0.300
Narok	0.222	0.182	0.268	0.199	0.166	0.237	0.201	0.176	0.224
Nyamira	0.398	0.352	0.446	0.358	0.313	0.403	0.311	0.278	0.344
Nyandarua	0.367	0.318	0.418	0.356	0.311	0.403	0.338	0.303	0.374
Nyeri	0.475	0.419	0.532	0.475	0.427	0.525	0.457	0.416	0.497
Samburu	0.143	0.113	0.180	0.128	0.102	0.158	0.102	0.078	0.129
Siaya	0.182	0.150	0.220	0.201	0.167	0.239	0.179	0.153	0.207
Taita Taveta	0.352	0.298	0.411	0.321	0.274	0.371	0.286	0.240	0.334
Tana River	0.127	0.101	0.158	0.114	0.090	0.144	0.097	0.078	0.115
Tharaka-Nithi	0.303	0.244	0.368	0.295	0.248	0.348	0.276	0.243	0.310
Trans-Nzoia	0.243	0.204	0.288	0.281	0.240	0.328	0.228	0.200	0.257
Turkana	0.150	0.100	0.218	0.102	0.077	0.134	0.069	0.049	0.092
Uasin Gishu	0.357	0.310	0.407	0.366	0.327	0.408	0.362	0.331	0.391
Vihiga	0.228	0.182	0.281	0.254	0.213	0.301	0.251	0.219	0.284
Wajir	0.089	0.058	0.134	0.113	0.087	0.145	0.088	0.068	0.110
West Pokot	0.147	0.116	0.186	0.140	0.110	0.173	0.123	0.100	0.148

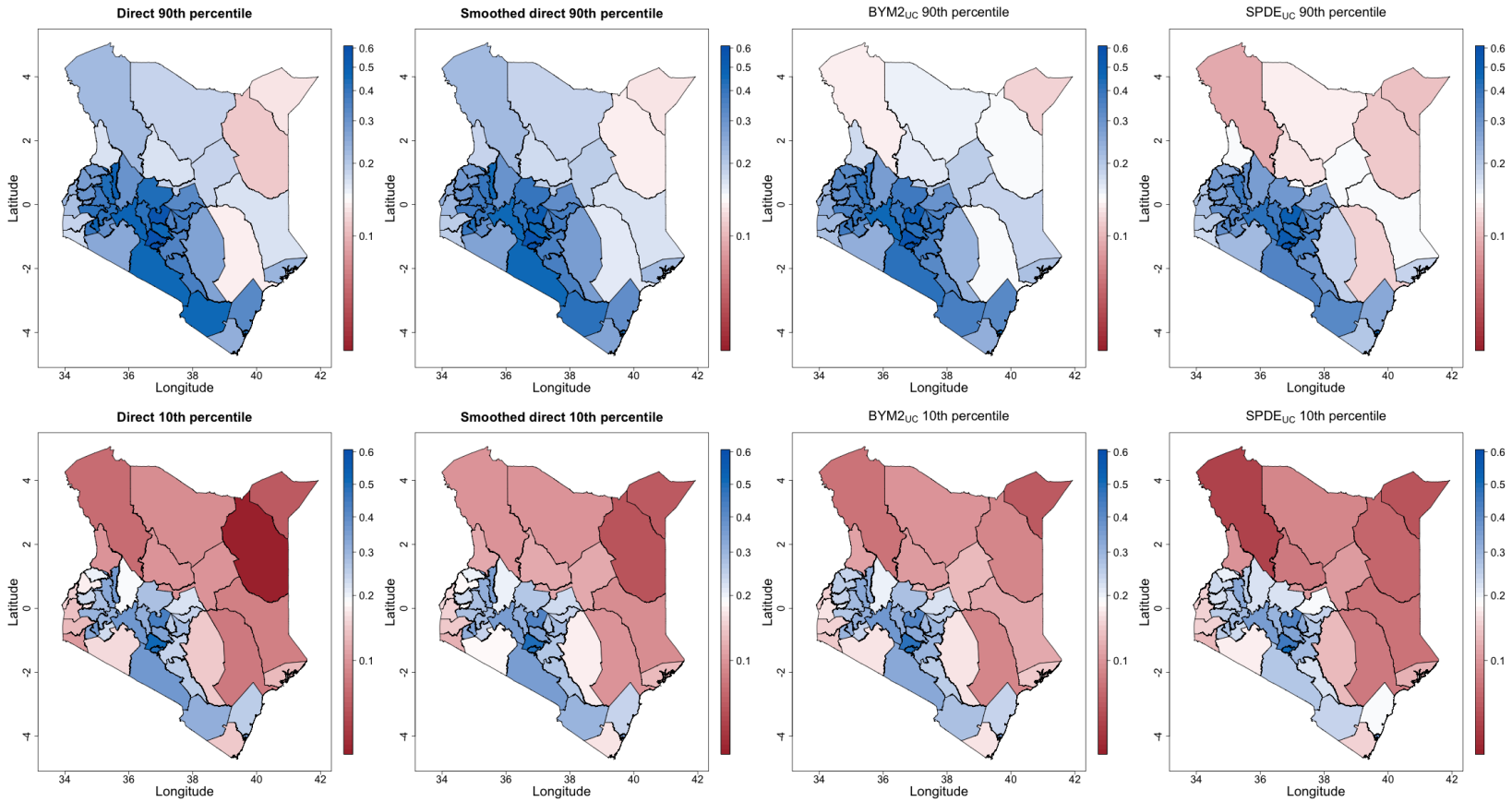


Figure A.1: Kenya county level 2014 secondary education 90th (top) and 10th (bottom) quantiles for women aged 20–29. The BYM2 and SPDE model predictions include both urban and cluster effects.

A.2 Additional Results: Application to NMRs in Kenya

Table A.2: Predicted neonatal mortality rates (NMRs) and 80% credible intervals for each of the 47 counties in Kenya.

	Smoothed Direct			***BYM2$_{UC}$***			***SPDE$_{UC}$***		
County	Est	Q10	Q90	Est	Q10	Q90	Est	Q10	Q90
Baringo	0.0194	0.0149	0.0242	0.0219	0.0172	0.0282	0.0206	0.0171	0.0241
Bomet	0.0224	0.0181	0.0281	0.0262	0.0204	0.0339	0.0234	0.0195	0.0275
Bungoma	0.0177	0.0128	0.0230	0.0179	0.0135	0.0235	0.0169	0.0140	0.0199
Busia	0.0207	0.0160	0.0262	0.0223	0.0168	0.0290	0.0176	0.0144	0.0211
Elgeyo Marakwet	0.0189	0.0142	0.0241	0.0210	0.0160	0.0272	0.0193	0.0159	0.0228
Embu	0.0209	0.0163	0.0264	0.0237	0.0181	0.0309	0.0221	0.0181	0.0264
Garissa	0.0243	0.0195	0.0313	0.0265	0.0207	0.0342	0.0227	0.0188	0.0273
Homa Bay	0.0212	0.0168	0.0264	0.0228	0.0180	0.0288	0.0221	0.0187	0.0256
Isiolo	0.0203	0.0159	0.0251	0.0236	0.0186	0.0300	0.0200	0.0168	0.0234
Kajiado	0.0191	0.0143	0.0241	0.0217	0.0165	0.0275	0.0219	0.0186	0.0254
Kakamega	0.0176	0.0128	0.0228	0.0178	0.0131	0.0236	0.0176	0.0147	0.0208
Kericho	0.0211	0.0170	0.0259	0.0244	0.0192	0.0309	0.0245	0.0207	0.0284
Kiambu	0.0227	0.0179	0.0292	0.0262	0.0203	0.0339	0.0257	0.0214	0.0303
Kilifi	0.0192	0.0146	0.0242	0.0218	0.0166	0.0286	0.0203	0.0164	0.0243
Kirinyaga	0.0210	0.0164	0.0265	0.0238	0.0180	0.0311	0.0230	0.0189	0.0277
Kisii	0.0196	0.0148	0.0248	0.0209	0.0159	0.0276	0.0223	0.0188	0.0260
Kisumu	0.0194	0.0145	0.0248	0.0213	0.0161	0.0276	0.0217	0.0183	0.0253
Kitui	0.0217	0.0173	0.0271	0.0235	0.0182	0.0301	0.0210	0.0173	0.0252
Kwale	0.0239	0.0182	0.0326	0.0250	0.0189	0.0328	0.0209	0.0170	0.0252
Laikipia	0.0204	0.0160	0.0254	0.0217	0.0167	0.0280	0.0207	0.0172	0.0245
Lamu	0.0246	0.0187	0.0339	0.0286	0.0217	0.0385	0.0244	0.0194	0.0301
Machakos	0.0223	0.0179	0.0280	0.0284	0.0221	0.0368	0.0235	0.0195	0.0277
Makueni	0.0206	0.0156	0.0265	0.0206	0.0151	0.0270	0.0203	0.0165	0.0244
Mandera	0.0183	0.0124	0.0252	0.0148	0.0097	0.0218	0.0184	0.0149	0.0221
Marsabit	0.0205	0.0160	0.0258	0.0222	0.0167	0.0286	0.0203	0.0172	0.0235
Meru	0.0234	0.0191	0.0296	0.0248	0.0192	0.0321	0.0199	0.0159	0.0240
Migori	0.0252	0.0205	0.0318	0.0292	0.0228	0.0375	0.0227	0.0187	0.0267
Mombasa	0.0199	0.0146	0.0262	0.0219	0.0158	0.0300	0.0217	0.0169	0.0270
Murang'a	0.0230	0.0183	0.0295	0.0257	0.0196	0.0340	0.0249	0.0206	0.0300

Table A.2: Predicted neonatal mortality rates (NMRs) and 80% credible intervals for each of the 47 counties in Kenya. (*continued*)

	Smoothed Direct			***BYM2$_{UC}$***			***SPDE$_{UC}$***		
County	Est	Q10	Q90	Est	Q10	Q90	Est	Q10	Q90
Nairobi	0.0240	0.0194	0.0307	0.0305	0.0234	0.0397	0.0258	0.0210	0.0310
Nakuru	0.0258	0.0208	0.0333	0.0277	0.0219	0.0352	0.0267	0.0224	0.0315
Nandi	0.0199	0.0156	0.0249	0.0215	0.0166	0.0280	0.0201	0.0168	0.0236
Narok	0.0202	0.0156	0.0253	0.0209	0.0161	0.0265	0.0223	0.0191	0.0257
Nyamira	0.0224	0.0179	0.0284	0.0284	0.0215	0.0373	0.0231	0.0197	0.0268
Nyandarua	0.0265	0.0209	0.0349	0.0314	0.0238	0.0423	0.0257	0.0215	0.0301
Nyeri	0.0197	0.0151	0.0248	0.0236	0.0183	0.0310	0.0237	0.0192	0.0283
Samburu	0.0179	0.0131	0.0230	0.0188	0.0141	0.0250	0.0179	0.0146	0.0217
Siaya	0.0178	0.0128	0.0233	0.0173	0.0124	0.0230	0.0188	0.0155	0.0223
Taita Taveta	0.0202	0.0153	0.0257	0.0214	0.0158	0.0282	0.0191	0.0149	0.0234
Tana River	0.0221	0.0178	0.0278	0.0262	0.0210	0.0329	0.0240	0.0203	0.0278
Tharaka-Nithi	0.0205	0.0158	0.0261	0.0231	0.0174	0.0311	0.0212	0.0170	0.0257
Trans-Nzoia	0.0179	0.0134	0.0229	0.0200	0.0152	0.0257	0.0179	0.0146	0.0214
Turkana	0.0210	0.0164	0.0266	0.0254	0.0199	0.0325	0.0215	0.0181	0.0254
Uasin Gishu	0.0183	0.0139	0.0231	0.0203	0.0154	0.0260	0.0196	0.0164	0.0230
Vihiga	0.0207	0.0161	0.0262	0.0239	0.0180	0.0319	0.0194	0.0161	0.0231
Wajir	0.0212	0.0171	0.0263	0.0215	0.0167	0.0274	0.0202	0.0167	0.0241
West Pokot	0.0201	0.0155	0.0254	0.0219	0.0168	0.0287	0.0190	0.0156	0.0227

Table A.2: Predicted neonatal mortality rates (NMRs) and 80% credible intervals for each of the 47 counties in Kenya. *(continued)*

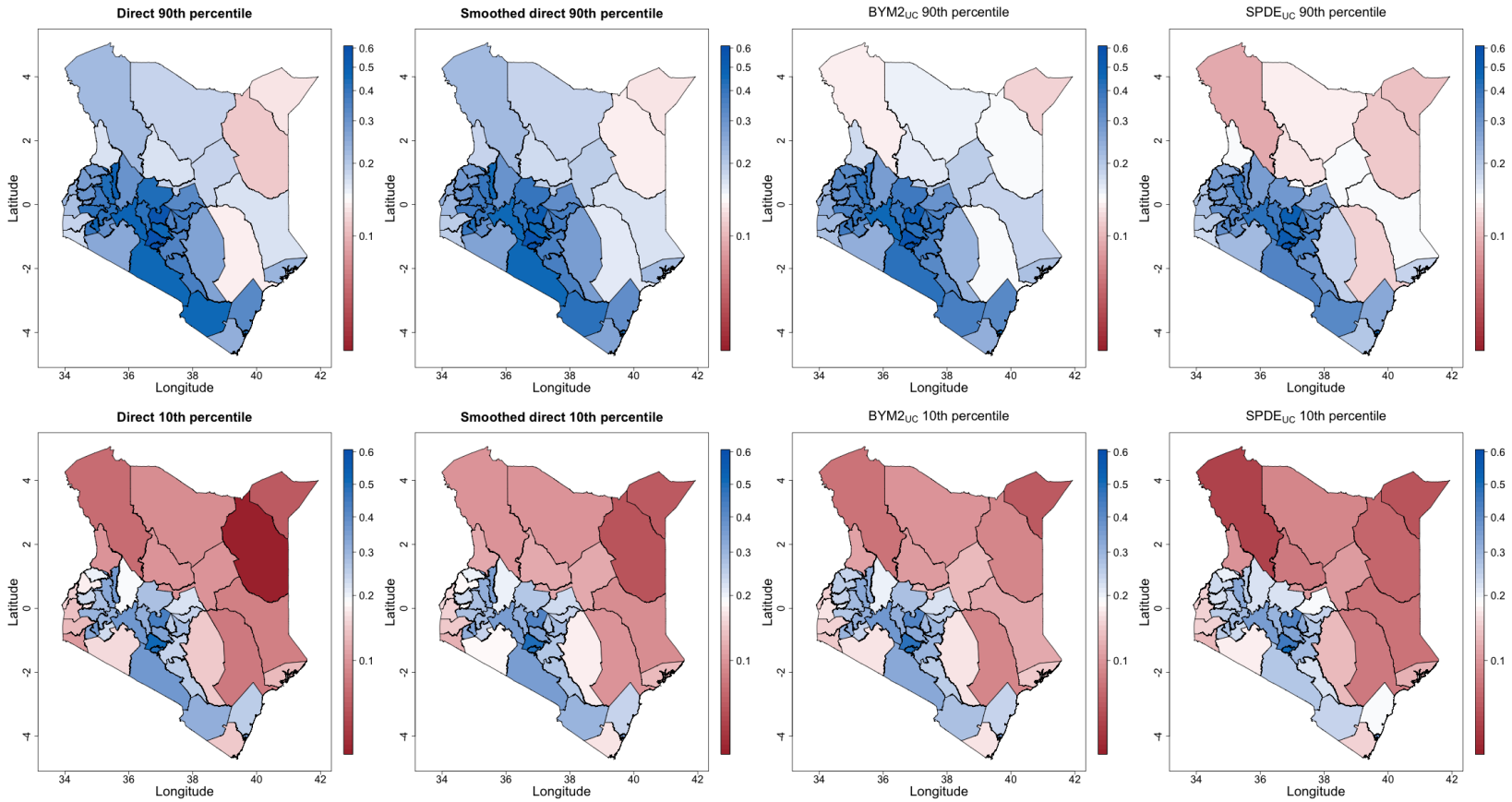


Figure A.2: Kenya county level neonatal mortality rate 90th (top) and 10th (bottom) quantiles from 2010-2014. The BYM2 and SPDE model predictions include both urban and cluster effects.

A.3 Simulation Details

A.3.1 Sampling

Since the number of enumeration areas selected within each of the 92 strata is fixed based on the empirical number of clusters in that strata for the 2014 DHS, it is sufficient to consider the sampling weights within any given stratum. Let H_c be the number of households in cluster c, and let N_c be the number of children that we sampled within that cluster. Since we use probability proportional to size (PPS) sampling or simple random sampling (SRS) to sample the clusters for the Stratified and Unstratified sampling designs, respectively, the probability of picking any given cluster is chosen to be either equal in the representative sampling case or proportional to the number of households within it in the PPS case. Given that n clusters are sampled from a given strata with N EAs in total, the probability of including cluster c is either:

$$p_c = \frac{nH_c}{\sum_c H_c}$$

in the PPS case or

$$p_c = \frac{n}{N}$$

in the SRS case. In the PPS case, we use Midzuno's method (Midzuno, 1951; Deville and Tille, 1998) for sampling without replacement in such a way as to match these inclusion probabilities, which in our case are all well-defined. The probability of including any given child in the sample given that the cluster was sampled is just the probability of selecting the household he/she is living in. Exactly 25 households are selected in each cluster, and so this probability is,

$$p_{ck|c} = \frac{25}{H_c},$$

where k is the index of the child. Since a given child can only be selected if its respective cluster is selected, the probability of including any given child in our sample is either

$$p_{ck} = p_c \times p_{ck|c} = \frac{25n}{\sum_c H_c}$$

under the Stratified design or

$$p_{ck} = p_c \times p_{ck|c} = \frac{25n}{NH_c}$$

for the Unstratified design, where the sum is taken to be over all clusters within the considered stratum. This yields a sampling weight for each child of $\frac{\sum_c H_c}{25n}$ or $\frac{NH_c}{25n}$ in the Stratified or Unstratified sampling cases respectively, and these can be summed over all the children sampled within any given cluster to yield respective cluster sampling weights of $N_c \sum_c H_c/25n$ or $N_c N H_c/25n$.

A.3.2 Additional Simulation Study Plots

150km Spatial Range, 0.15^2 *Spatial Variance*

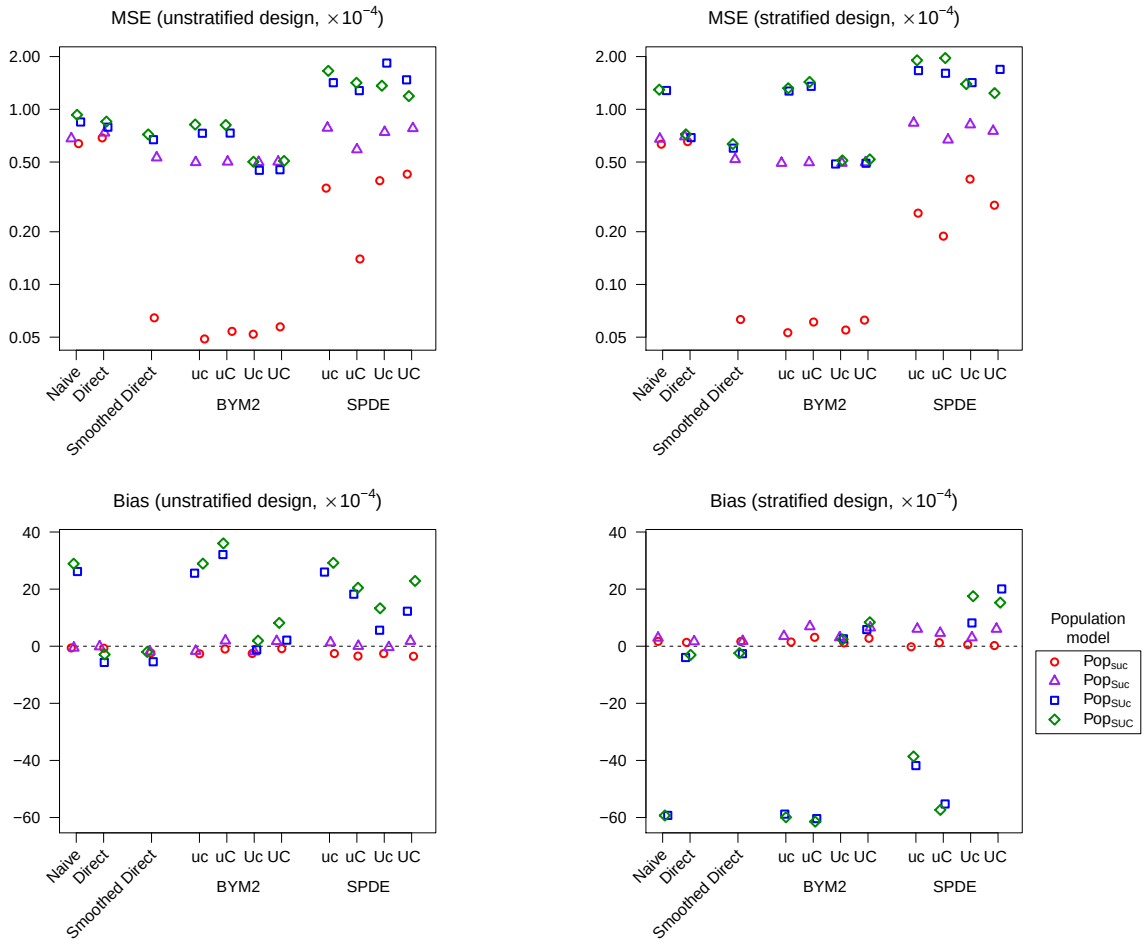


Figure A.3: County level scoring rules for each of the simulated populations and the main models considered with 150km spatial range and 0.15^2 spatial variance. Scoring rules for the simulated unstratified and stratified surveys are respectively in the left and right columns. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included respectively.

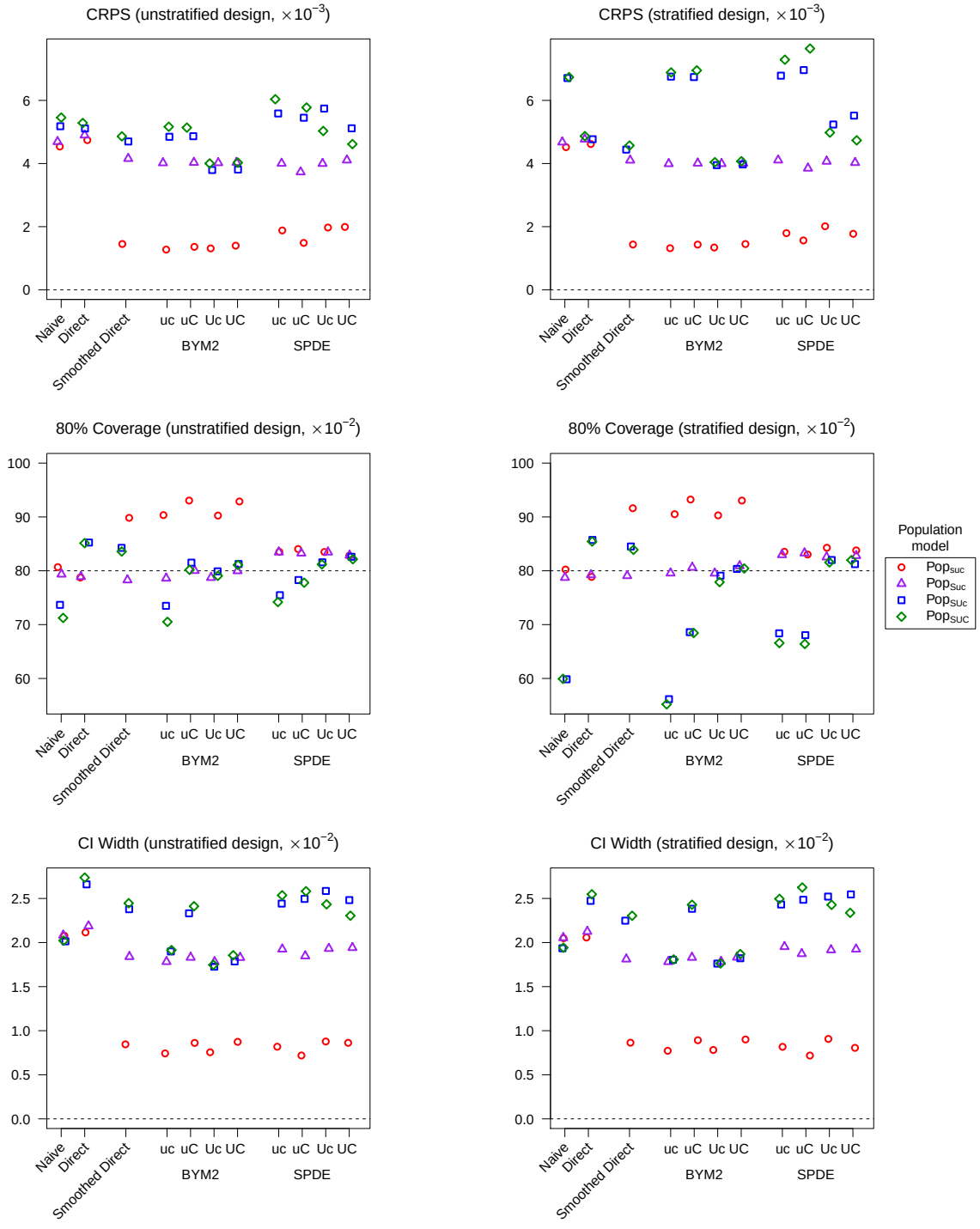


Figure A.4: County level scoring rules for each of the simulated populations and the main models considered with 150km spatial range and 0.15^2 spatial variance. Scoring rules for the simulated unstratified and stratified surveys are respectively in the left and right columns. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included respectively.

50km Spatial Range, 0.15^2 Spatial Variance

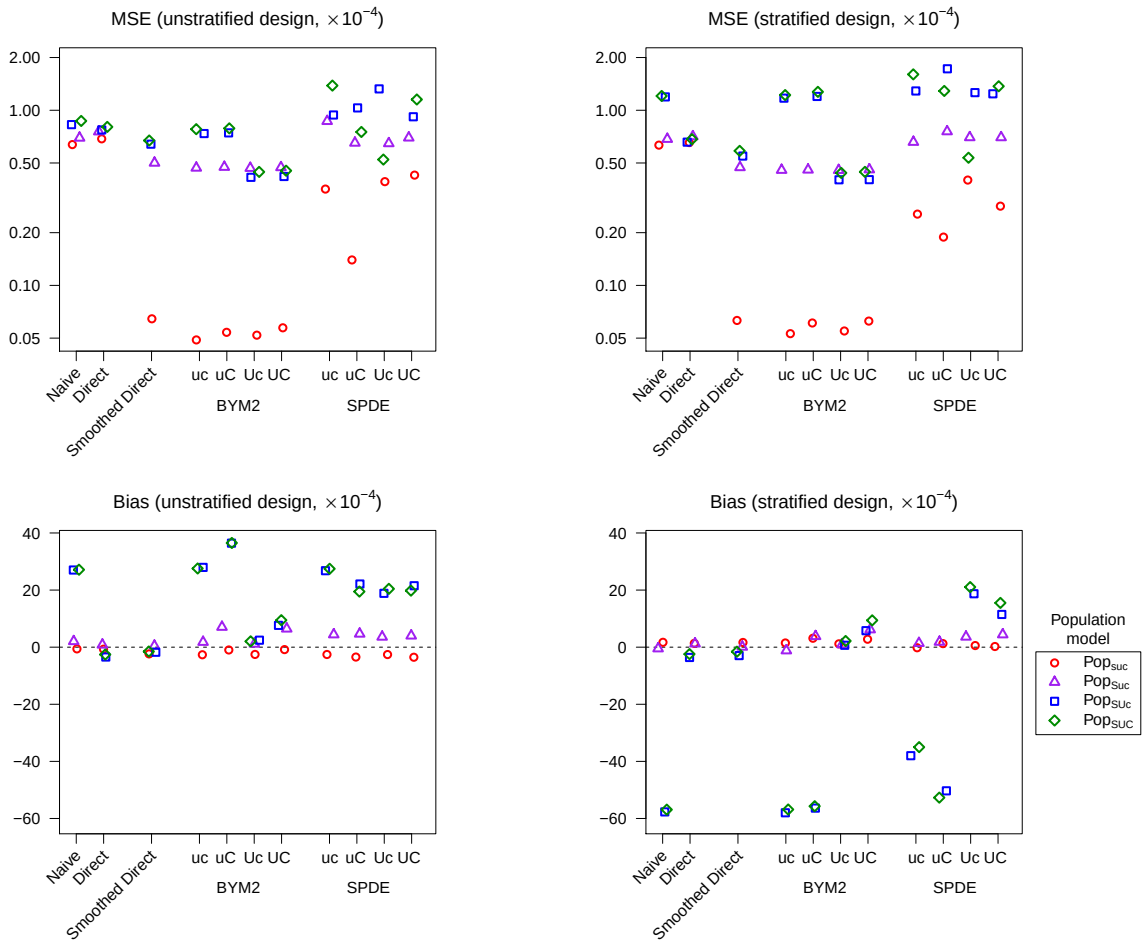


Figure A.5: County level scoring rules for each of the simulated populations and the main models considered with 50km spatial range and 0.15^2 spatial variance. Scoring rules for the simulated unstratified and stratified surveys are respectively in the left and right columns. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included respectively.

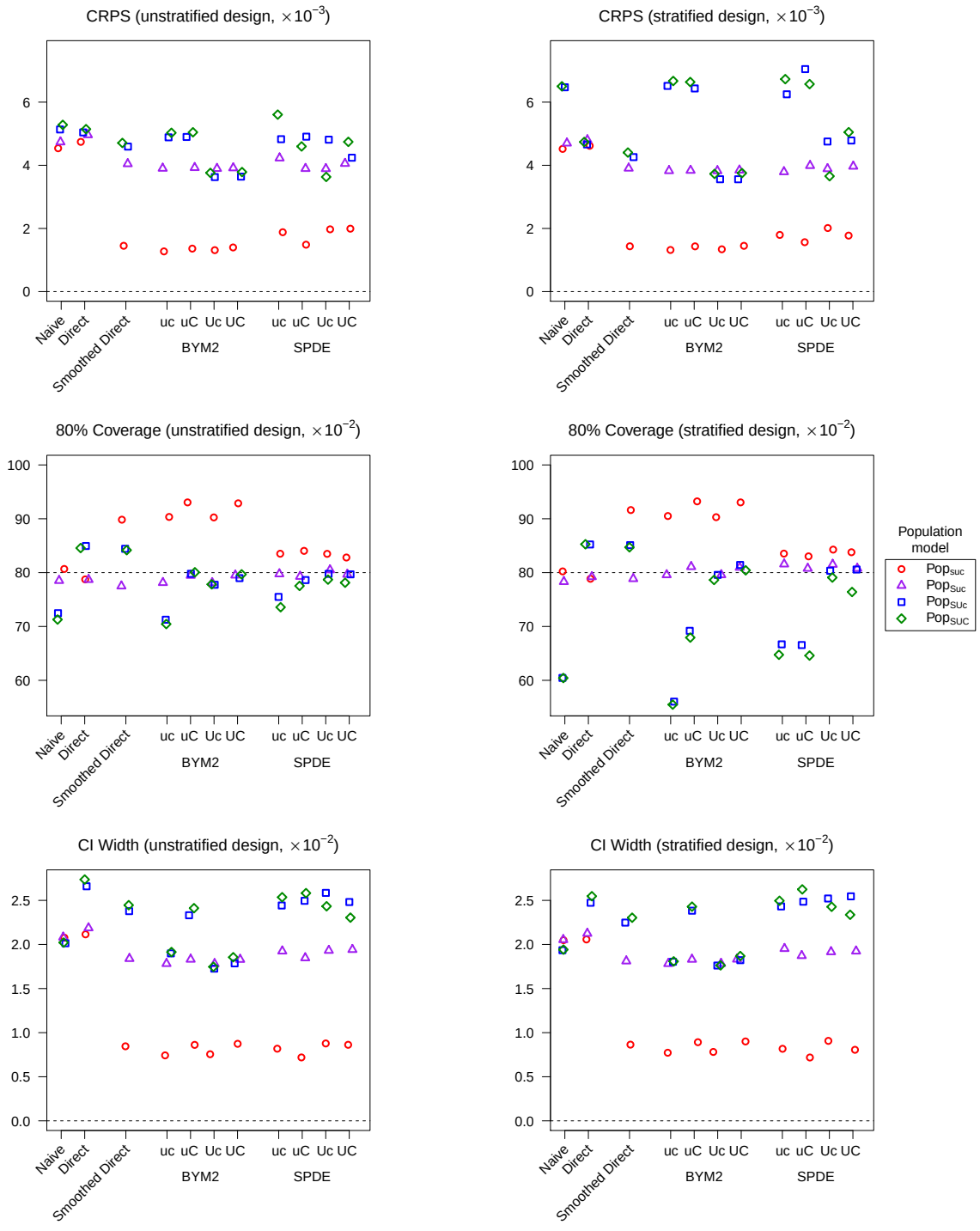


Figure A.6: County level scoring rules for each of the simulated populations and the main models considered with 50km spatial range and 0.15^2 spatial variance. Scoring rules for the simulated unstratified and stratified surveys are respectively in the left and right columns. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included respectively.

150km Spatial Range, 0.3^2 *Spatial Variance*

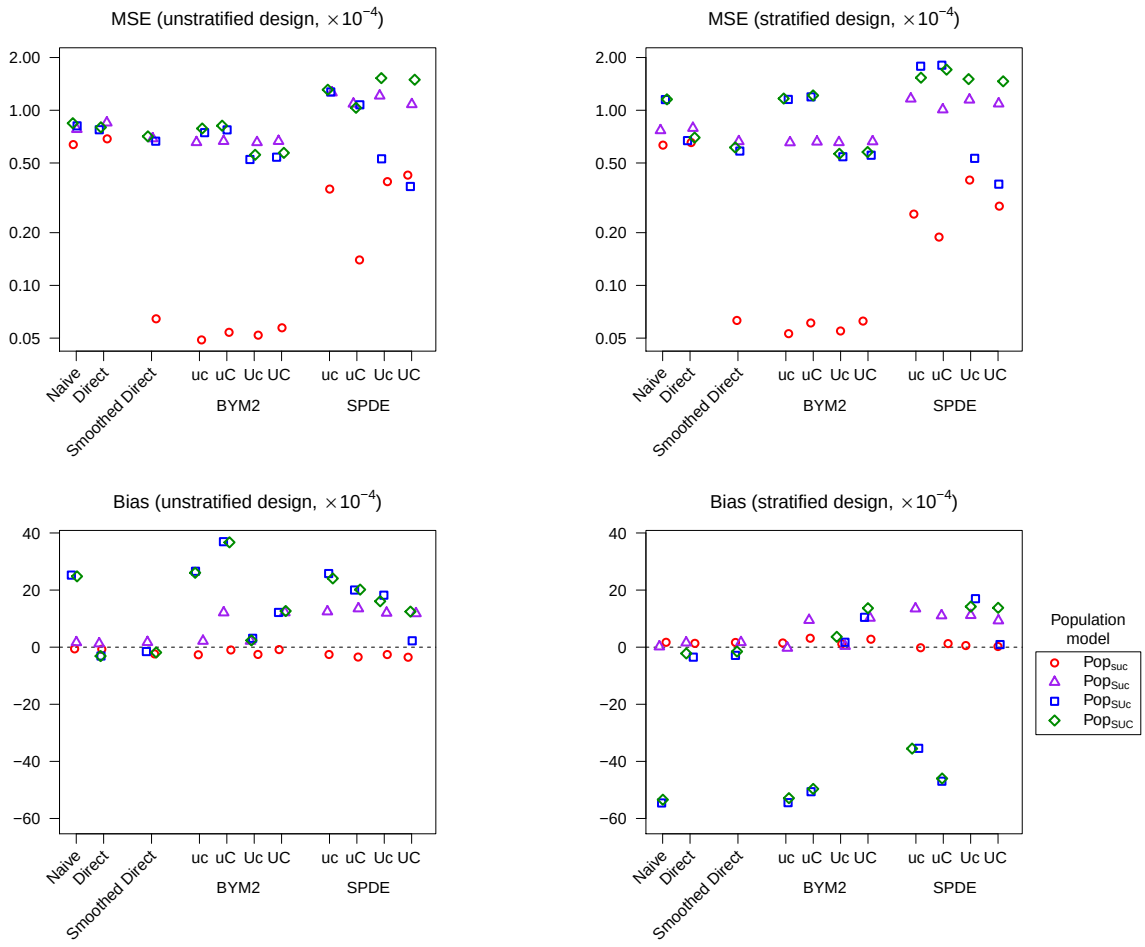


Figure A.7: County level scoring rules for each of the simulated populations and the main models considered with 150km spatial range and 0.3^2 spatial variance. Scoring rules for the simulated unstratified and stratified surveys are respectively in the left and right columns. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included respectively.

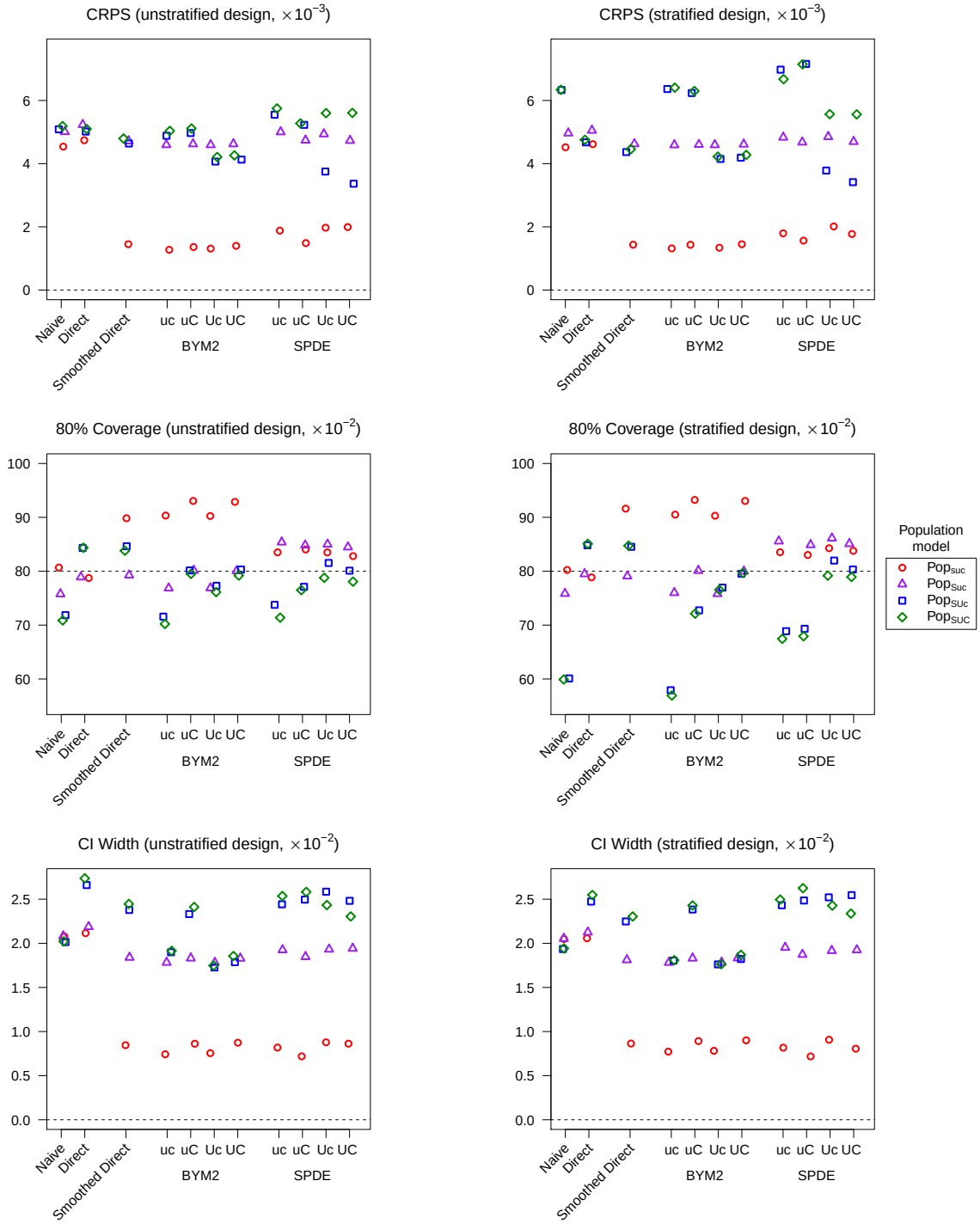


Figure A.8: County level scoring rules for each of the simulated populations and the main models considered with 150km spatial range and 0.3^2 spatial variance. Scoring rules for the simulated unstratified and stratified surveys are respectively in the left and right columns. The labels s/S, u/U, and c/C denote whether or not spatial, urban, and cluster effects are included respectively.

A.3.3 Simulation Result Tables

Unstratified Sampling Results

3.3.1.1 150km Spatial Range, 0.15^2 Spatial Variance

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-0.6	6.2	0.64	4.5	**81**	2.08
Direct						
	-0.7	*6.7*	*0.69*	*4.7*	**79**	*2.11*
Smoothed Direct						
	-2.4	0.48	0.06	1.4	90	0.85
BYM2						
uc	-2.6	**0.35**	**0.05**	**1.3**	90	0.74
uC	-1.0	0.40	**0.05**	1.4	*93*	0.86
Uc	-2.5	0.38	**0.05**	**1.3**	90	0.76
UC	-0.8	0.44	0.06	1.4	*93*	0.87
SPDE						
uc	-2.6	3.3	0.35	1.9	84	0.82
uC	*-3.5*	1.2	0.14	1.5	84	**0.72**
Uc	-2.6	3.7	0.39	2.0	84	0.88
UC	*-3.5*	4.0	0.43	2.0	83	0.86

Table A.3: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.4: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.755	0.01	-1.767	-1.754	-1.742
	Phi	0.278	0.238	0.032	0.207	0.653
	Tot. Var	0.00094	0.000744	0.00024	0.00074	0.00187
	Spatial Var	0.000237	0.000309	0.000017	0.000126	0.000599
	iid Var	0.0007	0.000639	0.00014	0.00052	0.00149
BYM2						
uc	Intercept	-1.755	0.01	-1.767	-1.767	-1.742
	Phi	0.283	0.241	0.033	0.21	0.663
	Tot. Var	0.001	0.001	0	0	0.001
	Spatial Var	0.000161	0.000233	0.000008	0.000076	0.000416
	iid Var	0.00046	0.000517	0.00006	0.00029	0.00107
	Tot. SD	0.0212	0.0109	0.0087	0.0196	0.0358
	Spatial SD	0.0099	0.00707	0.0026	0.0081	0.0196
	iid SD	0.0178	0.0102	0.0064	0.0158	0.0316
uC	Intercept	-1.756	0.01	-1.769	-1.769	-1.743
	Cluster Var	0.00233	0.00187	0.00048	0.00183	0.00483
	Phi	0.286	0.242	0.034	0.213	0.668
	Tot. Var	0.00063	0.000633	0.00011	0.00044	0.00138
	Spatial Var	0.000165	0.000238	0.000009	0.000079	0.000423
	iid Var	0.00047	0.000534	0.00006	0.00029	0.00108
	Cluster SD	0.0415	0.0188	0.0185	0.0396	0.067
	Tot. SD	0.0216	0.011	0.0091	0.0198	0.0361
	Spatial SD	0.0101	0.00711	0.0027	0.0084	0.0198
	iid SD	0.018	0.0103	0.0067	0.0161	0.0319
Uc	Intercept	-1.755	0.011	-1.769	-1.769	-1.741
	Urban	0.002	0.024	-0.029	-0.029	0.033
	Phi	0.283	0.242	0.033	0.21	0.663
	Tot. Var	0.001	0.001	0	0	0.001

Table A.4: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

			Est	SD	Q10	Q50	Q90
		Spatial Var	0.000163	0.000235	0.000008	0.000077	0.00042
		iid Var	0.00047	0.000519	0.00006	0.0003	0.00109
		Tot. SD	0.0213	0.0109	0.0087	0.0197	0.036
		Spatial SD	0.0099	0.00709	0.0025	0.0082	0.0197
		iid SD	0.0179	0.0102	0.0065	0.016	0.0318
	UC	Intercept	-1.757	0.011	-1.771	-1.771	-1.742
		Urban	0.002	0.024	-0.029	-0.029	0.033
		Cluster Var	0.00233	0.00187	0.00048	0.00183	0.00483
		Phi	0.286	0.243	0.034	0.213	0.668
		Tot. Var	0.00064	0.000635	0.00011	0.00045	0.00139
		Spatial Var	0.000167	0.000242	0.000009	0.00008	0.000428
		iid Var	0.00047	0.000534	0.00006	0.0003	0.00109
		Cluster SD	0.0415	0.0187	0.0185	0.0396	0.0669
		Tot. SD	0.0217	0.011	0.0092	0.0199	0.0363
		Spatial SD	0.0101	0.00718	0.0027	0.0084	0.0199
		iid SD	0.0181	0.0103	0.0067	0.0162	0.0319
SPDE							
	uc	Intercept	-1.756	0.015	-1.775	-1.756	-1.736
		Range	813	2458	36	638	3799
		Spatial Var	0.00119	0.0016	0.00014	0.00069	0.00267
		Spatial SD	0.0276	0.0163	0.0103	0.0242	0.0485
	uC	Intercept	-1.757	0.015	-1.776	-1.757	-1.737
		Range	822	2491	37	660	3934
		Spatial Var	0.00116	0.00162	0.00014	0.00066	0.00259
		Spatial SD	0.0272	0.0161	0.0103	0.0237	0.0476
		Cluster Var	0.0024	0.00189	0.0005	0.0017	0.0157
		Cluster SD	0.042	0.0187	0.016	0.032	0.12
	Uc	Intercept	-1.756	0.016	-1.776	-1.756	-1.736
		Urban	0.002	0.024	-0.03	0.002	0.033

Table A.4: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	845	2757	37	762	4707
	Spatial Var	0.00121	0.00161	0.00015	0.00071	0.00269
	Spatial SD	0.028	0.0164	0.0106	0.0246	0.0489
UC	Intercept	-1.757	0.016	-1.777	-1.757	-1.737
	Urban	0.002	0.024	-0.03	0.002	0.033
	Range	862	2861	37	795	4994
	Spatial Var	0.00118	0.00162	0.00015	0.00068	0.00263
	Spatial SD	0.0277	0.0163	0.0107	0.0241	0.0483
	Cluster Var	0.0024	0.00187	0.0005	0.0017	0.0153
	Cluster SD	0.041	0.0186	0.016	0.032	0.119

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-0.6	6.7	0.68	4.7	79	2.08
Direct						
	-0.1	7.2	0.73	*4.9*	79	*2.19*
Smoothed Direct						
	-2.0	5.1	0.53	4.2	78	1.84
BYM2						
uc	-1.7	**4.8**	**0.50**	4.0	79	**1.78**
uC	1.9	4.9	**0.50**	4.0	**80**	1.83
Uc	-1.9	**4.8**	**0.50**	4.0	79	**1.78**
UC	1.7	4.9	**0.50**	4.0	**80**	1.83
SPDE						
uc	1.3	*7.6*	*0.78*	4.0	*83*	1.92
uC	**0.0**	5.7	0.59	**3.7**	*83*	1.85
Uc	-0.4	7.1	0.74	4.0	*83*	1.93
UC	1.8	7.5	*0.78*	4.1	*83*	1.94

Table A.5: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.6: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.748	0.015	-1.766	-1.748	-1.729
	Phi	0.497	0.249	0.161	0.494	0.839
	Tot. Var	0.0134	0.00398	0.0088	0.0128	0.0186
	Spatial Var	0.0068	0.00437	0.0018	0.0061	0.0128
	iid Var	0.0065	0.00369	0.0021	0.0061	0.0114
BYM2						
uc	Intercept	-1.748	0.015	-1.766	-1.766	-1.73
	Phi	0.495	0.25	0.16	0.491	0.838
	Tot. Var	0.013	0.004	0.009	0.012	0.018
	Spatial Var	0.0066	0.00425	0.0018	0.0059	0.0124
	iid Var	0.0064	0.00358	0.002	0.006	0.0111
	Tot. SD	0.112	0.0165	0.092	0.111	0.134
	Spatial SD	0.076	0.0268	0.04	0.076	0.111
	iid SD	0.075	0.0237	0.043	0.076	0.105
uC	Intercept	-1.751	0.015	-1.771	-1.771	-1.732
	Cluster Var	0.0073	0.00352	0.0032	0.0067	0.012
	Phi	0.498	0.251	0.16	0.494	0.841
	Tot. Var	0.0127	0.00386	0.0083	0.0122	0.0178
	Spatial Var	0.0065	0.0042	0.0017	0.0058	0.0122
	iid Var	0.0062	0.00357	0.0019	0.0058	0.011
	Cluster SD	0.078	0.0215	0.051	0.078	0.107
	Tot. SD	0.111	0.0166	0.091	0.11	0.133
	Spatial SD	0.075	0.0267	0.04	0.075	0.11
	iid SD	0.074	0.0238	0.042	0.075	0.104
Uc	Intercept	-1.747	0.015	-1.766	-1.766	-1.728
	Urban	-0.006	0.026	-0.04	-0.04	0.027
	Phi	0.492	0.25	0.158	0.487	0.835
	Tot. Var	0.013	0.004	0.008	0.012	0.018

Table A.6: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0066	0.00424	0.0017	0.0059	0.0124
	iid Var	0.0063	0.00357	0.002	0.006	0.0111
	Tot. SD	0.112	0.0166	0.092	0.111	0.134
	Spatial SD	0.075	0.0269	0.04	0.075	0.11
	iid SD	0.075	0.0237	0.043	0.076	0.105
UC	Intercept	-1.75	0.016	-1.77	-1.77	-1.729
	Urban	-0.008	0.027	-0.043	-0.043	0.027
	Cluster Var	0.0074	0.00354	0.0033	0.0068	0.0121
	Phi	0.494	0.251	0.158	0.49	0.838
	Tot. Var	0.0126	0.00387	0.0082	0.0121	0.0178
	Spatial Var	0.0065	0.00418	0.0017	0.0057	0.0121
	iid Var	0.0062	0.00355	0.0019	0.0058	0.0109
	Cluster SD	0.079	0.0216	0.051	0.078	0.107
	Tot. SD	0.111	0.0168	0.09	0.109	0.133
	Spatial SD	0.074	0.0267	0.039	0.074	0.109
	iid SD	0.074	0.0238	0.042	0.075	0.104
SPDE						
uc	Intercept	-1.731	0.037	-1.779	-1.731	-1.683
	Range	160	44	111	153	218
	Spatial Var	0.0287	0.00795	0.0195	0.0276	0.0392
	Spatial SD	0.167	0.023	0.139	0.165	0.197
uC	Intercept	-1.732	0.038	-1.78	-1.732	-1.683
	Range	163	45	112	156	222
	Spatial Var	0.0283	0.008	0.0191	0.0272	0.0389
	Spatial SD	0.166	0.0232	0.138	0.164	0.196
	Cluster Var	0.0023	0.00191	0.0003	0.0013	0.0165
	Cluster SD	0.041	0.0193	0.013	0.028	0.125
Uc	Intercept	-1.73	0.038	-1.779	-1.73	-1.682
	Urban	-0.005	0.028	-0.04	-0.005	0.031

Table A.6: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	160	44	110	153	218
	Spatial Var	0.0287	0.00795	0.0195	0.0276	0.0392
	Spatial SD	0.167	0.023	0.139	0.165	0.197
UC	Intercept	-1.731	0.038	-1.78	-1.731	-1.683
	Urban	-0.005	0.028	-0.041	-0.005	0.031
	Range	163	45	112	156	222
	Spatial Var	0.0283	0.008	0.0191	0.0271	0.0389
	Spatial SD	0.166	0.0233	0.137	0.164	0.196
	Cluster Var	0.0023	0.00191	0.0004	0.0013	0.0162
	Cluster SD	0.041	0.0193	0.013	0.029	0.124

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	26.2	7.6	0.85	5.2	*74*	2.01
Direct						
	-5.7	7.7	0.79	5.1	85	*2.66*
Smoothed Direct						
	-5.4	6.5	0.67	4.7	84	2.38
BYM2						
uc	25.6	6.5	0.73	4.8	*74*	1.90
uC	*32.1*	6.1	0.73	4.9	82	2.33
Uc	**-1.2**	**4.3**	**0.45**	**3.8**	**80**	**1.73**
UC	2.1	4.4	**0.45**	**3.8**	81	1.79
SPDE						
uc	26.0	13.2	1.42	5.6	76	2.44
uC	18.2	12.1	1.28	5.5	78	2.50
Uc	5.6	*16.6*	*1.84*	*5.7*	82	2.58
UC	12.3	13.8	1.47	5.1	83	2.48

Table A.7: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.8: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.866	0.019	-1.89	-1.866	-1.843
	Phi	0.744	0.201	0.444	0.794	0.961
	Tot. Var	0.0477	0.0127	0.033	0.0459	0.0645
	Spatial Var	0.0361	0.0152	0.0174	0.0349	0.0561
	iid Var	0.0116	0.0092	0.0019	0.0093	0.0246
BYM2						
uc	Intercept	-1.841	0.017	-1.862	-1.862	-1.821
	Phi	0.754	0.196	0.461	0.804	0.963
	Tot. Var	0.049	0.013	0.035	0.047	0.066
	Spatial Var	0.0378	0.0156	0.0185	0.0367	0.0581
	iid Var	0.0114	0.00901	0.0019	0.0092	0.0241
	Tot. SD	0.22	0.0277	0.186	0.217	0.256
	Spatial SD	0.189	0.0414	0.135	0.191	0.241
	iid SD	0.097	0.0425	0.042	0.094	0.154
uC	Intercept	-1.882	0.021	-1.908	-1.908	-1.855
	Cluster Var	0.0801	0.00901	0.0688	0.0798	0.0919
	Phi	0.738	0.204	0.432	0.788	0.96
	Tot. Var	0.0474	0.0126	0.033	0.0457	0.0642
	Spatial Var	0.0356	0.0152	0.0169	0.0344	0.0554
	iid Var	0.0118	0.00939	0.002	0.0095	0.0251
	Cluster SD	0.282	0.0159	0.262	0.282	0.303
	Tot. SD	0.216	0.0282	0.181	0.213	0.253
	Spatial SD	0.184	0.0414	0.129	0.185	0.235
	iid SD	0.099	0.0434	0.043	0.096	0.158
Uc	Intercept	-1.68	0.014	-1.698	-1.698	-1.663
	Urban	-0.995	0.035	-1.04	-1.04	-0.95
	Phi	0.765	0.184	0.494	0.812	0.962
	Tot. Var	0.021	0.006	0.014	0.02	0.028

Table A.8: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.016	0.00642	0.0082	0.0154	0.0244
	iid Var	0.0047	0.00379	0.00082	0.00375	0.00994
	Tot. SD	0.142	0.0193	0.119	0.141	0.168
	Spatial SD	0.123	0.0258	0.09	0.124	0.156
	iid SD	0.0616	0.0274	0.0269	0.0596	0.0987
UC	Intercept	-1.684	0.015	-1.703	-1.703	-1.664
	Urban	-0.995	0.036	-1.041	-1.041	-0.949
	Cluster Var	0.0073	0.00365	0.0031	0.0067	0.0122
	Phi	0.767	0.185	0.495	0.813	0.963
	Tot. Var	0.0204	0.00569	0.0139	0.0196	0.028
	Spatial Var	0.0158	0.00635	0.0081	0.0152	0.0242
	iid Var	0.00463	0.00379	0.00077	0.00365	0.00986
	Cluster SD	0.079	0.0221	0.051	0.078	0.108
	Tot. SD	0.141	0.0194	0.118	0.14	0.167
	Spatial SD	0.123	0.0256	0.09	0.123	0.155
	iid SD	0.061	0.0275	0.0263	0.0589	0.0984
SPDE						
uc	Intercept	-1.785	0.029	-1.822	-1.785	-1.749
	Range	41	7	33	41	51
	Spatial Var	0.109	0.0159	0.089	0.108	0.13
	Spatial SD	0.328	0.0239	0.298	0.327	0.359
uC	Intercept	-1.808	0.031	-1.848	-1.808	-1.768
	Range	52	10	40	51	65
	Spatial Var	0.088	0.0138	0.071	0.087	0.106
	Spatial SD	0.295	0.0233	0.266	0.294	0.325
	Cluster Var	0.0459	0.00778	0.0363	0.0455	0.0562
	Cluster SD	0.213	0.0182	0.19	0.212	0.237
Uc	Intercept	-1.7	0.044	-1.757	-1.7	-1.643
	Urban	-0.998	0.035	-1.044	-0.998	-0.953

Table A.8: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	204	59	139	193	282
	Spatial Var	0.031	0.00915	0.0206	0.0296	0.0431
	Spatial SD	0.174	0.0253	0.143	0.171	0.207
UC	Intercept	-1.701	0.045	-1.758	-1.701	-1.643
	Urban	-0.999	0.036	-1.044	-0.999	-0.953
	Range	209	61	142	198	289
	Spatial Var	0.0306	0.00925	0.0202	0.0292	0.0429
	Spatial SD	0.173	0.0257	0.142	0.17	0.207
	Cluster Var	0.0025	0.00203	0.0004	0.0015	0.0176
	Cluster SD	0.043	0.0199	0.015	0.032	0.129

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	28.9	8.3	0.93	5.5	71	2.02
Direct						
	-2.9	8.3	0.85	5.3	85	*2.74*
Smoothed Direct						
	-2.0	7.0	0.72	4.9	84	2.45
BYM2						
uc	28.9	7.2	0.82	5.2	*70*	1.91
uC	*36.0*	6.7	0.81	5.1	**80**	2.41
Uc	2.0	**4.8**	**0.50**	**4.0**	79	**1.75**
UC	8.2	**4.8**	0.51	**4.0**	81	1.86
SPDE						
uc	29.2	*15.5*	*1.66*	*6.0*	74	2.54
uC	20.5	13.3	1.42	5.8	78	2.58
Uc	13.3	11.9	1.36	5.0	81	2.43
UC	22.9	10.9	1.19	4.6	82	2.30

Table A.9: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.10: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.855	0.02	-1.879	-1.855	-1.831
	Phi	0.737	0.202	0.435	0.785	0.958
	Tot. Var	0.0489	0.0131	0.0339	0.0471	0.0662
	Spatial Var	0.0366	0.0156	0.0176	0.0354	0.0571
	iid Var	0.0123	0.00958	0.0021	0.01	0.0257
BYM2						
uc	Intercept	-1.83	0.018	-1.852	-1.852	-1.809
	Phi	0.738	0.2	0.442	0.785	0.956
	Tot. Var	0.051	0.013	0.036	0.05	0.069
	Spatial Var	0.0387	0.0163	0.0186	0.0375	0.06
	iid Var	0.0127	0.0096	0.0024	0.0104	0.0261
	Tot. SD	0.225	0.0282	0.19	0.222	0.262
	Spatial SD	0.192	0.0427	0.135	0.193	0.245
	iid SD	0.102	0.0432	0.046	0.1	0.161
uC	Intercept	-1.876	0.022	-1.904	-1.904	-1.849
	Cluster Var	0.093	0.00946	0.081	0.092	0.105
	Phi	0.719	0.208	0.41	0.764	0.951
	Tot. Var	0.0495	0.0131	0.0344	0.0476	0.0669
	Spatial Var	0.0362	0.0157	0.0169	0.0349	0.0568
	iid Var	0.0133	0.0101	0.0024	0.011	0.0273
	Cluster SD	0.304	0.0155	0.284	0.304	0.324
	Tot. SD	0.22	0.0288	0.185	0.218	0.258
	Spatial SD	0.185	0.0426	0.129	0.186	0.238
	iid SD	0.105	0.0443	0.048	0.103	0.165
Uc	Intercept	-1.669	0.014	-1.687	-1.687	-1.651
	Urban	-0.996	0.035	-1.041	-1.041	-0.951
	Phi	0.74	0.19	0.461	0.782	0.95
	Tot. Var	0.022	0.006	0.015	0.021	0.03

Table A.10: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0162	0.0067	0.0081	0.0156	0.025
	iid Var	0.0054	0.00411	0.0011	0.0045	0.0111
	Tot. SD	0.145	0.0196	0.122	0.144	0.171
	Spatial SD	0.124	0.0268	0.089	0.124	0.158
	iid SD	0.067	0.0279	0.031	0.065	0.104
UC	Intercept	-1.675	0.016	-1.696	-1.696	-1.655
	Urban	-0.997	0.036	-1.043	-1.043	-0.95
	Cluster Var	0.0168	0.00502	0.0107	0.0163	0.0234
	Phi	0.742	0.192	0.46	0.785	0.953
	Tot. Var	0.0212	0.00591	0.0144	0.0204	0.029
	Spatial Var	0.0159	0.00663	0.0079	0.0152	0.0246
	iid Var	0.0053	0.00412	0.001	0.0043	0.011
	Cluster SD	0.125	0.0201	0.099	0.125	0.151
	Tot. SD	0.144	0.0198	0.119	0.142	0.17
	Spatial SD	0.123	0.0267	0.088	0.123	0.156
	iid SD	0.066	0.0282	0.03	0.064	0.104
SPDE						
uc	Intercept	-1.77	0.029	-1.808	-1.771	-1.734
	Range	39	7	31	39	48
	Spatial Var	0.118	0.0174	0.097	0.117	0.141
	Spatial SD	0.342	0.0251	0.311	0.341	0.375
uC	Intercept	-1.798	0.032	-1.84	-1.799	-1.757
	Range	53	10	40	52	66
	Spatial Var	0.09	0.0142	0.072	0.088	0.108
	Spatial SD	0.298	0.0237	0.268	0.297	0.329
	Cluster Var	0.0584	0.00829	0.048	0.058	0.0693
	Cluster SD	0.24	0.0172	0.218	0.24	0.263
Uc	Intercept	-1.686	0.042	-1.741	-1.687	-1.632
	Urban	-0.999	0.035	-1.045	-0.999	-0.954

Table A.10: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	186	54	127	176	256
	Spatial Var	0.0316	0.00885	0.0214	0.0303	0.0433
	Spatial SD	0.175	0.0244	0.146	0.173	0.207
UC	Intercept	-1.69	0.044	-1.747	-1.69	-1.633
	Urban	-1	0.036	-1.046	-1	-0.954
	Range	204	60	137	193	282
	Spatial Var	0.0304	0.00922	0.02	0.029	0.0427
	Spatial SD	0.172	0.0257	0.141	0.17	0.206
	Cluster Var	0.0102	0.0042	0.0053	0.0097	0.0204
	Cluster SD	0.094	0.0217	0.066	0.093	0.14

3.3.1.2 50km Spatial Range, 0.15^2 Spatial Variance

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	2.0	6.8	0.70	4.7	*78*	2.07
Direct						
	0.8	7.4	0.76	*5.0*	79	*2.22*
Smoothed Direct						
	0.4	4.8	0.50	4.0	*78*	1.76
BYM2						
uc	1.8	**4.5**	**0.47**	**3.9**	*78*	1.71
uC	*7.0*	**4.5**	**0.47**	**3.9**	**80**	1.76
Uc	1.1	**4.5**	**0.47**	**3.9**	*78*	**1.70**
UC	6.4	**4.5**	**0.47**	**3.9**	**80**	1.76
SPDE						
uc	4.4	*8.3*	*0.87*	4.2	**80**	1.86
uC	4.7	6.2	0.65	**3.9**	79	1.77
Uc	3.6	6.3	0.65	**3.9**	**80**	1.78
UC	4.0	6.7	0.70	4.0	**80**	1.86

Table A.11: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.12: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.763	0.015	-1.782	-1.763	-1.745
	Phi	0.377	0.257	0.068	0.338	0.757
	Tot. Var	0.0094	0.0031	0.0059	0.009	0.0136
	Spatial Var	0.00366	0.003	0.00055	0.00292	0.00781
	iid Var	0.00579	0.00305	0.00209	0.0055	0.00981
BYM2						
uc	Intercept	-1.763	0.014	-1.781	-1.781	-1.745
	Phi	0.374	0.257	0.067	0.333	0.753
	Tot. Var	0.01	0.003	0.006	0.009	0.014
	Spatial Var	0.00371	0.00305	0.00054	0.00296	0.00793
	iid Var	0.00597	0.00305	0.00219	0.00572	0.00996
	Tot. SD	0.096	0.0152	0.078	0.095	0.116
	Spatial SD	0.0541	0.025	0.0219	0.0526	0.0878
	iid SD	0.0731	0.021	0.0444	0.0744	0.0989
uC	Intercept	-1.768	0.015	-1.787	-1.787	-1.749
	Cluster Var	0.0124	0.00459	0.0069	0.0119	0.0185
	Phi	0.377	0.261	0.065	0.335	0.761
	Tot. Var	0.0093	0.00306	0.0058	0.0088	0.0133
	Spatial Var	0.00359	0.003	0.00051	0.00283	0.00772
	iid Var	0.00569	0.003	0.00202	0.00541	0.00964
	Cluster SD	0.107	0.0211	0.08	0.107	0.135
	Tot. SD	0.094	0.0155	0.075	0.093	0.115
	Spatial SD	0.0531	0.0249	0.0212	0.0515	0.0866
	iid SD	0.0712	0.0211	0.0425	0.0723	0.0973
Uc	Intercept	-1.759	0.015	-1.778	-1.778	-1.74
	Urban	-0.02	0.026	-0.054	-0.054	0.013
	Phi	0.367	0.255	0.065	0.324	0.745
	Tot. Var	0.009	0.003	0.006	0.009	0.014

Table A.12: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.00358	0.00298	0.00052	0.00282	0.00772
	iid Var	0.00592	0.003	0.00222	0.00567	0.00986
	Tot. SD	0.096	0.0152	0.077	0.094	0.115
	Spatial SD	0.053	0.0248	0.0214	0.0514	0.0866
	iid SD	0.0728	0.0207	0.0448	0.0741	0.0984
UC	Intercept	-1.764	0.016	-1.784	-1.784	-1.743
	Urban	-0.021	0.027	-0.056	-0.056	0.013
	Cluster Var	0.0124	0.0046	0.0069	0.0119	0.0186
	Phi	0.37	0.259	0.063	0.326	0.754
	Tot. Var	0.0091	0.00303	0.0057	0.0087	0.0131
	Spatial Var	0.00345	0.0029	0.00049	0.00272	0.00748
	iid Var	0.00566	0.00297	0.00203	0.00538	0.00953
	Cluster SD	0.107	0.0212	0.08	0.107	0.135
	Tot. SD	0.093	0.0155	0.074	0.092	0.114
	Spatial SD	0.052	0.0245	0.0207	0.0504	0.0853
	iid SD	0.071	0.0209	0.0428	0.0721	0.0967
SPDE						
uc	Intercept	-1.753	0.022	-1.781	-1.753	-1.725
	Range	72	21	48	68	99
	Spatial Var	0.0236	0.00574	0.0168	0.023	0.0313
	Spatial SD	0.152	0.0186	0.128	0.151	0.176
uC	Intercept	-1.755	0.022	-1.783	-1.755	-1.726
	Range	76	22	50	72	105
	Spatial Var	0.0223	0.00565	0.0156	0.0216	0.0298
	Spatial SD	0.147	0.0188	0.124	0.146	0.172
	Cluster Var	0.0043	0.00291	0.0012	0.0037	0.0218
	Cluster SD	0.059	0.0222	0.029	0.054	0.143
Uc	Intercept	-1.752	0.022	-1.78	-1.752	-1.723
	Urban	-0.012	0.028	-0.048	-0.012	0.024

Table A.12: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	72	21	48	68	99
	Spatial Var	0.0236	0.00575	0.0167	0.0229	0.0312
	Spatial SD	0.151	0.0187	0.128	0.15	0.176
UC	Intercept	-1.753	0.023	-1.782	-1.753	-1.724
	Urban	-0.013	0.029	-0.05	-0.013	0.024
	Range	76	22	50	72	105
	Spatial Var	0.0222	0.00567	0.0155	0.0215	0.0297
	Spatial SD	0.147	0.0189	0.123	0.145	0.171
	Cluster Var	0.0043	0.00292	0.0012	0.0037	0.0229
	Cluster SD	0.059	0.0222	0.029	0.055	0.147

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	27.0	7.4	0.83	*5.1*	72	1.96
Direct						
	-3.5	7.5	0.77	5.0	85	*2.58*
Smoothed Direct						
	-1.8	6.2	0.64	4.6	84	2.32
BYM2						
uc	27.9	6.4	0.74	4.9	*71*	1.85
uC	*36.4*	5.9	0.74	4.9	**80**	2.27
Uc	2.4	4.0	**0.41**	**3.6**	78	**1.55**
UC	7.6	4.0	0.42	**3.6**	79	1.60
SPDE						
uc	26.8	8.4	0.94	4.8	76	2.16
uC	22.1	9.6	1.03	4.9	79	2.28
Uc	18.9	*11.4*	*1.32*	4.8	**80**	2.18
UC	21.5	8.1	0.92	4.2	**80**	2.03

Table A.13: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.14: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.936	0.02	-1.96	-1.936	-1.912
	Phi	0.74	0.226	0.39	0.807	0.972
	Tot. Var	0.0542	0.0145	0.0375	0.0522	0.0735
	Spatial Var	0.0411	0.0186	0.0174	0.0401	0.0653
	iid Var	0.0131	0.0114	0.0016	0.0099	0.0293
BYM2						
uc	Intercept	-1.911	0.017	-1.931	-1.931	-1.891
	Phi	0.767	0.211	0.445	0.834	0.976
	Tot. Var	0.055	0.014	0.038	0.053	0.073
	Spatial Var	0.0428	0.018	0.0198	0.042	0.0661
	iid Var	0.0118	0.0105	0.0014	0.0086	0.0268
	Tot. SD	0.231	0.0293	0.196	0.229	0.27
	Spatial SD	0.201	0.0463	0.139	0.204	0.257
	iid SD	0.096	0.0481	0.036	0.091	0.162
uC	Intercept	-1.953	0.021	-1.979	-1.979	-1.927
	Cluster Var	0.0829	0.00963	0.0707	0.0826	0.0954
	Phi	0.753	0.219	0.418	0.818	0.975
	Tot. Var	0.0535	0.0155	0.0365	0.0508	0.0741
	Spatial Var	0.0412	0.0187	0.0182	0.0397	0.0653
	iid Var	0.0123	0.0112	0.0014	0.0091	0.0281
	Cluster SD	0.287	0.0168	0.265	0.287	0.309
	Tot. SD	0.229	0.0322	0.191	0.225	0.272
	Spatial SD	0.196	0.048	0.133	0.199	0.255
	iid SD	0.098	0.0497	0.035	0.093	0.166
Uc	Intercept	-1.749	0.016	-1.77	-1.77	-1.729
	Urban	-0.998	0.035	-1.044	-1.044	-0.953
	Phi	0.233	0.199	0.032	0.174	0.533
	Tot. Var	0.009	0.003	0.006	0.009	0.013

Table A.14: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.00212	0.00208	0.00025	0.00145	0.00494
	iid Var	0.007	0.00293	0.0035	0.0066	0.0108
	Tot. SD	0.093	0.0155	0.074	0.092	0.114
	Spatial SD	0.0401	0.0209	0.0153	0.0371	0.0692
	iid SD	0.081	0.0179	0.058	0.081	0.103
UC	Intercept	-1.754	0.017	-1.775	-1.775	-1.733
	Urban	-1	0.036	-1.047	-1.047	-0.954
	Cluster Var	0.0135	0.00495	0.0076	0.013	0.0201
	Phi	0.24	0.205	0.032	0.18	0.55
	Tot. Var	0.0087	0.00303	0.0052	0.0082	0.0127
	Spatial Var	0.00208	0.00206	0.00024	0.00143	0.00488
	iid Var	0.0066	0.00292	0.0032	0.0062	0.0104
	Cluster SD	0.112	0.0219	0.084	0.111	0.14
	Tot. SD	0.091	0.0159	0.071	0.09	0.112
	Spatial SD	0.0397	0.0208	0.015	0.0367	0.0688
	iid SD	0.078	0.0183	0.055	0.078	0.101
SPDE						
uc	Intercept	-1.835	0.025	-1.867	-1.835	-1.802
	Range	30	5	24	30	37
	Spatial Var	0.127	0.0215	0.101	0.125	0.156
	Spatial SD	0.355	0.0298	0.318	0.353	0.394
uC	Intercept	-1.858	0.027	-1.892	-1.858	-1.823
	Range	36	7	27	35	45
	Spatial Var	0.101	0.0177	0.079	0.099	0.124
	Spatial SD	0.315	0.0277	0.281	0.314	0.351
	Cluster Var	0.0464	0.00809	0.0363	0.0459	0.0571
	Cluster SD	0.214	0.0188	0.19	0.214	0.238
Uc	Intercept	-1.744	0.021	-1.771	-1.744	-1.717
	Urban	-1.003	0.038	-1.052	-1.003	-0.955

Table A.14: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	58	15	41	56	79
	Spatial Var	0.026	0.0063	0.0184	0.0253	0.0343
	Spatial SD	0.159	0.0195	0.134	0.158	0.184
UC	Intercept	-1.745	0.021	-1.772	-1.745	-1.718
	Urban	-1.004	0.038	-1.053	-1.004	-0.955
	Range	61	17	42	59	83
	Spatial Var	0.0245	0.0062	0.0172	0.0237	0.0327
	Spatial SD	0.154	0.0197	0.13	0.153	0.18
	Cluster Var	0.0042	0.00294	0.0011	0.0032	0.0246
	Cluster SD	0.057	0.0227	0.025	0.049	0.154

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	27.1	7.8	0.87	5.3	71	1.97
Direct						
	-2.6	7.9	0.80	5.1	85	*2.63*
Smoothed Direct						
	-1.5	6.5	0.67	4.7	84	2.38
BYM2						
uc	27.6	6.9	0.78	5.0	*70*	1.86
uC	*36.5*	6.4	0.79	5.0	**80**	2.33
Uc	2.0	4.3	**0.44**	3.8	78	**1.60**
UC	9.5	4.3	0.45	3.8	**80**	1.68
SPDE						
uc	27.4	*12.8*	*1.38*	*5.6*	74	2.30
uC	19.5	7.0	0.75	4.6	78	2.23
Uc	20.4	4.7	0.52	**3.6**	79	2.01
UC	19.8	10.6	1.15	4.7	78	2.18

Table A.15: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.16: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance.

			Est	SD	Q10	Q50	Q90
Smoothed Direct							
hspace1em		Intercept	-1.932	0.021	-1.957	-1.932	-1.907
		Phi	0.72	0.233	0.359	0.784	0.966
		Tot. Var	0.0563	0.0152	0.0389	0.0543	0.0765
		Spatial Var	0.0416	0.0196	0.0167	0.0406	0.0671
		iid Var	0.0147	0.0122	0.002	0.0115	0.0321
BYM2							
	uc	Intercept	-1.907	0.018	-1.929	-1.929	-1.886
		Phi	0.742	0.219	0.409	0.805	0.969
		Tot. Var	0.057	0.015	0.04	0.055	0.076
		Spatial Var	0.0434	0.0191	0.019	0.0424	0.0682
		iid Var	0.0136	0.0113	0.0019	0.0106	0.0297
		Tot. SD	0.236	0.0299	0.2	0.234	0.276
		Spatial SD	0.202	0.0488	0.136	0.205	0.261
		iid SD	0.104	0.0489	0.042	0.1	0.171
	uC	Intercept	-1.953	0.022	-1.981	-1.981	-1.926
		Cluster Var	0.093	0.00988	0.081	0.093	0.106
		Phi	0.726	0.227	0.38	0.787	0.967
		Tot. Var	0.0555	0.016	0.0381	0.0526	0.0767
		Spatial Var	0.0413	0.0195	0.0174	0.0398	0.0665
		iid Var	0.0142	0.0121	0.0019	0.0111	0.0311
		Cluster SD	0.304	0.0162	0.284	0.304	0.325
		Tot. SD	0.233	0.0325	0.195	0.229	0.276
		Spatial SD	0.196	0.05	0.129	0.199	0.257
		iid SD	0.106	0.0508	0.041	0.103	0.175
	Uc	Intercept	-1.745	0.017	-1.767	-1.767	-1.724
		Urban	-1.002	0.036	-1.047	-1.047	-0.956
		Phi	0.224	0.196	0.029	0.164	0.518
		Tot. Var	0.011	0.003	0.007	0.01	0.015

Table A.16: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

			Est	SD	Q10	Q50	Q90
		Spatial Var	0.0024	0.00239	0.00027	0.00162	0.00565
		iid Var	0.0082	0.00331	0.0043	0.0079	0.0126
		Tot. SD	0.101	0.0159	0.081	0.1	0.122
		Spatial SD	0.0426	0.0225	0.0159	0.0393	0.0741
		iid SD	0.088	0.0186	0.064	0.088	0.111
	UC	Intercept	-1.753	0.017	-1.775	-1.775	-1.731
		Urban	-1.005	0.037	-1.052	-1.052	-0.958
		Cluster Var	0.0215	0.0057	0.0145	0.021	0.029
		Phi	0.233	0.205	0.028	0.17	0.545
		Tot. Var	0.0099	0.00337	0.006	0.0094	0.0143
		Spatial Var	0.0023	0.00234	0.00024	0.00153	0.00548
		iid Var	0.0076	0.00328	0.0037	0.0072	0.0118
		Cluster SD	0.144	0.0198	0.118	0.143	0.169
		Tot. SD	0.097	0.0166	0.077	0.096	0.119
		Spatial SD	0.0415	0.0224	0.015	0.0382	0.0729
		iid SD	0.084	0.0192	0.059	0.084	0.108
SPDE							
	uc	Intercept	-1.83	0.026	-1.864	-1.83	-1.797
		Range	30	5	24	29	36
		Spatial Var	0.14	0.0237	0.112	0.138	0.172
		Spatial SD	0.372	0.0313	0.333	0.37	0.413
	uC	Intercept	-1.857	0.028	-1.892	-1.857	-1.821
		Range	36	7	28	36	45
		Spatial Var	0.105	0.0184	0.083	0.103	0.13
		Spatial SD	0.322	0.0281	0.287	0.321	0.359
		Cluster Var	0.0549	0.00848	0.0443	0.0545	0.066
		Cluster SD	0.233	0.0182	0.209	0.232	0.256
	Uc	Intercept	-1.74	0.021	-1.768	-1.74	-1.713
		Urban	-1.004	0.038	-1.053	-1.004	-0.955

Table A.16: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	54	14	38	53	72
	Spatial Var	0.0303	0.00703	0.0218	0.0295	0.0396
	Spatial SD	0.172	0.0201	0.147	0.171	0.198
UC	Intercept	-1.743	0.022	-1.771	-1.743	-1.715
	Urban	-1.005	0.039	-1.055	-1.005	-0.956
	Range	60	16	41	58	81
	Spatial Var	0.0262	0.0067	0.0183	0.0253	0.035
	Spatial SD	0.159	0.0206	0.134	0.158	0.186
	Cluster Var	0.0106	0.00466	0.0052	0.0101	0.021
	Cluster SD	0.096	0.0234	0.065	0.095	0.143

3.3.1.3 150km Spatial Range, 0.3^2 Spatial Variance

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	1.7	7.6	0.78	5.0	76	2.06
Direct						
	1.3	8.3	0.85	*5.2*	79	2.33
Smoothed Direct						
	1.7	6.7	0.69	4.7	79	2.13
BYM2						
uc	2.1	**6.3**	**0.66**	**4.6**	77	**1.93**
uC	12.1	**6.3**	0.67	**4.6**	**80**	2.12
Uc	2.2	**6.3**	**0.66**	**4.6**	77	**1.93**
UC	12.1	**6.3**	0.67	**4.6**	**80**	2.12
SPDE						
uc	12.4	*12.0*	*1.26*	5.0	*85*	*2.67*
uC	*13.5*	10.4	1.09	4.7	*85*	2.52
Uc	12	11.6	1.21	4.9	*85*	2.62
UC	11.8	10.3	1.08	4.7	84	2.56

Table A.17: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance. "Cvg" stands for coverage.

Table A.18: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.771	0.015	-1.789	-1.771	-1.752
	Phi	0.804	0.166	0.562	0.85	0.973
	Tot. Var	0.0407	0.0103	0.0288	0.0393	0.0543
	Spatial Var	0.0329	0.0118	0.0186	0.032	0.0482
	iid Var	0.0078	0.00662	0.0012	0.0059	0.0169
BYM2						
uc	Intercept	-1.772	0.014	-1.789	-1.789	-1.755
	Phi	0.802	0.163	0.564	0.847	0.971
	Tot. Var	0.0412	0.0102	0.0294	0.0398	0.0547
	Spatial Var	0.0333	0.0118	0.0189	0.0324	0.0487
	iid Var	0.0079	0.00655	0.0013	0.0061	0.017
	Tot. SD	0.201	0.0245	0.171	0.199	0.234
	Spatial SD	0.179	0.0328	0.137	0.18	0.22
	iid SD	0.08	0.0363	0.034	0.077	0.129
uC	Intercept	-1.782	0.018	-1.803	-1.803	-1.76
	Cluster Var	0.0278	0.0058	0.0206	0.0275	0.0354
	Phi	0.811	0.165	0.57	0.86	0.977
	Tot. Var	0.0407	0.0103	0.0288	0.0393	0.0543
	Spatial Var	0.0332	0.0119	0.0189	0.0323	0.0487
	iid Var	0.0074	0.00659	0.001	0.0055	0.0166
	Cluster SD	0.165	0.0175	0.143	0.165	0.188
	Tot. SD	0.2	0.0249	0.169	0.198	0.233
	Spatial SD	0.179	0.033	0.137	0.179	0.22
	iid SD	0.076	0.0373	0.03	0.072	0.128
Uc	Intercept	-1.772	0.015	-1.791	-1.791	-1.753
	Urban	-0.001	0.027	-0.035	-0.035	0.034
	Phi	0.802	0.163	0.564	0.847	0.971
	Tot. Var	0.0412	0.0102	0.0294	0.0398	0.0548

Table A.18: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

			Est	SD	Q10	Q50	Q90
		Spatial Var	0.0333	0.0118	0.0189	0.0324	0.0487
		iid Var	0.0079	0.00655	0.0013	0.0062	0.017
		Tot. SD	0.201	0.0245	0.171	0.199	0.234
		Spatial SD	0.179	0.0328	0.137	0.18	0.221
		iid SD	0.08	0.0363	0.034	0.077	0.129
	UC	Intercept	-1.781	0.019	-1.805	-1.805	-1.758
		Urban	-0.002	0.029	-0.039	-0.039	0.036
		Cluster Var	0.0278	0.0058	0.0206	0.0275	0.0354
		Phi	0.811	0.165	0.57	0.859	0.976
		Tot. Var	0.0406	0.0103	0.0288	0.0393	0.0543
		Spatial Var	0.0332	0.0119	0.0188	0.0322	0.0487
		iid Var	0.0074	0.00659	0.001	0.0056	0.0166
		Cluster SD	0.165	0.0175	0.143	0.165	0.188
		Tot. SD	0.2	0.0249	0.169	0.198	0.233
		Spatial SD	0.179	0.033	0.137	0.179	0.22
		iid SD	0.076	0.0373	0.03	0.073	0.128
SPDE							
	uc	Intercept	-1.801	0.061	-1.88	-1.801	-1.723
		Range	150	29	117	146	188
		Spatial Var	0.092	0.0203	0.069	0.09	0.12
		Spatial SD	0.302	0.0329	0.262	0.299	0.345
	uC	Intercept	-1.802	0.061	-1.88	-1.802	-1.724
		Range	154	29	119	150	193
		Spatial Var	0.091	0.0205	0.067	0.089	0.119
		Spatial SD	0.3	0.0334	0.259	0.297	0.344
		Cluster Var	0.0034	0.00267	0.0007	0.0026	0.027
		Cluster SD	0.051	0.0222	0.022	0.044	0.161
	Uc	Intercept	-1.801	0.061	-1.88	-1.801	-1.723
		Urban	-0.002	0.029	-0.04	-0.002	0.035

Table A.18: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	150	29	117	146	188
	Spatial Var	0.093	0.0203	0.069	0.09	0.12
	Spatial SD	0.302	0.0328	0.262	0.299	0.345
UC	Intercept	-1.802	0.062	-1.881	-1.802	-1.723
	Urban	-0.003	0.03	-0.041	-0.003	0.035
	Range	153	29	119	150	192
	Spatial Var	0.091	0.0205	0.067	0.089	0.119
	Spatial SD	0.3	0.0334	0.259	0.297	0.344
	Cluster Var	0.0034	0.00267	0.0007	0.0026	0.0272
	Cluster SD	0.051	0.0222	0.022	0.045	0.162

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	25.2	7.4	0.82	5.1	*72*	1.89
Direct						
	-3.1	7.6	0.77	5.0	84	*2.54*
Smoothed Direct						
	-1.5	6.5	0.67	4.6	85	2.36
BYM2						
uc	26.6	6.6	0.74	4.9	*72*	1.82
uC	*37.0*	6.2	0.77	5.0	**80**	2.32
Uc	3.1	5.1	0.52	4.1	77	1.72
UC	12.2	5.1	0.54	4.1	**80**	1.87
SPDE						
uc	25.8	*11.8*	*1.27*	*5.5*	74	2.41
uC	20.0	10.1	1.07	5.2	77	2.39
Uc	18.2	4.8	0.53	3.8	82	2.13
UC	2.2	**3.5**	**0.37**	**3.4**	**80**	**1.53**

Table A.19: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance. "Cvg" stands for coverage.

Table A.20: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-2.036	0.024	-2.065	-2.036	-2.006
	Phi	0.718	0.215	0.393	0.768	0.955
	Tot. Var	0.083	0.0213	0.059	0.08	0.111
	Spatial Var	0.0611	0.0272	0.027	0.0593	0.0967
	iid Var	0.022	0.0166	0.004	0.0183	0.0455
BYM2						
uc	Intercept	-2.01	0.022	-2.036	-2.036	-1.985
	Phi	0.732	0.207	0.422	0.782	0.957
	Tot. Var	0.083	0.0207	0.059	0.08	0.111
	Spatial Var	0.0625	0.0268	0.0288	0.0609	0.0973
	iid Var	0.0208	0.0157	0.0038	0.0172	0.043
	Tot. SD	0.286	0.035	0.243	0.283	0.332
	Spatial SD	0.243	0.056	0.168	0.246	0.312
	iid SD	0.132	0.0553	0.06	0.129	0.206
uC	Intercept	-2.061	0.026	-2.093	-2.093	-2.029
	Cluster Var	0.102	0.0107	0.089	0.102	0.116
	Phi	0.718	0.214	0.398	0.767	0.954
	Tot. Var	0.081	0.021	0.057	0.078	0.109
	Spatial Var	0.0598	0.0263	0.0269	0.058	0.094
	iid Var	0.0216	0.0164	0.0039	0.0179	0.0446
	Cluster SD	0.319	0.0167	0.298	0.319	0.341
	Tot. SD	0.283	0.0358	0.239	0.279	0.33
	Spatial SD	0.237	0.0562	0.162	0.24	0.306
	iid SD	0.134	0.0568	0.06	0.132	0.21
Uc	Intercept	-1.846	0.024	-1.877	-1.877	-1.815
	Urban	-1.024	0.039	-1.074	-1.074	-0.975
	Phi	0.313	0.216	0.066	0.269	0.633
	Tot. Var	0.0323	0.00822	0.0228	0.0312	0.0432

Table A.20: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0105	0.00835	0.0018	0.0083	0.0221
	iid Var	0.0218	0.00837	0.0112	0.0215	0.0326
	Tot. SD	0.178	0.0224	0.151	0.176	0.207
	Spatial SD	0.092	0.0404	0.041	0.089	0.148
	iid SD	0.144	0.0298	0.104	0.146	0.18
UC	Intercept	-1.856	0.025	-1.887	-1.887	-1.825
	Urban	-1.028	0.04	-1.079	-1.079	-0.976
	Cluster Var	0.0287	0.00647	0.0207	0.0283	0.0372
	Phi	0.318	0.224	0.063	0.273	0.652
	Tot. Var	0.0312	0.00824	0.0217	0.0301	0.0421
	Spatial Var	0.0103	0.00833	0.0017	0.0081	0.0219
	iid Var	0.0209	0.00845	0.0103	0.0206	0.0318
	Cluster SD	0.167	0.0192	0.143	0.167	0.192
	Tot. SD	0.175	0.0228	0.147	0.173	0.205
	Spatial SD	0.091	0.0407	0.04	0.088	0.147
	iid SD	0.14	0.0308	0.1	0.143	0.178
SPDE						
uc	Intercept	-1.932	0.039	-1.983	-1.932	-1.882
	Range	53	8	43	52	64
	Spatial Var	0.154	0.019	0.13	0.153	0.179
	Spatial SD	0.391	0.0243	0.36	0.39	0.422
uC	Intercept	-1.952	0.042	-2.005	-1.952	-1.899
	Range	62	11	49	61	76
	Spatial Var	0.135	0.0187	0.112	0.134	0.16
	Spatial SD	0.367	0.0254	0.335	0.366	0.4
	Cluster Var	0.0453	0.00851	0.0348	0.0447	0.0566
	Cluster SD	0.211	0.0201	0.185	0.21	0.237
Uc	Intercept	-1.842	0.051	-1.907	-1.842	-1.777
	Urban	-1.022	0.041	-1.074	-1.022	-0.97

Table A.20: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	126	24	97	123	158
	Spatial Var	0.0776	0.016	0.0585	0.0758	0.0989
	Spatial SD	0.277	0.0284	0.242	0.275	0.314
UC	Intercept	-1.843	0.051	-1.908	-1.843	-1.778
	Urban	-1.022	0.041	-1.075	-1.022	-0.97
	Range	128	25	99	125	162
	Spatial Var	0.0765	0.0162	0.0573	0.0746	0.098
	Spatial SD	0.275	0.0289	0.239	0.273	0.313
	Cluster Var	0.0036	0.00296	0.0007	0.0026	0.0336
	Cluster SD	0.053	0.0236	0.021	0.044	0.18

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	24.8	7.7	0.84	5.2	71	1.90
Direct						
	-3.1	7.8	0.80	5.1	84	*2.58*
Smoothed Direct						
	-1.9	7.0	0.71	4.8	84	2.39
BYM2						
uc	26.1	7.1	0.79	5.0	*70*	1.83
uC	*36.7*	6.7	0.82	5.1	**80**	2.35
Uc	2.4	5.5	**0.56**	**4.2**	76	**1.73**
UC	12.7	5.5	0.57	4.3	79	1.91
SPDE						
uc	24.1	12.4	1.31	*5.8*	71	2.41
uC	20.2	9.8	1.03	5.3	76	2.43
Uc	16.1	*14.2*	*1.53*	5.6	79	2.54
UC	12.4	13.8	1.49	5.6	78	2.52

Table A.21: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated unstratified surveys for the naive and direct models and 100 for the others for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance. "Cvg" stands for coverage.

Table A.22: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-2.032	0.025	-2.062	-2.031	-2.001
	Phi	0.705	0.218	0.377	0.754	0.95
	Tot. Var	0.083	0.0214	0.059	0.08	0.112
	Spatial Var	0.0602	0.0274	0.0259	0.0584	0.0961
	iid Var	0.0231	0.0169	0.0045	0.0196	0.0468
BYM2						
uc	Intercept	-2.007	0.023	-2.033	-2.033	-1.98
	Phi	0.721	0.208	0.41	0.769	0.952
	Tot. Var	0.084	0.0208	0.06	0.081	0.112
	Spatial Var	0.0621	0.027	0.0281	0.0604	0.0974
	iid Var	0.0219	0.016	0.0043	0.0185	0.0446
	Tot. SD	0.287	0.0351	0.245	0.285	0.334
	Spatial SD	0.242	0.0565	0.166	0.245	0.312
	iid SD	0.136	0.0552	0.064	0.134	0.21
uC	Intercept	-2.061	0.027	-2.094	-2.094	-2.028
	Cluster Var	0.11	0.011	0.096	0.11	0.124
	Phi	0.708	0.215	0.385	0.755	0.949
	Tot. Var	0.082	0.0211	0.058	0.079	0.111
	Spatial Var	0.0597	0.0266	0.0262	0.0579	0.0945
	iid Var	0.0228	0.0168	0.0044	0.0192	0.0464
	Cluster SD	0.331	0.0167	0.309	0.331	0.352
	Tot. SD	0.284	0.0359	0.241	0.282	0.332
	Spatial SD	0.237	0.0569	0.161	0.24	0.307
	iid SD	0.138	0.0567	0.064	0.137	0.214
Uc	Intercept	-1.842	0.025	-1.873	-1.873	-1.811
	Urban	-1.031	0.039	-1.081	-1.081	-0.981
	Phi	0.309	0.212	0.066	0.267	0.622
	Tot. Var	0.033	0.00836	0.0233	0.0319	0.0441

Table A.22: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0105	0.00836	0.0019	0.0084	0.0222
	iid Var	0.0225	0.0085	0.0117	0.0221	0.0334
	Tot. SD	0.18	0.0225	0.152	0.178	0.21
	Spatial SD	0.093	0.0403	0.042	0.09	0.148
	iid SD	0.146	0.0298	0.107	0.148	0.182
UC	Intercept	-1.854	0.025	-1.885	-1.885	-1.822
	Urban	-1.034	0.04	-1.086	-1.086	-0.983
	Cluster Var	0.0341	0.00676	0.0257	0.0337	0.0429
	Phi	0.316	0.222	0.062	0.272	0.646
	Tot. Var	0.0318	0.00839	0.0221	0.0307	0.043
	Spatial Var	0.0104	0.00842	0.0017	0.0083	0.0222
	iid Var	0.0214	0.00861	0.0105	0.021	0.0325
	Cluster SD	0.182	0.0185	0.159	0.182	0.206
	Tot. SD	0.177	0.023	0.148	0.175	0.207
	Spatial SD	0.092	0.0409	0.04	0.089	0.147
	iid SD	0.142	0.0311	0.101	0.144	0.18
SPDE						
uc	Intercept	-1.927	0.039	-1.977	-1.927	-1.877
	Range	51	8	42	51	62
	Spatial Var	0.161	0.0198	0.136	0.16	0.187
	Spatial SD	0.4	0.0248	0.369	0.399	0.432
uC	Intercept	-1.95	0.042	-2.004	-1.95	-1.897
	Range	62	11	49	61	76
	Spatial Var	0.138	0.0192	0.114	0.136	0.163
	Spatial SD	0.37	0.0259	0.337	0.369	0.403
	Cluster Var	0.0542	0.00888	0.0432	0.0537	0.066
	Cluster SD	0.231	0.0192	0.207	0.231	0.256
Uc	Intercept	-1.838	0.05	-1.902	-1.838	-1.774
	Urban	-1.021	0.041	-1.074	-1.021	-0.969

Table A.22: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated unstratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	123	24	94	120	155
	Spatial Var	0.077	0.0157	0.0582	0.0753	0.0979
	Spatial SD	0.276	0.0281	0.241	0.274	0.312
UC	Intercept	-1.841	0.051	-1.906	-1.841	-1.776
	Urban	-1.023	0.041	-1.076	-1.023	-0.971
	Range	130	27	99	127	166
	Spatial Var	0.0744	0.0161	0.0553	0.0725	0.0958
	Spatial SD	0.271	0.0292	0.235	0.269	0.309
	Cluster Var	0.0094	0.00458	0.0042	0.0089	0.0224
	Cluster SD	0.089	0.0244	0.058	0.089	0.147

Stratified Sampling Results

3.3.2.1 150*km Spatial Range,* 0.15^2 *Spatial Variance*

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	1.7	6.2	0.63	4.5	**80**	2.05
Direct						
	1.3	*6.4*	*0.65*	*4.6*	79	*2.06*
Smoothed Direct						
	1.6	0.48	0.06	1.4	92	0.86
BYM2						
uc	1.5	**0.39**	**0.05**	**1.3**	90	0.77
uC	*3.1*	0.45	0.06	1.4	*93*	0.89
Uc	1.2	0.41	**0.05**	**1.3**	90	0.78
UC	2.8	0.48	0.06	1.4	*93*	0.90
SPDE						
uc	**-0.2**	2.3	0.26	1.8	84	0.82
uC	1.2	1.7	0.19	1.6	83	**0.72**
Uc	0.6	3.8	0.40	2.0	84	0.91
UC	**0.2**	2.6	0.28	1.8	84	0.81

Table A.23: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.24: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.751	0.01	-1.764	-1.751	-1.739
	Phi	0.274	0.235	0.032	0.204	0.642
	Tot. Var	0.00099	0.00076	0.00026	0.00079	0.00195
	Spatial Var	0.000244	0.000317	0.000017	0.00013	0.000614
	iid Var	0.00074	0.000655	0.00015	0.00055	0.00157
BYM2						
uc	Intercept	-1.752	0.01	-1.764	-1.764	-1.739
	Phi	0.282	0.24	0.033	0.21	0.659
	Tot. Var	0.001	0.001	0	0.001	0.002
	Spatial Var	0.000181	0.000248	0.000011	0.00009	0.000464
	iid Var	0.00053	0.000558	0.00008	0.00035	0.0012
	Tot. SD	0.0228	0.0111	0.0099	0.0212	0.0377
	Spatial SD	0.0106	0.00731	0.0029	0.0088	0.0207
	iid SD	0.0192	0.0105	0.0073	0.0173	0.0334
uC	Intercept	-1.753	0.01	-1.766	-1.766	-1.74
	Cluster Var	0.00233	0.00185	0.00052	0.00183	0.00479
	Phi	0.285	0.241	0.034	0.213	0.663
	Tot. Var	0.00073	0.000676	0.00015	0.00053	0.00154
	Spatial Var	0.000191	0.000267	0.000012	0.000095	0.000478
	iid Var	0.00054	0.000569	0.00008	0.00036	0.00121
	Cluster SD	0.0415	0.0186	0.0187	0.0395	0.0667
	Tot. SD	0.0233	0.0111	0.0105	0.0216	0.0381
	Spatial SD	0.0109	0.00747	0.0031	0.0091	0.021
	iid SD	0.0195	0.0105	0.0077	0.0176	0.0337
Uc	Intercept	-1.753	0.012	-1.768	-1.768	-1.738
	Urban	0.004	0.02	-0.022	-0.022	0.03
	Phi	0.282	0.24	0.033	0.209	0.659
	Tot. Var	0.001	0.001	0	0.001	0.002

Table A.24: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

			Est	SD	Q10	Q50	Q90
		Spatial Var	0.000183	0.000255	0.000011	0.00009	0.000467
		iid Var	0.00053	0.000554	0.00008	0.00035	0.0012
		Tot. SD	0.0228	0.0111	0.0099	0.0212	0.0378
		Spatial SD	0.0106	0.00739	0.0029	0.0088	0.0207
		iid SD	0.0191	0.0105	0.0074	0.0173	0.0334
	UC	Intercept	-1.754	0.012	-1.769	-1.769	-1.739
		Urban	0.004	0.021	-0.022	-0.022	0.031
		Cluster Var	0.00236	0.00186	0.00053	0.00186	0.00484
		Phi	0.284	0.241	0.034	0.212	0.663
		Tot. Var	0.00073	0.000677	0.00015	0.00053	0.00154
		Spatial Var	0.00019	0.000262	0.000012	0.000096	0.000483
		iid Var	0.00054	0.000573	0.00008	0.00036	0.00121
		Cluster SD	0.0416	0.0186	0.0188	0.0396	0.0668
		Tot. SD	0.0233	0.0111	0.0105	0.0216	0.0381
		Spatial SD	0.0109	0.00744	0.0031	0.0091	0.0211
		iid SD	0.0195	0.0105	0.0077	0.0176	0.0336
SPDE							
	uc	Intercept	-1.753	0.016	-1.773	-1.753	-1.732
		Range	875	2489	37	618	3629
		Spatial Var	0.00133	0.00162	0.00019	0.00081	0.00296
		Spatial SD	0.0296	0.0168	0.0117	0.0261	0.0512
	uC	Intercept	-1.754	0.016	-1.774	-1.754	-1.733
		Range	881	2522	40	650	3814
		Spatial Var	0.00131	0.00166	0.00019	0.00078	0.00292
		Spatial SD	0.0293	0.0169	0.0116	0.0257	0.0508
		Cluster Var	0.0023	0.00179	0.0005	0.0015	0.0152
		Cluster SD	0.041	0.0183	0.015	0.03	0.118
	Uc	Intercept	-1.754	0.017	-1.776	-1.754	-1.732
		Urban	0.003	0.021	-0.023	0.003	0.03

Table A.24: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with constant risk (Pop_{suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	885	2638	40	623	3678
	Spatial Var	0.00138	0.00174	0.0002	0.00083	0.00307
	Spatial SD	0.03	0.0172	0.0117	0.0263	0.0519
UC	Intercept	-1.755	0.017	-1.777	-1.755	-1.733
	Urban	0.003	0.021	-0.024	0.003	0.03
	Range	906	2767	42	696	4103
	Spatial Var	0.00136	0.00177	0.00019	0.0008	0.00302
	Spatial SD	0.0296	0.0172	0.0117	0.0259	0.0514
	Cluster Var	0.0023	0.00179	0.0005	0.0015	0.0151
	Cluster SD	0.041	0.0183	0.015	0.03	0.118

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	2.9	6.6	0.68	4.7	79	2.05
Direct						
	1.6	6.8	0.70	*4.8*	79	*2.12*
Smoothed Direct						
	1.8	5.0	0.52	4.1	79	1.81
BYM2						
uc	3.5	**4.8**	**0.49**	4.0	**80**	**1.78**
uC	*6.9*	**4.8**	0.50	4.0	81	1.83
Uc	3.0	**4.8**	**0.49**	4.0	**80**	**1.78**
UC	6.5	**4.8**	0.50	4.0	81	1.83
SPDE						
uc	6.0	*8.1*	*0.84*	4.1	*83*	1.95
uC	4.6	6.4	0.67	**3.8**	*83*	1.87
Uc	3.0	7.8	0.82	4.1	82	1.92
UC	6.0	7.3	0.75	4.0	*83*	1.92

Table A.25: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.26: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.745	0.015	-1.763	-1.745	-1.726
	Phi	0.499	0.249	0.163	0.496	0.841
	Tot. Var	0.0137	0.00403	0.0091	0.0132	0.0191
	Spatial Var	0.007	0.00446	0.0019	0.0063	0.0131
	iid Var	0.0067	0.00376	0.0021	0.0063	0.0117
BYM2						
uc	Intercept	-1.745	0.015	-1.763	-1.763	-1.726
	Phi	0.503	0.248	0.169	0.501	0.841
	Tot. Var	0.014	0.004	0.009	0.013	0.019
	Spatial Var	0.0073	0.00456	0.002	0.0065	0.0135
	iid Var	0.0068	0.00377	0.0021	0.0064	0.0118
	Tot. SD	0.117	0.0167	0.096	0.116	0.139
	Spatial SD	0.08	0.0274	0.043	0.08	0.115
	iid SD	0.078	0.0242	0.045	0.079	0.108
uC	Intercept	-1.747	0.015	-1.767	-1.767	-1.728
	Cluster Var	0.0067	0.00338	0.0028	0.0062	0.0113
	Phi	0.506	0.249	0.169	0.505	0.845
	Tot. Var	0.0138	0.00406	0.0091	0.0132	0.0192
	Spatial Var	0.0072	0.00449	0.002	0.0065	0.0132
	iid Var	0.0066	0.00376	0.002	0.0062	0.0116
	Cluster SD	0.076	0.0214	0.048	0.075	0.104
	Tot. SD	0.116	0.0168	0.095	0.114	0.138
	Spatial SD	0.079	0.0272	0.043	0.079	0.114
	iid SD	0.077	0.0243	0.043	0.078	0.107
Uc	Intercept	-1.746	0.016	-1.766	-1.766	-1.726
	Urban	0.005	0.021	-0.022	-0.022	0.033
	Phi	0.506	0.248	0.171	0.505	0.844
	Tot. Var	0.014	0.004	0.009	0.014	0.02

Table A.26: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0074	0.00461	0.002	0.0066	0.0137
	iid Var	0.0068	0.00379	0.0021	0.0064	0.0118
	Tot. SD	0.117	0.0168	0.097	0.116	0.139
	Spatial SD	0.08	0.0275	0.044	0.08	0.116
	iid SD	0.077	0.0243	0.044	0.079	0.108
UC	Intercept	-1.749	0.017	-1.77	-1.77	-1.728
	Urban	0.005	0.022	-0.024	-0.024	0.033
	Cluster Var	0.0067	0.00338	0.0028	0.0062	0.0113
	Phi	0.509	0.249	0.172	0.508	0.847
	Tot. Var	0.0139	0.0041	0.0092	0.0133	0.0193
	Spatial Var	0.0073	0.00454	0.002	0.0065	0.0135
	iid Var	0.0066	0.00379	0.002	0.0062	0.0117
	Cluster SD	0.076	0.0214	0.048	0.075	0.104
	Tot. SD	0.116	0.0169	0.095	0.115	0.138
	Spatial SD	0.08	0.0273	0.043	0.08	0.115
	iid SD	0.077	0.0245	0.043	0.078	0.107
SPDE						
uc	Intercept	-1.725	0.039	-1.774	-1.725	-1.675
	Range	164	45	114	157	224
	Spatial Var	0.0297	0.0083	0.0202	0.0286	0.0408
	Spatial SD	0.17	0.0236	0.141	0.168	0.201
uC	Intercept	-1.725	0.039	-1.775	-1.725	-1.676
	Range	167	46	116	160	228
	Spatial Var	0.0294	0.00836	0.0198	0.0282	0.0405
	Spatial SD	0.169	0.0238	0.14	0.167	0.201
	Cluster Var	0.0021	0.00178	0.0003	0.0012	0.0166
	Cluster SD	0.039	0.0187	0.012	0.026	0.125
Uc	Intercept	-1.725	0.039	-1.775	-1.725	-1.675
	Urban	0.002	0.023	-0.027	0.002	0.032

Table A.26: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	164	45	114	157	223
	Spatial Var	0.0298	0.00832	0.0202	0.0286	0.0409
	Spatial SD	0.17	0.0236	0.142	0.169	0.201
UC	Intercept	-1.726	0.039	-1.776	-1.726	-1.675
	Urban	0.002	0.023	-0.028	0.002	0.032
	Range	167	46	116	160	228
	Spatial Var	0.0295	0.00838	0.0199	0.0283	0.0406
	Spatial SD	0.17	0.0239	0.14	0.168	0.201
	Cluster Var	0.0021	0.00178	0.0003	0.0011	0.0162
	Cluster SD	0.039	0.0187	0.012	0.026	0.124

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-59.3	9.2	1.28	6.7	60	1.93
Direct						
	-4.0	6.7	0.69	4.8	86	2.47
Smoothed Direct						
	-2.6	5.9	0.6	4.4	84	2.25
BYM2						
uc	-58.8	9.1	1.27	6.8	*56*	1.80
uC	*-60.4*	9.8	1.35	6.7	69	2.38
Uc	**2.6**	**4.7**	**0.49**	**4.0**	79	**1.76**
UC	5.8	4.8	**0.49**	**4.0**	**80**	1.82
SPDE						
uc	-41.8	14.6	1.66	6.8	68	2.43
uC	-55.2	12.8	1.61	*7.0*	68	2.49
Uc	8.2	12.4	1.42	5.2	82	2.52
UC	20.1	*15.1*	*1.69*	5.5	81	*2.55*

Table A.27: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.28: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.864	0.019	-1.887	-1.864	-1.842
	Phi	0.745	0.201	0.444	0.796	0.962
	Tot. Var	0.0494	0.0129	0.0345	0.0476	0.0665
	Spatial Var	0.0375	0.0157	0.018	0.0364	0.0579
	iid Var	0.0119	0.00942	0.002	0.0096	0.0252
BYM2						
uc	Intercept	-1.912	0.016	-1.931	-1.931	-1.893
	Phi	0.765	0.193	0.478	0.817	0.967
	Tot. Var	0.042	0.011	0.03	0.041	0.057
	Spatial Var	0.0329	0.0134	0.0164	0.032	0.0504
	iid Var	0.0093	0.00763	0.0015	0.0073	0.0201
	Tot. SD	0.203	0.026	0.172	0.201	0.238
	Spatial SD	0.177	0.0379	0.127	0.178	0.224
	iid SD	0.087	0.0394	0.037	0.084	0.141
uC	Intercept	-1.989	0.021	-2.015	-2.015	-1.963
	Cluster Var	0.14	0.0128	0.124	0.14	0.157
	Phi	0.747	0.202	0.444	0.798	0.964
	Tot. Var	0.0378	0.0107	0.0254	0.0364	0.052
	Spatial Var	0.0286	0.0122	0.0137	0.0276	0.0447
	iid Var	0.0091	0.00759	0.0014	0.0071	0.0198
	Cluster SD	0.374	0.0171	0.352	0.373	0.396
	Tot. SD	0.192	0.0269	0.159	0.19	0.227
	Spatial SD	0.165	0.037	0.116	0.166	0.211
	iid SD	0.086	0.0394	0.036	0.084	0.14
Uc	Intercept	-1.679	0.014	-1.697	-1.697	-1.661
	Urban	-0.989	0.028	-1.025	-1.025	-0.953
	Phi	0.75	0.189	0.471	0.795	0.957
	Tot. Var	0.022	0.006	0.015	0.021	0.029

Table A.28: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0163	0.00671	0.0082	0.0157	0.0252
	iid Var	0.0052	0.00407	0.0009	0.0042	0.0109
	Tot. SD	0.145	0.0196	0.121	0.144	0.171
	Spatial SD	0.125	0.0267	0.09	0.125	0.158
	iid SD	0.065	0.0281	0.029	0.063	0.103
UC	Intercept	-1.682	0.016	-1.702	-1.702	-1.661
	Urban	-0.99	0.029	-1.026	-1.026	-0.953
	Cluster Var	0.0071	0.00362	0.003	0.0065	0.012
	Phi	0.752	0.19	0.472	0.797	0.958
	Tot. Var	0.0213	0.00592	0.0145	0.0205	0.0292
	Spatial Var	0.0162	0.00668	0.0081	0.0155	0.0249
	iid Var	0.0051	0.00405	0.0009	0.0041	0.0108
	Cluster SD	0.078	0.0223	0.049	0.077	0.107
	Tot. SD	0.144	0.0198	0.12	0.143	0.17
	Spatial SD	0.124	0.0266	0.09	0.124	0.157
	iid SD	0.065	0.0281	0.029	0.063	0.103
SPDE						
uc	Intercept	-1.823	0.029	-1.86	-1.823	-1.786
	Range	28	5	23	28	34
	Spatial Var	0.193	0.0311	0.156	0.19	0.234
	Spatial SD	0.437	0.0349	0.394	0.435	0.483
uC	Intercept	-1.865	0.032	-1.906	-1.865	-1.825
	Range	36	7	28	36	45
	Spatial Var	0.14	0.0236	0.112	0.138	0.172
	Spatial SD	0.372	0.0312	0.334	0.371	0.413
	Cluster Var	0.0765	0.0095	0.0645	0.0761	0.089
	Cluster SD	0.276	0.0172	0.253	0.275	0.298
Uc	Intercept	-1.699	0.044	-1.756	-1.699	-1.643
	Urban	-0.994	0.029	-1.032	-0.994	-0.957

Table A.28: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	196	56	134	186	269
	Spatial Var	0.0314	0.00914	0.021	0.03	0.0435
	Spatial SD	0.175	0.0251	0.144	0.173	0.208
UC	Intercept	-1.7	0.044	-1.757	-1.7	-1.643
	Urban	-0.994	0.029	-1.032	-0.995	-0.957
	Range	201	58	137	190	276
	Spatial Var	0.031	0.00921	0.0205	0.0295	0.0432
	Spatial SD	0.174	0.0254	0.143	0.171	0.207
	Cluster Var	0.0026	0.00205	0.0005	0.0016	0.0166
	Cluster SD	0.043	0.0197	0.015	0.031	0.126

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-59.3	9.3	1.29	6.7	60	1.94
Direct						
	-3.0	7.0	0.72	4.9	85	2.55
Smoothed Direct						
	-2.4	6.2	0.63	4.6	84	2.30
BYM2						
uc	-59.9	9.5	1.32	6.9	*55*	1.81
uC	*-61.4*	10.4	1.43	7.0	68	2.43
Uc	**2.4**	**4.9**	**0.51**	**4.0**	78	**1.76**
UC	8.4	**4.9**	0.52	4.1	**80**	1.87
SPDE						
uc	-38.6	*17.2*	1.90	7.3	67	2.50
uC	-57.3	16.0	*1.96*	*7.6*	66	*2.62*
Uc	17.5	12.9	1.39	5.0	82	2.43
UC	15.3	10.9	1.24	4.7	82	2.34

Table A.29: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.30: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.855	0.019	-1.878	-1.855	-1.832
	Phi	0.745	0.2	0.445	0.795	0.962
	Tot. Var	0.0481	0.0127	0.0335	0.0463	0.0649
	Spatial Var	0.0364	0.0153	0.0175	0.0352	0.0565
	iid Var	0.0117	0.00922	0.002	0.0095	0.0247
BYM2						
uc	Intercept	-1.903	0.016	-1.922	-1.922	-1.884
	Phi	0.764	0.192	0.479	0.814	0.965
	Tot. Var	0.041	0.011	0.029	0.04	0.055
	Spatial Var	0.0319	0.013	0.016	0.0309	0.0489
	iid Var	0.0092	0.00746	0.0015	0.0074	0.0197
	Tot. SD	0.201	0.0257	0.17	0.199	0.235
	Spatial SD	0.174	0.0372	0.126	0.176	0.221
	iid SD	0.087	0.0388	0.037	0.084	0.14
uC	Intercept	-1.986	0.021	-2.013	-2.013	-1.96
	Cluster Var	0.155	0.013	0.139	0.155	0.172
	Phi	0.745	0.201	0.446	0.795	0.962
	Tot. Var	0.0364	0.0105	0.0243	0.035	0.0503
	Spatial Var	0.0275	0.0119	0.0131	0.0263	0.0431
	iid Var	0.0089	0.00734	0.0014	0.0071	0.0192
	Cluster SD	0.393	0.0165	0.372	0.393	0.415
	Tot. SD	0.189	0.0269	0.155	0.187	0.224
	Spatial SD	0.161	0.0365	0.114	0.162	0.207
	iid SD	0.085	0.0386	0.036	0.083	0.138
Uc	Intercept	-1.669	0.014	-1.687	-1.687	-1.651
	Urban	-0.996	0.028	-1.032	-1.032	-0.96
	Phi	0.744	0.19	0.465	0.787	0.953
	Tot. Var	0.021	0.006	0.014	0.02	0.029

Table A.30: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0157	0.00653	0.0078	0.0151	0.0243
	iid Var	0.0052	0.00399	0.001	0.0043	0.0108
	Tot. SD	0.143	0.0194	0.119	0.142	0.169
	Spatial SD	0.122	0.0264	0.088	0.123	0.155
	iid SD	0.065	0.0277	0.03	0.064	0.103
UC	Intercept	-1.674	0.016	-1.695	-1.695	-1.654
	Urban	-0.997	0.029	-1.034	-1.034	-0.959
	Cluster Var	0.0161	0.00506	0.0099	0.0156	0.0228
	Phi	0.748	0.191	0.467	0.793	0.957
	Tot. Var	0.0204	0.00577	0.0138	0.0196	0.0281
	Spatial Var	0.0154	0.00644	0.0077	0.0148	0.0239
	iid Var	0.005	0.00397	0.0009	0.0041	0.0105
	Cluster SD	0.123	0.0205	0.096	0.123	0.149
	Tot. SD	0.141	0.0197	0.117	0.14	0.167
	Spatial SD	0.121	0.0263	0.087	0.121	0.154
	iid SD	0.064	0.0279	0.028	0.062	0.101
SPDE						
uc	Intercept	-1.813	0.029	-1.851	-1.813	-1.776
	Range	27	4	22	27	33
	Spatial Var	0.205	0.0337	0.164	0.201	0.249
	Spatial SD	0.45	0.0368	0.404	0.447	0.498
uC	Intercept	-1.863	0.032	-1.904	-1.863	-1.822
	Range	37	7	28	36	46
	Spatial Var	0.138	0.0237	0.109	0.136	0.17
	Spatial SD	0.37	0.0317	0.33	0.368	0.411
	Cluster Var	0.092	0.0102	0.079	0.091	0.105
	Cluster SD	0.302	0.0168	0.28	0.302	0.324
Uc	Intercept	-1.691	0.043	-1.746	-1.691	-1.636
	Urban	-0.999	0.029	-1.037	-0.999	-0.962

Table A.30: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	191	56	129	180	265
	Spatial Var	0.0303	0.00878	0.0203	0.029	0.042
	Spatial SD	0.172	0.0246	0.142	0.17	0.204
UC	Intercept	-1.694	0.045	-1.752	-1.694	-1.637
	Urban	-1	0.03	-1.038	-1	-0.962
	Range	209	63	140	198	292
	Spatial Var	0.0293	0.00915	0.019	0.0278	0.0414
	Spatial SD	0.169	0.0259	0.137	0.166	0.203
	Cluster Var	0.0101	0.00428	0.0051	0.0097	0.0194
	Cluster SD	0.094	0.0221	0.065	0.093	0.137

3.3.2.2 50km Spatial Range, 0.15^2 Spatial Variance

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-0.5	6.7	0.68	4.7	*78*	2.04
Direct						
	1.2	6.9	0.71	*4.8*	79	*2.15*
Smoothed Direct						
	0.1	4.6	0.47	3.9	79	1.74
BYM2						
uc	-1.2	**4.4**	0.46	**3.8**	**80**	**1.70**
uC	3.8	**4.4**	0.46	**3.8**	81	1.75
Uc	0.9	**4.4**	0.46	**3.8**	**80**	**1.70**
UC	*6.2*	**4.4**	0.46	**3.8**	81	1.75
SPDE						
uc	1.4	6.3	0.66	**3.8**	*82*	1.77
uC	1.8	*7.3*	*0.76*	4.0	81	1.80
Uc	3.6	6.8	0.70	3.9	*82*	1.79
UC	4.4	6.7	0.70	4.0	81	1.81

Table A.31: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.32: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.764	0.015	-1.782	-1.763	-1.745
	Phi	0.36	0.258	0.057	0.313	0.746
	Tot. Var	0.0095	0.00307	0.006	0.0091	0.0136
	Spatial Var	0.00355	0.00301	0.00047	0.00277	0.00774
	iid Var	0.00597	0.00305	0.00218	0.00572	0.01
BYM2						
uc	Intercept	-1.766	0.015	-1.784	-1.784	-1.747
	Phi	0.355	0.256	0.057	0.307	0.737
	Tot. Var	0.01	0.003	0.006	0.009	0.014
	Spatial Var	0.00364	0.00308	0.00049	0.00283	0.00795
	iid Var	0.0062	0.00311	0.0024	0.006	0.0103
	Tot. SD	0.098	0.0151	0.079	0.097	0.118
	Spatial SD	0.0533	0.0254	0.0209	0.0516	0.088
	iid SD	0.075	0.0209	0.047	0.077	0.101
uC	Intercept	-1.77	0.015	-1.789	-1.789	-1.751
	Cluster Var	0.0119	0.00439	0.0067	0.0114	0.0178
	Phi	0.358	0.259	0.055	0.309	0.746
	Tot. Var	0.0095	0.00308	0.006	0.0091	0.0136
	Spatial Var	0.00353	0.00302	0.00045	0.00274	0.00774
	iid Var	0.00598	0.00308	0.00216	0.00573	0.00999
	Cluster SD	0.105	0.0207	0.078	0.104	0.131
	Tot. SD	0.096	0.0154	0.077	0.095	0.116
	Spatial SD	0.0524	0.0253	0.0201	0.0507	0.0868
	iid SD	0.0735	0.0211	0.0448	0.0748	0.0993
Uc	Intercept	-1.758	0.016	-1.778	-1.778	-1.738
	Urban	-0.025	0.022	-0.053	-0.053	0.002
	Phi	0.348	0.253	0.055	0.299	0.729
	Tot. Var	0.01	0.003	0.006	0.009	0.014

Table A.32: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

			Est	SD	Q10	Q50	Q90
		Spatial Var	0.0035	0.00299	0.00047	0.0027	0.00765
		iid Var	0.0062	0.00305	0.0024	0.0059	0.0102
		Tot. SD	0.097	0.015	0.078	0.096	0.116
		Spatial SD	0.0522	0.0251	0.0204	0.0504	0.0863
		iid SD	0.075	0.0206	0.047	0.076	0.1
	UC	Intercept	-1.762	0.016	-1.783	-1.783	-1.742
		Urban	-0.026	0.022	-0.055	-0.055	0.003
		Cluster Var	0.0119	0.00438	0.0067	0.0114	0.0177
		Phi	0.351	0.257	0.053	0.301	0.739
		Tot. Var	0.0093	0.00304	0.0058	0.0089	0.0133
		Spatial Var	0.00339	0.00293	0.00043	0.00261	0.00746
		iid Var	0.00592	0.00304	0.00218	0.00566	0.00987
		Cluster SD	0.104	0.0207	0.078	0.104	0.131
		Tot. SD	0.095	0.0154	0.076	0.093	0.115
		Spatial SD	0.0513	0.0249	0.0197	0.0495	0.0852
		iid SD	0.0731	0.0208	0.045	0.0744	0.0987
SPDE							
	uc	Intercept	-1.752	0.023	-1.782	-1.753	-1.723
		Range	75	21	51	72	102
		Spatial Var	0.023	0.00552	0.0164	0.0224	0.0303
		Spatial SD	0.15	0.0181	0.127	0.149	0.173
	uC	Intercept	-1.754	0.023	-1.784	-1.754	-1.725
		Range	79	22	53	76	108
		Spatial Var	0.0218	0.00543	0.0153	0.0211	0.029
		Spatial SD	0.146	0.0183	0.123	0.145	0.169
		Cluster Var	0.0043	0.00279	0.0013	0.0036	0.0208
		Cluster SD	0.058	0.0215	0.029	0.053	0.14
	Uc	Intercept	-1.749	0.023	-1.778	-1.749	-1.719
		Urban	-0.019	0.024	-0.049	-0.019	0.012

Table A.32: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	75	21	51	72	102
	Spatial Var	0.0228	0.00551	0.0162	0.0222	0.0301
	Spatial SD	0.149	0.0182	0.126	0.148	0.173
UC	Intercept	-1.75	0.023	-1.78	-1.75	-1.72
	Urban	-0.019	0.024	-0.05	-0.019	0.011
	Range	79	23	53	76	109
	Spatial Var	0.0215	0.00543	0.0151	0.0209	0.0288
	Spatial SD	0.145	0.0184	0.122	0.144	0.169
	Cluster Var	0.0043	0.0028	0.0013	0.0037	0.0215
	Cluster SD	0.058	0.0215	0.029	0.053	0.143

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-57.7	8.5	1.19	6.5	60	1.88
Direct						
	-3.6	6.4	0.66	4.7	85	2.40
Smoothed Direct						
	-3.0	5.3	0.55	4.3	85	2.19
BYM2						
uc	*-58.0*	8.2	1.17	6.5	*56*	1.76
uC	-56.4	8.7	1.20	6.4	69	2.32
Uc	**0.7**	3.9	**0.40**	3.6	**80**	**1.59**
UC	5.8	**3.8**	**0.40**	3.6	81	1.64
SPDE						
uc	-38.0	11.2	1.29	6.2	67	2.25
uC	-50.3	*14.4*	*1.73*	*7.0*	67	*2.51*
Uc	18.7	11.6	1.26	4.8	**80**	2.20
UC	11.5	10.8	1.24	4.8	81	2.25

Table A.33: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.34: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.938	0.019	-1.96	-1.938	-1.915
	Phi	0.756	0.219	0.418	0.823	0.975
	Tot. Var	0.0553	0.0145	0.0386	0.0534	0.0745
	Spatial Var	0.0427	0.0185	0.0191	0.0418	0.0667
	iid Var	0.0126	0.0112	0.0015	0.0093	0.0287
BYM2						
uc	Intercept	-1.986	0.017	-2.007	-2.007	-1.966
	Phi	0.733	0.229	0.38	0.799	0.97
	Tot. Var	0.046	0.012	0.032	0.044	0.062
	Spatial Var	0.0345	0.0158	0.0143	0.0338	0.0551
	iid Var	0.0113	0.00954	0.0015	0.0087	0.025
	Tot. SD	0.212	0.0275	0.178	0.21	0.248
	Spatial SD	0.179	0.0457	0.117	0.183	0.234
	iid SD	0.094	0.0455	0.036	0.091	0.157
uC	Intercept	-2.066	0.022	-2.093	-2.093	-2.039
	Cluster Var	0.147	0.0126	0.131	0.146	0.163
	Phi	0.704	0.244	0.33	0.767	0.967
	Tot. Var	0.0412	0.0127	0.0274	0.0389	0.058
	Spatial Var	0.0298	0.0152	0.0111	0.0285	0.0495
	iid Var	0.0114	0.00972	0.0014	0.0089	0.0248
	Cluster SD	0.382	0.0164	0.361	0.382	0.404
	Tot. SD	0.2	0.0299	0.165	0.197	0.24
	Spatial SD	0.165	0.0463	0.103	0.168	0.222
	iid SD	0.095	0.046	0.035	0.092	0.157
Uc	Intercept	-1.75	0.017	-1.771	-1.771	-1.728
	Urban	-1.009	0.029	-1.047	-1.047	-0.972
	Phi	0.219	0.191	0.03	0.161	0.506
	Tot. Var	0.01	0.003	0.006	0.009	0.014

Table A.34: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.00216	0.00215	0.00026	0.00146	0.00508
	iid Var	0.0076	0.00311	0.0039	0.0073	0.0116
	Tot. SD	0.097	0.0158	0.077	0.096	0.118
	Spatial SD	0.0405	0.0212	0.0155	0.0372	0.0703
	iid SD	0.084	0.0181	0.062	0.084	0.107
UC	Intercept	-1.754	0.017	-1.776	-1.776	-1.732
	Urban	-1.011	0.03	-1.049	-1.049	-0.973
	Cluster Var	0.0129	0.0049	0.0071	0.0123	0.0195
	Phi	0.227	0.197	0.031	0.167	0.523
	Tot. Var	0.0093	0.00321	0.0057	0.0088	0.0135
	Spatial Var	0.00212	0.00211	0.00025	0.00144	0.00497
	iid Var	0.0072	0.00311	0.0035	0.0068	0.0112
	Cluster SD	0.109	0.0222	0.08	0.108	0.137
	Tot. SD	0.094	0.0162	0.074	0.093	0.116
	Spatial SD	0.04	0.0211	0.0152	0.0369	0.0696
	iid SD	0.082	0.0185	0.058	0.082	0.105
SPDE						
uc	Intercept	-1.875	0.028	-1.911	-1.875	-1.84
	Range	24	4	19	24	29
	Spatial Var	0.217	0.0387	0.171	0.213	0.269
	Spatial SD	0.463	0.0409	0.413	0.461	0.517
uC	Intercept	-1.92	0.029	-1.957	-1.92	-1.882
	Range	29	5	23	29	36
	Spatial Var	0.158	0.0293	0.122	0.155	0.196
	Spatial SD	0.394	0.0365	0.349	0.392	0.442
	Cluster Var	0.0794	0.0102	0.0666	0.079	0.0928
	Cluster SD	0.281	0.0181	0.257	0.28	0.304
Uc	Intercept	-1.745	0.022	-1.772	-1.745	-1.717
	Urban	-1.012	0.031	-1.053	-1.012	-0.972

Table A.34: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	57	15	40	55	76
	Spatial Var	0.0267	0.00643	0.0189	0.026	0.0352
	Spatial SD	0.161	0.0196	0.137	0.16	0.187
UC	Intercept	-1.746	0.022	-1.774	-1.746	-1.718
	Urban	-1.013	0.031	-1.053	-1.013	-0.973
	Range	59	15	41	57	79
	Spatial Var	0.0252	0.00634	0.0177	0.0245	0.0336
	Spatial SD	0.157	0.0198	0.132	0.156	0.183
	Cluster Var	0.0039	0.00285	0.001	0.0031	0.0266
	Cluster SD	0.054	0.0225	0.024	0.046	0.16

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-56.9	8.7	1.21	6.5	60	1.89
Direct						
	-2.4	6.7	0.68	4.7	85	*2.45*
Smoothed Direct						
	-1.6	5.7	0.59	4.4	85	2.24
BYM2						
uc	*-56.9*	8.8	1.22	*6.7*	*56*	1.77
uC	-55.7	9.5	1.27	6.6	68	2.36
Uc	2.2	**4.2**	**0.44**	3.7	79	**1.62**
UC	9.4	**4.2**	**0.44**	3.8	**80**	1.70
SPDE						
uc	-35.0	*14.4*	*1.60*	*6.7*	65	2.33
uC	-52.7	9.9	1.29	6.6	65	2.33
Uc	21.1	4.7	0.54	**3.6**	79	2.03
UC	15.5	12.2	1.37	5.0	76	2.21

Table A.35: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance. "Cvg" stands for coverage.

Table A.36: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.931	0.02	-1.955	-1.931	-1.908
	Phi	0.738	0.226	0.389	0.805	0.971
	Tot. Var	0.055	0.0146	0.0382	0.0531	0.0744
	Spatial Var	0.0416	0.0188	0.0177	0.0407	0.0659
	iid Var	0.0135	0.0116	0.0017	0.0102	0.0301
BYM2						
uc	Intercept	-1.98	0.018	-2.001	-2.001	-1.959
	Phi	0.713	0.236	0.351	0.777	0.965
	Tot. Var	0.046	0.012	0.032	0.044	0.062
	Spatial Var	0.0338	0.0161	0.0132	0.033	0.0549
	iid Var	0.0122	0.00989	0.0018	0.0097	0.0263
	Tot. SD	0.212	0.0277	0.179	0.21	0.249
	Spatial SD	0.177	0.0471	0.112	0.181	0.234
	iid SD	0.099	0.0457	0.04	0.096	0.161
uC	Intercept	-2.065	0.022	-2.093	-2.093	-2.037
	Cluster Var	0.157	0.013	0.141	0.156	0.174
	Phi	0.681	0.251	0.3	0.739	0.961
	Tot. Var	0.041	0.0126	0.0273	0.0387	0.0576
	Spatial Var	0.0287	0.0152	0.0101	0.0273	0.0486
	iid Var	0.0122	0.01	0.0016	0.0099	0.0262
	Cluster SD	0.395	0.0164	0.375	0.395	0.417
	Tot. SD	0.2	0.0297	0.165	0.196	0.24
	Spatial SD	0.162	0.0473	0.098	0.164	0.22
	iid SD	0.099	0.0462	0.039	0.097	0.161
Uc	Intercept	-1.744	0.017	-1.766	-1.766	-1.722
	Urban	-1.007	0.029	-1.044	-1.044	-0.97
	Phi	0.22	0.193	0.029	0.16	0.509
	Tot. Var	0.011	0.003	0.007	0.01	0.015

Table A.36: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.00235	0.00235	0.00027	0.00158	0.00554
	iid Var	0.0083	0.00331	0.0043	0.0079	0.0126
	Tot. SD	0.101	0.016	0.082	0.1	0.122
	Spatial SD	0.0423	0.0224	0.0159	0.0389	0.0735
	iid SD	0.088	0.0186	0.065	0.088	0.112
UC	Intercept	-1.751	0.018	-1.774	-1.774	-1.729
	Urban	-1.01	0.03	-1.048	-1.048	-0.971
	Cluster Var	0.0209	0.00573	0.0139	0.0205	0.0285
	Phi	0.229	0.202	0.028	0.167	0.536
	Tot. Var	0.0098	0.00338	0.006	0.0093	0.0143
	Spatial Var	0.00226	0.0023	0.00024	0.0015	0.00539
	iid Var	0.0075	0.00328	0.0037	0.0072	0.0118
	Cluster SD	0.141	0.0202	0.115	0.141	0.167
	Tot. SD	0.097	0.0167	0.077	0.096	0.119
	Spatial SD	0.0412	0.0222	0.0151	0.0379	0.0725
	iid SD	0.084	0.0192	0.06	0.084	0.108
SPDE						
uc	Intercept	-1.871	0.028	-1.907	-1.871	-1.835
	Range	24	4	19	24	29
	Spatial Var	0.228	0.041	0.179	0.224	0.283
	Spatial SD	0.475	0.0423	0.422	0.472	0.53
uC	Intercept	-1.92	0.03	-1.959	-1.921	-1.882
	Range	30	6	23	29	37
	Spatial Var	0.157	0.0294	0.121	0.154	0.196
	Spatial SD	0.393	0.0368	0.348	0.391	0.442
	Cluster Var	0.09	0.0105	0.077	0.089	0.104
	Cluster SD	0.299	0.0176	0.276	0.298	0.321
Uc	Intercept	-1.741	0.022	-1.769	-1.741	-1.713
	Urban	-1.008	0.031	-1.048	-1.008	-0.968

Table A.36: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	55	14	38	53	73
	Spatial Var	0.0296	0.00702	0.0211	0.0288	0.0389
	Spatial SD	0.17	0.0203	0.144	0.169	0.196
UC	Intercept	-1.744	0.022	-1.772	-1.744	-1.716
	Urban	-1.009	0.032	-1.049	-1.009	-0.968
	Range	60	16	41	57	81
	Spatial Var	0.0255	0.00668	0.0177	0.0247	0.0344
	Spatial SD	0.158	0.0207	0.132	0.156	0.185
	Cluster Var	0.0103	0.00465	0.0049	0.0098	0.0232
	Cluster SD	0.094	0.0238	0.063	0.094	0.15

3.3.2.3 150km Spatial Range, 0.3^2 Spatial Variance

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	0.2	7.5	0.77	*5.0*	76	2.03
Direct						
	1.6	7.7	0.79	*5.0*	**80**	2.27
Smoothed Direct						
	1.7	6.5	0.66	**4.6**	79	2.07
BYM2						
uc	-0.3	6.4	**0.65**	**4.6**	76	**1.90**
uC	9.5	**6.3**	0.66	**4.6**	**80**	2.08
Uc	0.4	6.4	0.66	**4.6**	76	**1.90**
UC	10.2	**6.3**	0.66	**4.6**	**80**	2.09
SPDE						
uc	*13.5*	*11.1*	*1.16*	4.8	*86*	2.60
uC	11.1	9.7	1.01	4.7	85	2.53
Uc	11.2	*11.1*	1.15	4.8	*86*	*2.61*
UC	9.3	10.5	1.09	4.7	85	2.50

Table A.37: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance. "Cvg" stands for coverage.

Table A.38: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-1.771	0.015	-1.789	-1.771	-1.753
	Phi	0.807	0.164	0.566	0.853	0.973
	Tot. Var	0.0403	0.0101	0.0286	0.039	0.0538
	Spatial Var	0.0328	0.0116	0.0186	0.0318	0.048
	iid Var	0.0076	0.00647	0.0011	0.0058	0.0166
BYM2						
uc	Intercept	-1.774	0.014	-1.791	-1.791	-1.757
	Phi	0.807	0.161	0.574	0.851	0.972
	Tot. Var	0.041	0.0101	0.0293	0.0397	0.0544
	Spatial Var	0.0334	0.0117	0.0192	0.0325	0.0486
	iid Var	0.0077	0.00639	0.0012	0.0059	0.0166
	Tot. SD	0.201	0.0244	0.171	0.199	0.233
	Spatial SD	0.18	0.0323	0.138	0.18	0.22
	iid SD	0.078	0.0358	0.033	0.075	0.128
uC	Intercept	-1.783	0.017	-1.805	-1.805	-1.762
	Cluster Var	0.0267	0.00561	0.0197	0.0264	0.034
	Phi	0.815	0.162	0.578	0.863	0.977
	Tot. Var	0.0406	0.0102	0.0288	0.0393	0.0542
	Spatial Var	0.0333	0.0117	0.0191	0.0324	0.0486
	iid Var	0.0073	0.00647	0.001	0.0054	0.0163
	Cluster SD	0.162	0.0173	0.139	0.162	0.184
	Tot. SD	0.2	0.0247	0.169	0.198	0.232
	Spatial SD	0.179	0.0325	0.138	0.18	0.22
	iid SD	0.076	0.037	0.029	0.072	0.127
Uc	Intercept	-1.772	0.016	-1.791	-1.791	-1.753
	Urban	-0.008	0.022	-0.036	-0.036	0.02
	Phi	0.808	0.16	0.575	0.852	0.972
	Tot. Var	0.0409	0.0101	0.0293	0.0396	0.0544

Table A.38: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

			Est	SD	Q10	Q50	Q90
		Spatial Var	0.0333	0.0116	0.0193	0.0324	0.0484
		iid Var	0.0076	0.0064	0.0012	0.0059	0.0165
		Tot. SD	0.201	0.0244	0.171	0.199	0.233
		Spatial SD	0.179	0.0322	0.139	0.18	0.22
		iid SD	0.078	0.0359	0.033	0.075	0.127
	UC	Intercept	-1.781	0.019	-1.804	-1.804	-1.757
		Urban	-0.008	0.024	-0.039	-0.039	0.022
		Cluster Var	0.0267	0.00561	0.0197	0.0264	0.0341
		Phi	0.816	0.161	0.581	0.864	0.977
		Tot. Var	0.0405	0.0102	0.0287	0.0391	0.054
		Spatial Var	0.0333	0.0117	0.0191	0.0323	0.0485
		iid Var	0.0072	0.00639	0.001	0.0054	0.0161
		Cluster SD	0.162	0.0173	0.139	0.162	0.184
		Tot. SD	0.199	0.0247	0.169	0.197	0.232
		Spatial SD	0.179	0.0324	0.138	0.179	0.22
		iid SD	0.075	0.0367	0.03	0.072	0.126
SPDE							
	uc	Intercept	-1.8	0.061	-1.878	-1.8	-1.722
		Range	150	29	117	146	188
		Spatial Var	0.09	0.0199	0.067	0.087	0.116
		Spatial SD	0.298	0.0327	0.258	0.295	0.341
	uC	Intercept	-1.8	0.061	-1.878	-1.801	-1.723
		Range	153	30	119	150	193
		Spatial Var	0.089	0.0201	0.065	0.086	0.116
		Spatial SD	0.296	0.0331	0.255	0.293	0.339
		Cluster Var	0.0032	0.00253	0.0007	0.0024	0.0263
		Cluster SD	0.049	0.0215	0.021	0.042	0.159
	Uc	Intercept	-1.799	0.061	-1.877	-1.799	-1.721
		Urban	-0.006	0.025	-0.038	-0.007	0.025

Table A.38: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population without urban or cluster effects (Pop_{Suc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	150	29	117	146	188
	Spatial Var	0.09	0.0199	0.067	0.087	0.116
	Spatial SD	0.298	0.0327	0.258	0.295	0.34
UC	Intercept	-1.799	0.061	-1.878	-1.8	-1.721
	Urban	-0.007	0.025	-0.039	-0.007	0.025
	Range	153	30	119	150	193
	Spatial Var	0.089	0.0201	0.065	0.086	0.116
	Spatial SD	0.296	0.0331	0.255	0.293	0.339
	Cluster Var	0.0032	0.00255	0.0007	0.0024	0.0262
	Cluster SD	0.049	0.0216	0.021	0.042	0.159

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-54.6	8.4	1.15	6.3	60	1.81
Direct						
	-3.5	6.6	0.67	4.7	85	2.38
Smoothed Direct						
	-2.9	5.7	0.58	4.4	84	2.23
BYM2						
uc	-54.5	8.4	1.15	6.4	*58*	1.73
uC	-50.6	9.2	1.19	6.2	73	2.37
Uc	1.7	5.3	0.54	4.2	77	1.75
UC	10.4	5.3	0.55	4.2	**80**	1.90
SPDE						
uc	-35.4	*16.3*	1.78	7.0	69	2.47
uC	-47.0	15.7	*1.81*	*7.2*	69	*2.56*
Uc	17.0	4.9	0.53	3.8	82	2.17
UC	**0.9**	**3.6**	**0.38**	**3.4**	**80**	**1.55**

Table A.39: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance. "Cvg" stands for coverage.

Table A.40: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-2.037	0.023	-2.065	-2.037	-2.009
	Phi	0.733	0.209	0.416	0.786	0.959
	Tot. Var	0.084	0.0211	0.059	0.081	0.112
	Spatial Var	0.0627	0.0269	0.0288	0.0613	0.0978
	iid Var	0.0209	0.0162	0.0036	0.0171	0.0438
BYM2						
uc	Intercept	-2.086	0.021	-2.11	-2.11	-2.062
	Phi	0.739	0.203	0.435	0.79	0.958
	Tot. Var	0.0726	0.0181	0.0517	0.0701	0.0967
	Spatial Var	0.0549	0.0232	0.0259	0.0535	0.0853
	iid Var	0.0177	0.0135	0.0032	0.0145	0.0368
	Tot. SD	0.267	0.0329	0.227	0.265	0.311
	Spatial SD	0.228	0.0515	0.16	0.231	0.292
	iid SD	0.121	0.0514	0.055	0.119	0.191
uC	Intercept	-2.175	0.026	-2.207	-2.207	-2.143
	Cluster Var	0.166	0.014	0.148	0.166	0.184
	Phi	0.727	0.211	0.411	0.777	0.958
	Tot. Var	0.0668	0.0181	0.0464	0.0641	0.0909
	Spatial Var	0.0496	0.0219	0.0226	0.048	0.0783
	iid Var	0.0172	0.0135	0.0029	0.014	0.0361
	Cluster SD	0.407	0.0173	0.385	0.407	0.429
	Tot. SD	0.256	0.034	0.215	0.253	0.301
	Spatial SD	0.216	0.0509	0.149	0.218	0.279
	iid SD	0.119	0.0519	0.052	0.117	0.189
Uc	Intercept	-1.847	0.025	-1.878	-1.878	-1.816
	Urban	-1.029	0.031	-1.069	-1.069	-0.99
	Phi	0.308	0.21	0.067	0.267	0.618
	Tot. Var	0.0327	0.00832	0.0231	0.0316	0.0438

Table A.40: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0104	0.00824	0.0019	0.0084	0.022
	iid Var	0.0223	0.0084	0.0117	0.0219	0.0331
	Tot. SD	0.179	0.0225	0.152	0.177	0.209
	Spatial SD	0.092	0.0399	0.042	0.089	0.147
	iid SD	0.145	0.0295	0.107	0.147	0.181
UC	Intercept	-1.857	0.025	-1.888	-1.888	-1.825
	Urban	-1.032	0.032	-1.074	-1.074	-0.991
	Cluster Var	0.0274	0.00648	0.0195	0.027	0.036
	Phi	0.316	0.218	0.065	0.274	0.638
	Tot. Var	0.0317	0.00837	0.0221	0.0305	0.0428
	Spatial Var	0.0104	0.00829	0.0018	0.0082	0.0219
	iid Var	0.0213	0.00843	0.0108	0.0209	0.0322
	Cluster SD	0.163	0.0198	0.138	0.163	0.189
	Tot. SD	0.176	0.0229	0.148	0.174	0.206
	Spatial SD	0.092	0.0404	0.04	0.089	0.147
	iid SD	0.142	0.0303	0.102	0.144	0.179
SPDE						
uc	Intercept	-1.973	0.038	-2.021	-1.973	-1.925
	Range	39	6	32	38	46
	Spatial Var	0.22	0.0269	0.186	0.218	0.255
	Spatial SD	0.467	0.0287	0.431	0.466	0.504
uC	Intercept	-2.013	0.04	-2.064	-2.013	-1.961
	Range	47	8	37	47	58
	Spatial Var	0.175	0.0239	0.146	0.174	0.207
	Spatial SD	0.417	0.0285	0.381	0.416	0.454
	Cluster Var	0.079	0.0107	0.0656	0.0785	0.0932
	Cluster SD	0.28	0.0191	0.255	0.28	0.305
Uc	Intercept	-1.843	0.052	-1.91	-1.843	-1.777
	Urban	-1.025	0.033	-1.068	-1.025	-0.983

Table A.40: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population with spatial and urban effects but without cluster effects (Pop_{SUc}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	129	26	99	126	163
	Spatial Var	0.0774	0.0165	0.0579	0.0755	0.0994
	Spatial SD	0.276	0.0293	0.24	0.274	0.315
UC	Intercept	-1.844	0.052	-1.911	-1.844	-1.777
	Urban	-1.026	0.033	-1.069	-1.026	-0.983
	Range	132	27	101	128	167
	Spatial Var	0.0762	0.0166	0.0566	0.0742	0.0983
	Spatial SD	0.274	0.0297	0.237	0.272	0.313
	Cluster Var	0.0037	0.00295	0.0008	0.0027	0.0347
	Cluster SD	0.052	0.0234	0.021	0.043	0.183

	Bias ($\times 10^{-4}$)	**Var** ($\times 10^{-5}$)	**MSE** ($\times 10^{-4}$)	**CRPS** ($\times 10^{-3}$)	**80% Cvg** ($\times 10^{-2}$)	**CI Width** ($\times 10^{-2}$)
Naive						
	-53.4	8.6	1.15	6.3	60	1.82
Direct						
	-2.2	6.8	0.70	4.8	85	2.43
Smoothed Direct						
	-1.5	6.0	0.61	4.5	85	2.27
BYM2						
uc	-52.9	8.7	1.17	6.4	*57*	**1.74**
uC	-49.6	9.6	1.22	6.3	72	2.40
Uc	3.6	5.5	**0.56**	**4.2**	77	1.75
UC	13.7	5.5	0.58	4.3	**80**	1.94
SPDE						
uc	-35.5	13.9	1.53	6.7	67	2.36
uC	-46.0	*14.7*	*1.70*	*7.1*	68	*2.54*
Uc	14.2	13.7	1.51	5.6	79	2.49
UC	13.8	13.7	1.46	5.6	79	*2.54*

Table A.41: Scoring rules for all models when predictions are aggregated to the county level. Based on 250 simulated stratified surveys for the naive and direct models and 100 for the others for the simulated population with all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance. "Cvg" stands for coverage.

Table A.42: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 50km spatial range and 0.15^2 spatial variance.

		Est	SD	Q10	Q50	Q90
Smoothed Direct						
	Intercept	-2.031	0.024	-2.06	-2.031	-2.003
	Phi	0.722	0.215	0.396	0.774	0.957
	Tot. Var	0.083	0.0211	0.059	0.08	0.111
	Spatial Var	0.0614	0.0271	0.0273	0.0598	0.0968
	iid Var	0.0217	0.0165	0.0039	0.0179	0.045
BYM2						
uc	Intercept	-2.08	0.021	-2.105	-2.105	-2.055
	Phi	0.728	0.208	0.416	0.779	0.956
	Tot. Var	0.0726	0.0182	0.0517	0.0702	0.0968
	Spatial Var	0.0542	0.0235	0.0247	0.0528	0.0848
	iid Var	0.0184	0.0137	0.0034	0.0153	0.0378
	Tot. SD	0.267	0.0329	0.227	0.265	0.311
	Spatial SD	0.226	0.0525	0.156	0.229	0.291
	iid SD	0.124	0.0515	0.057	0.122	0.194
uC	Intercept	-2.173	0.026	-2.205	-2.205	-2.141
	Cluster Var	0.174	0.0142	0.156	0.174	0.193
	Phi	0.716	0.215	0.393	0.766	0.955
	Tot. Var	0.0665	0.018	0.0461	0.0639	0.0904
	Spatial Var	0.0487	0.0219	0.0216	0.047	0.0774
	iid Var	0.0178	0.0138	0.003	0.0146	0.0371
	Cluster SD	0.417	0.017	0.395	0.417	0.439
	Tot. SD	0.255	0.0339	0.215	0.252	0.3
	Spatial SD	0.214	0.0515	0.146	0.216	0.278
	iid SD	0.122	0.0521	0.054	0.12	0.192
Uc	Intercept	-1.841	0.025	-1.873	-1.873	-1.809
	Urban	-1.027	0.031	-1.067	-1.067	-0.988
	Phi	0.29	0.209	0.057	0.244	0.601
	Tot. Var	0.0331	0.00837	0.0234	0.032	0.0443

Table A.42: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Spatial Var	0.0099	0.00819	0.0016	0.0077	0.0214
	iid Var	0.0232	0.00857	0.0124	0.0229	0.0342
	Tot. SD	0.18	0.0225	0.153	0.179	0.21
	Spatial SD	0.09	0.0403	0.039	0.086	0.145
	iid SD	0.149	0.0294	0.11	0.151	0.184
UC	Intercept	-1.853	0.025	-1.885	-1.885	-1.821
	Urban	-1.031	0.032	-1.072	-1.072	-0.989
	Cluster Var	0.0336	0.00688	0.0251	0.0332	0.0427
	Phi	0.3	0.221	0.054	0.251	0.631
	Tot. Var	0.0318	0.00838	0.0222	0.0307	0.0429
	Spatial Var	0.0099	0.00832	0.0015	0.0076	0.0215
	iid Var	0.022	0.00868	0.011	0.0216	0.0331
	Cluster SD	0.181	0.0189	0.157	0.181	0.206
	Tot. SD	0.176	0.0229	0.148	0.175	0.207
	Spatial SD	0.089	0.0412	0.037	0.085	0.145
	iid SD	0.144	0.0309	0.103	0.146	0.181
SPDE						
uc	Intercept	-1.969	0.037	-2.017	-1.969	-1.922
	Range	37	5	30	37	44
	Spatial Var	0.226	0.0282	0.191	0.224	0.263
	Spatial SD	0.473	0.0296	0.436	0.472	0.512
uC	Intercept	-2.014	0.04	-2.065	-2.014	-1.962
	Range	47	8	37	46	58
	Spatial Var	0.174	0.0241	0.144	0.172	0.206
	Spatial SD	0.416	0.0288	0.379	0.415	0.453
	Cluster Var	0.089	0.0111	0.075	0.088	0.104
	Cluster SD	0.297	0.0186	0.273	0.297	0.321
Uc	Intercept	-1.841	0.05	-1.905	-1.841	-1.777
	Urban	-1.019	0.033	-1.062	-1.019	-0.977

Table A.42: Parameter summary statistics for all spatial smoothing models when predictions are aggregated to the county level. Averaged over 100 simulated stratified surveys for the simulated population all effects present in the risk model (Pop_{SUC}) with 150km spatial range and 0.3^2 spatial variance. *(continued)*

		Est	SD	Q10	Q50	Q90
	Range	123	25	94	120	156
	Spatial Var	0.0762	0.0159	0.0572	0.0744	0.0973
	Spatial SD	0.274	0.0285	0.239	0.272	0.311
UC	Intercept	-1.843	0.051	-1.909	-1.843	-1.778
	Urban	-1.021	0.034	-1.064	-1.021	-0.978
	Range	130	27	99	127	166
	Spatial Var	0.0736	0.0162	0.0544	0.0717	0.0952
	Spatial SD	0.269	0.0295	0.233	0.267	0.308
	Cluster Var	0.0091	0.00459	0.0039	0.0086	0.025
	Cluster SD	0.088	0.0248	0.056	0.088	0.155

Appendix B

APPENDIX OF CHAPTER 4

B.1 Additional Results for the Simulation Study of Neonatal Mortality

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	0.005	0.010	0.006	62	0.019
SC	0.004	0.009	0.005	69	0.020
SCP	0.004	0.009	0.005	71	0.021
Total Deaths					
S	95	241	124	54	332
SC	83	190	96	71	382
SCP	83	190	96	73	390
Relative Prevalence					
S	0.042	0.108	0.060	42	0.110
SC	0.028	0.103	0.051	64	0.162
SCP	0.029	0.104	0.05	76	0.216

Table B.1: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = -2.9$, $\beta^{\text{URB}} = -1$ under the representative sampling scheme.

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	0.004	0.006	0.004	57	0.011
SC	0.003	0.005	0.003	68	0.011
SCP	0.003	0.005	0.003	70	0.012
Total Deaths					
S	75	169	80	50	200
SC	53	124	57	69	217
SCP	53	124	57	71	225
Relative Prevalence					
S	-0.017	0.243	0.139	45	0.245
SC	-0.012	0.245	0.129	62	0.375
SCP	-0.006	0.243	0.125	78	0.527

Table B.2: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = -3.9$, $\beta^{\text{URB}} = 0$ under the stratified sampling scheme.

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	0.005	0.006	0.004	51	0.011
SC	0.003	0.005	0.003	66	0.011
SCP	0.003	0.005	0.003	68	0.012
Total Deaths					
S	88	174	83	46	197
SC	64	124	57	67	214
SCP	64	124	57	70	222
Relative Prevalence					
S	0.000	0.249	0.130	51	0.261
SC	0.006	0.251	0.123	66	0.394
SCP	0.011	0.25	0.120	81	0.542

Table B.3: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = -3.9$, $\beta^{\text{URB}} = 0$ under the representative sampling scheme.

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	0.000	0.034	0.019	70	0.071
SC	0.002	0.031	0.018	76	0.074
SCP	0.002	0.031	0.018	76	0.075
Total Deaths					
S	12	1303	696	46	1267
SC	31	738	395	78	1697
SCP	31	738	395	78	1713
Relative Prevalence					
S	0.008	0.058	0.032	59	0.083
SC	0.015	0.059	0.031	75	0.128
SCP	0.015	0.06	0.031	78	0.137

Table B.4: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = 0$, $\beta^{\text{URB}} = -1$ under the stratified sampling scheme.

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	-0.003	0.033	0.019	69	0.067
SC	-0.001	0.031	0.017	76	0.070
SCP	-0.001	0.031	0.017	76	0.071
Total Deaths					
S	-44	1300	700	46	1202
SC	-27	755	391	77	1663
SCP	-27	755	391	77	1680
Relative Prevalence					
S	0.003	0.059	0.033	57	0.081
SC	0.010	0.06	0.031	74	0.125
SCP	0.010	0.06	0.031	77	0.134

Table B.5: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = 0$, $\beta^{\text{URB}} = -1$ under the representative sampling scheme.

	Bias	RMSE	CRPS	80% Coverage	80% CI Width
Prevalence					
S	0.001	0.03	0.016	77	0.072
SC	0.001	0.03	0.016	79	0.074
SCP	0.001	0.03	0.016	80	0.075
Total Deaths					
S	28	1789	837	44	1304
SC	30	773	393	79	1759
SCP	30	773	393	80	1776
Relative Prevalence					
S	0.000	0.077	0.039	57	0.099
SC	0.002	0.077	0.037	79	0.160
SCP	0.002	0.077	0.037	82	0.172

Table B.6: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = 0$, $\beta^{\text{URB}} = 0$ under the stratified sampling scheme.

	Bias	RMSE	CRPS	80% Coverage	Width
Prevalence					
S	0.000	0.030	0.017	76	0.070
SC	0.000	0.029	0.016	78	0.072
SCP	0.000	0.029	0.016	78	0.073
Total Deaths					
S	8	1744	837	42	1256
SC	12	775	392	79	1724
SCP	12	775	392	79	1742
Relative Prevalence					
S	0.012	0.077	0.041	54	0.098
SC	0.013	0.078	0.039	75	0.161
SCP	0.014	0.079	0.039	79	0.174

Table B.7: Bias, root mean squared error, continuous rank probability score, 80% credible interval coverage, and 80% credible interval achieved coverage for the S, SC, and SCP aggregation models in the simulation study, calculated for neonatal mortality prevalence, total deaths, and relative prevalence (urban versus rural) at the constituency level. Results were simulated with $\beta_0 = 0$, $\beta^{\text{URB}} = 0$ under the representative sampling scheme.

B.2 Additional Results for the Application to Neonatal Mortality Rates

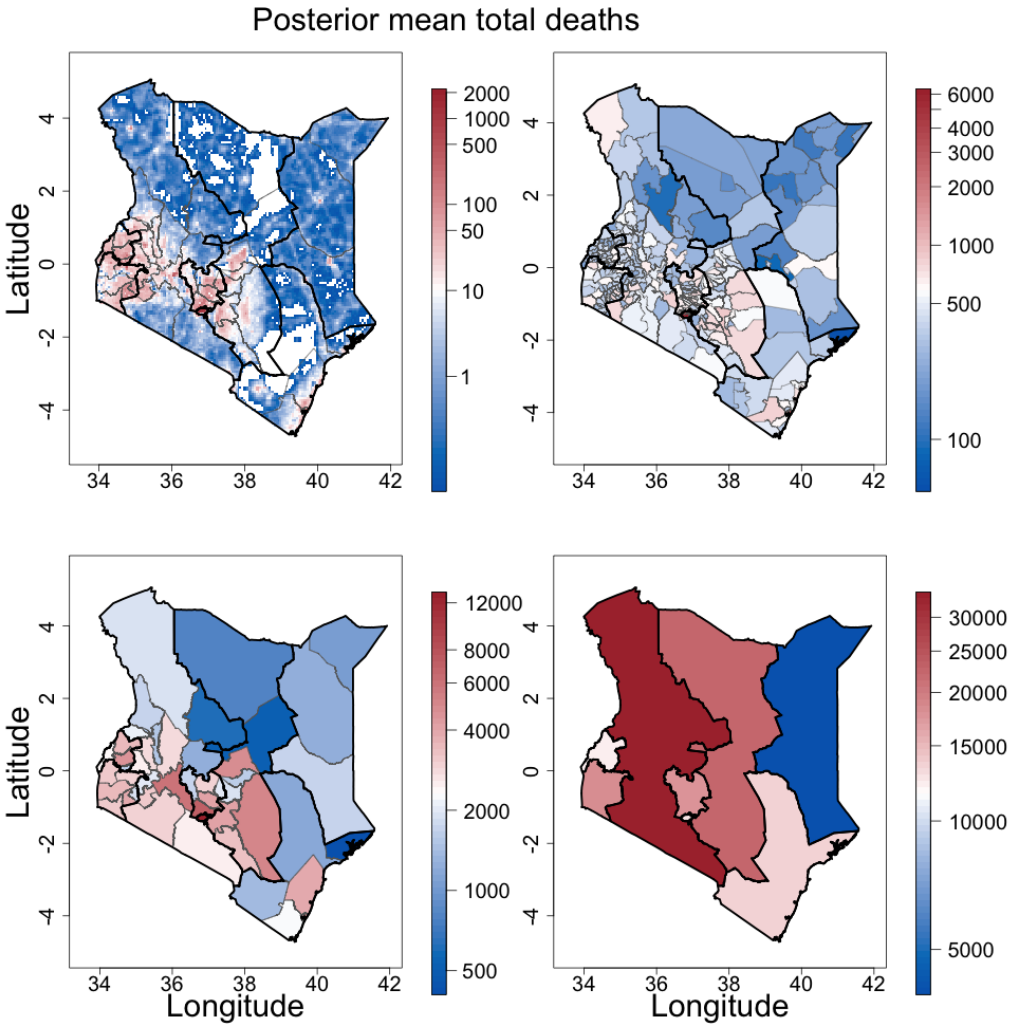


Figure B.1: Predicted total neonatal deaths for the SCP model at the 5km × 5km pixel (top left), constituency (top right), county (bottom left), and province (bottom right) levels. Province and county borders are shown as black and grey lines respectively.

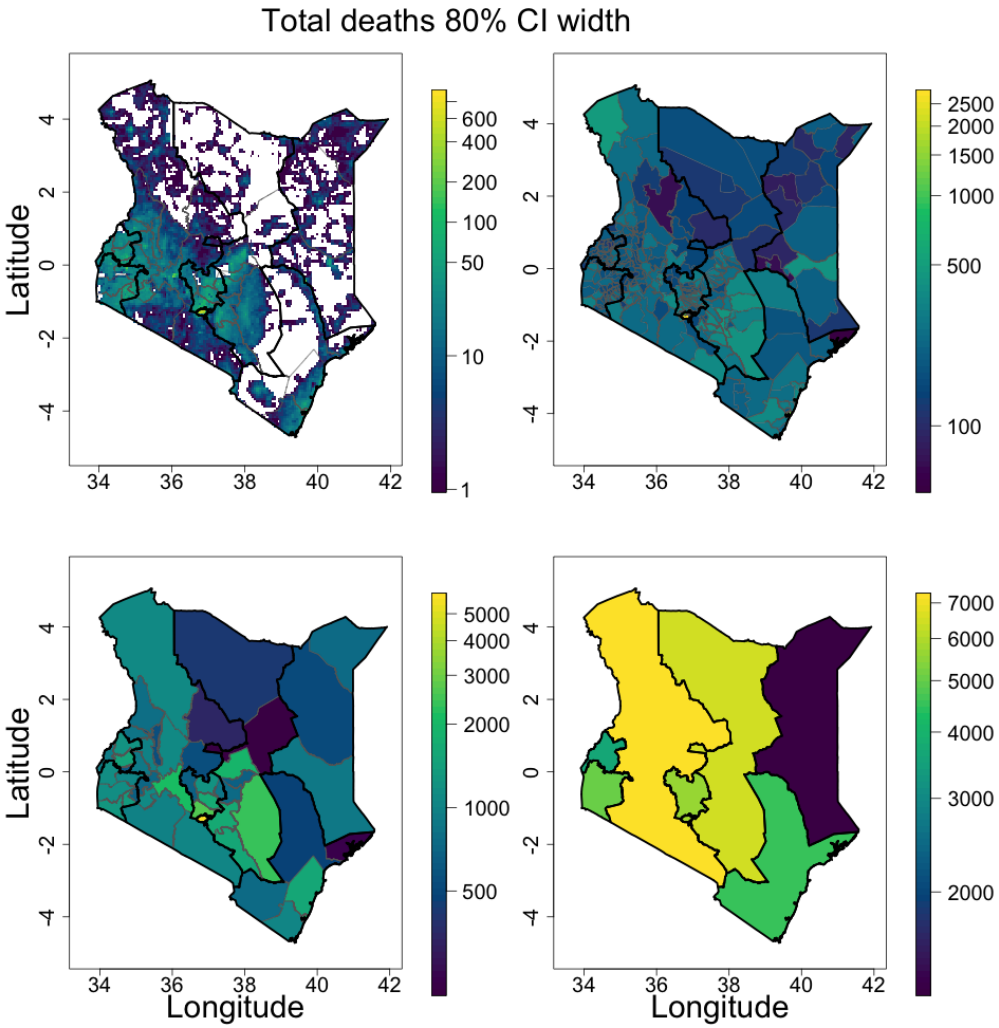


Figure B.2: Predicted total neonatal deaths 80% credible interval width for the SCP model at the 5km × 5km pixel (top left), constituency (top right), county (bottom left), and province (bottom right) levels. Province and county borders are shown as black and grey lines respectively.

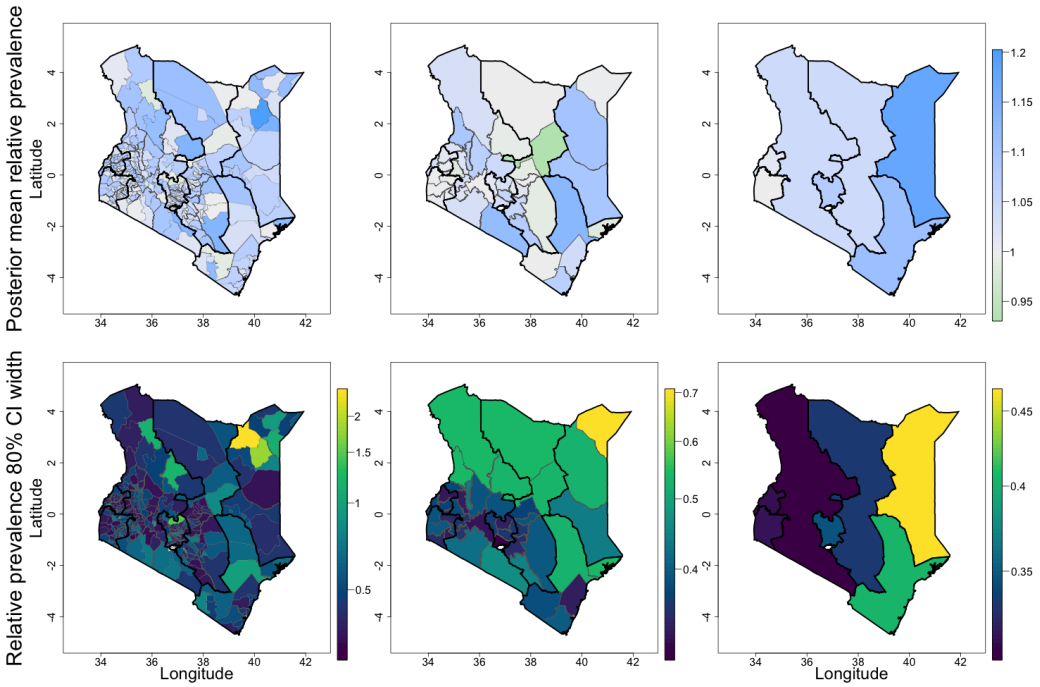


Figure B.3: Predicted neonatal mortality relative prevalence central estimates (top row) and 80% credible interval widths (bottom row) for the SCP model at the constituency (left column), county (middle column), and province (right column) levels. Province and county borders are shown as black and grey lines respectively.

B.3 Drawing from the Posterior of SCP CPAM

To draw a sample from the posterior of p_i and p_{ig} for the SCP CPAM for area i, we can draw a sample from the joint distribution of $\boldsymbol{\eta}^S_{i:}$ for all g and $\boldsymbol{\epsilon}_{i:}$ for all c from the point process model, and then draw samples from the joint distribution of $\boldsymbol{s}_{i:}$, $\boldsymbol{N}_{i:}$, and $\boldsymbol{Y}_{i:}$ for all c conditional on the joint draw from the point level model. We can then aggregate with (4.4) and (4.5). For sufficiently fine g, η^S will not change significantly over A^g, the spatial domain of pixel g. In that case, we can consider only the values of η^S at the centroid of each

pixel g. This would provide a considerable computational advantage if, for instance, η^S is represented as a linear combination of basis functions, since the basis matrix would not need to be recomputed for each draw of $\boldsymbol{s}_{i:}$. The following algorithm illustrates this process:

Algorithm 1 Draw $p_i^{(j)},\ p_{ig}^{(j)}$ from posterior $p_i, p_{ig}|\mathcal{Y}$

1: $\mathbf{s}_{i:}^{(j)} \leftarrow \mathbf{s}_{i:}|q(\cdot)$
2: $\boldsymbol{\epsilon}_{i:}^{(j)} \leftarrow \boldsymbol{\epsilon}_{i:}|\mathcal{Y}$
3: $\mathbf{N}_{i:}^{(j)} \leftarrow \mathbf{N}_{i:}|\mathcal{W}$
4: **for all** $g \in A_i$ **do**
5: $\quad (\eta_{ig}^S)^{(j)} \leftarrow \eta_{i:}^S|\boldsymbol{s}_{ig}, \mathcal{Y}$
6: $\quad N_{ig}^{(j)} \leftarrow \sum_{c \in C^g} N_{ic}^{(j)}$
7: $\quad Y_{ig}^{(j)} \leftarrow Y_{ig}|\mathbf{N}_{i:}, \boldsymbol{\mu}_{i:}$
8: $\quad p_{ig}^{(j)} \leftarrow \sum_{c \in C^g} \frac{N_{ic}^{(j)}}{N_{ig}^{(j)}} \times \frac{Y_{ic}^{(j)}}{N_{ic}^{(j)}}$
9: **end for**
10: $N_i^{(j)} \leftarrow \mathbf{1}^T \mathbf{N}_{i:}^{(j)}$
11: $p_i^{(j)} \leftarrow \sum_{g \in A_i} \frac{N_{ig}^{(j)}}{N_i^{(j)}} \times p_{ig}^{(j)}$

B.4 Deriving the Expectations Under the Superpopulation Model

For the SC aggregation model, the areal aggregation equation can be shown to be an expectation over the superpopulation model of the areal aggregation equation for the SCP

aggregation model:

$$\begin{aligned}
E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:}}\left[p_i\right] &= E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:}}\left[\sum_{c\in B_i}\frac{N_{ic}}{N_i}\times\frac{Y_{ic}}{N_{ic}}\right] \\
&= E_{\boldsymbol{N}_{i:}}\left[\sum_{c\in B_i}\frac{N_{ic}}{N_i}\times E_{\boldsymbol{Y}_{i:}}\left[\frac{Y_{ic}}{N_{ic}}\right]\right] \\
&= \sum_{c\in B_i} E_{\boldsymbol{N}_{i:}}\left[\frac{N_{ic}}{N_i}\right]\times\mu_{ic} \\
&= \sum_{k} E_{\boldsymbol{N}_{i:}}\left[\frac{N_{ic}}{N_i}\right]\sum_{c\in B_i^k}\mu_{ic} \\
&= \sum_{k} Q_i^k\frac{1}{|C_i^k|}\sum_{c\in C_i^k}\mu_{ic}.
\end{aligned}$$

We can then take expectation over EA locations and the cluster level random effects to obtain the S model areal aggregation equation:

$$\begin{aligned}
E_{\boldsymbol{Y}_{i:},\boldsymbol{N}_{i:},\boldsymbol{\epsilon}_{i:},\boldsymbol{s}_{i:}}\left[p_i\right] &= E_{\boldsymbol{\epsilon}_{i:},\boldsymbol{s}_{i:}}\left[\sum_{k} Q_i^k\frac{1}{|B_i^k|}\sum_{c\in B_i^k}\mu_{ic}\right] \\
&= \sum_{k} Q_i^k\frac{1}{|B_i^k|}E_{\boldsymbol{s}_{i:}}\left[\sum_{c\in B_i^k}\tilde{\mu}(\boldsymbol{s}_{ic})\right] \\
&= \sum_{k} Q_i^k E_{\boldsymbol{s}_{i1}}\left[\tilde{\mu}(\boldsymbol{s}_{i1})\right] \\
&= \sum_{k} Q_i^k\int_{A_i^k}\tilde{\mu}(\boldsymbol{s})\times\tilde{q}_{A_i^k}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} \\
&= \int_{A_i}\tilde{\mu}(\boldsymbol{s})\times q_{A_i}(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}.
\end{aligned}$$

Derivations for the pixel level aggregation equations are similar, and assume that spatial variation and variation in general and overall population density do not vary significantly within any pixel due to their small spatial area.

B.5 Optimal Direct Estimate Validation Weights

Under certain assumptions, there is a relatively simple closed form for the validation weights, a_i for $i = 1, \ldots, N$. Let $\boldsymbol{M}$ and $\boldsymbol{D}$ represent the N-vectors of model based and direct estimates respectively, and let $\boldsymbol{y}$ represent the true population prevalences in the N areas. Further assume that $E[M_i] = \mu_i$, $\text{Var}(M_i) = \nu_i^2$, $E[D_i] = y_i$, and $\text{Var}(D_i) = \sigma_i^2$. Under the validation scheme given in Section 4.4, we have that $M_i \perp\!\!\!\perp D_i$ and $D_i \perp\!\!\!\perp D_j$ for all $i \neq j$. In order to obtain a close form for the validation weights, we will later assume that $M_i \perp\!\!\!\perp M_j$ and $M_i \perp\!\!\!\perp D_j$ for $i \neq j$. In the meantime, let $\text{Var}(\boldsymbol{M}) = \Sigma_{\boldsymbol{M}}$, $\text{Var}(\boldsymbol{D}) = \Sigma_{\boldsymbol{D}}$, and $\text{Cov}(\boldsymbol{M}, \boldsymbol{D}) = \Sigma_{\boldsymbol{MD}}$.

We select $\boldsymbol{a}$ as,

$$\boldsymbol{a} = \arg\min_{\boldsymbol{a}} \text{Var}\left(\boldsymbol{a}^T(\boldsymbol{M} - \boldsymbol{D}) - \frac{1}{N}\mathbf{1}^T(\boldsymbol{M} - \boldsymbol{y})\right),$$

subject to the constraints that $\mathbf{1}^T\boldsymbol{a} = 1$, and $a_i \geq 0$ for all $i = 1, \ldots, N$. We choose the above objective function since our goal is to estimate $\frac{1}{N}\mathbf{1}^T(\boldsymbol{M} - \boldsymbol{y})$ for which $\boldsymbol{a}^T(\boldsymbol{M} - \boldsymbol{D})$ is unbiased.

To find $\boldsymbol{a}$, we will minimize the dual objective function:

$$\begin{aligned} L(\boldsymbol{a}, \lambda, \boldsymbol{\gamma}) &= \text{Var}\left(\boldsymbol{a}^T(\boldsymbol{M} - \boldsymbol{D}) - \frac{1}{N}\mathbf{1}^T(\boldsymbol{M} - \boldsymbol{y})\right) + \lambda(\mathbf{1}^T\boldsymbol{a} - 1) - \boldsymbol{a}^T\boldsymbol{\gamma} \\ &= (\boldsymbol{a} - \frac{1}{N}\mathbf{1})^T\Sigma_{\boldsymbol{M}}(\boldsymbol{a} - \frac{1}{N}\mathbf{1}) + \boldsymbol{a}^T\Sigma_{\mathbf{D}}\boldsymbol{a} - 2(\boldsymbol{a} - \frac{1}{N}\mathbf{1})^T\Sigma_{\boldsymbol{MD}}\boldsymbol{a} + \lambda(\mathbf{1}^T\boldsymbol{a} - 1) - \boldsymbol{a}^T\boldsymbol{\gamma} \\ &\propto \boldsymbol{a}^T(\Sigma_{\boldsymbol{M}} + \Sigma_{\boldsymbol{D}} - 2\Sigma_{\boldsymbol{MD}})\boldsymbol{a} + \frac{2}{N}\mathbf{1}^T(\Sigma_{\boldsymbol{MD}} - \Sigma_{\boldsymbol{M}})\boldsymbol{a} + \lambda(\mathbf{1}^T\boldsymbol{a} - 1) - \boldsymbol{a}^T\boldsymbol{\gamma}. \end{aligned}$$

Differentiating with respect to $\boldsymbol{a}$ yields:

$$\frac{\partial}{\partial \boldsymbol{a}} L(\boldsymbol{a}, \lambda, \boldsymbol{\gamma}) = 2(\Sigma_{\boldsymbol{M}} + \Sigma_{\boldsymbol{D}} - 2\Sigma_{\boldsymbol{MD}})\boldsymbol{a} + \frac{2}{N}(\Sigma_{\boldsymbol{MD}} - \Sigma_{\boldsymbol{M}})\mathbf{1} + \lambda\mathbf{1} - \boldsymbol{\gamma}$$

If we assume that $M_i \perp\!\!\!\perp M_j$ and $M_i \perp\!\!\!\perp D_j$ for $i \neq j$, then we can ignore the off diagonal terms of the covariance matrices. Define $W = \text{diag}(\boldsymbol{\sigma})$ and $v = \text{diag}(\boldsymbol{\nu})$. Then the above gradient

becomes:

$$\frac{\partial}{\partial \boldsymbol{a}} L(\boldsymbol{a}, \lambda, \boldsymbol{\gamma}) = 2(V^2 + W^2)\boldsymbol{a} - \frac{2}{N} V^2 \mathbf{1} + \lambda \mathbf{1} - \boldsymbol{\gamma}$$

At the optimum we require $\frac{\partial}{\partial \boldsymbol{a}} L(\boldsymbol{a}, \lambda, \boldsymbol{\gamma}) = \mathbf{0}$. This implies that:

$$\boldsymbol{a} = (V^2 + W^2)^{-1} \left(\frac{1}{N} V^2 \mathbf{1} - \frac{1}{2} \lambda \mathbf{1} + \boldsymbol{\gamma} \right)$$

Dual feasibility requires that $\gamma_i \geq 0$, which, taken in combination with our requirement that $a > 0$ implies $\boldsymbol{\gamma} = \mathbf{0}$, so,

$$\boldsymbol{a} = (V^2 + W^2)^{-1} \left(\frac{1}{N} V^2 \mathbf{1} - \frac{1}{2} \lambda \mathbf{1} \right),$$

at the optimum. Hence,

$$a_i = \frac{\nu_i^2 + C}{N(\nu_i^2 + \sigma_i^2)}$$

for some constant C. Since we require $\mathbf{1}^T \boldsymbol{a} = 1$, we get that,

$$C = \frac{\sum_{j=1}^{N} \frac{\sigma_j^2}{\nu_j^2 + \sigma_j^2}}{\sum_{j=1} \frac{1}{\nu_j^2 + \sigma_j^2}},$$

and so:

$$a_i = \frac{\nu_i^2 + \frac{\sum_{j=1}^{N} \frac{\sigma_j^2}{\nu_j^2 + \sigma_j^2}}{\sum_{j=1} \frac{1}{\nu_j^2 + \sigma_j^2}}}{N(\nu_i^2 + \sigma_i^2)}. \tag{B.1}$$

In other words, we find that the validation weights should be roughly proportional to the signal-to-noise ratio, where the signal use the variance of the model based estimates, and the noise is the variance of the direct estimates.

Appendix C

APPENDIX OF CHAPTER 5

C.1 Relevant Correlation Scales for Spatial Integration

Long-range correlations are especially important when calculating predictions of certain areal averages. With a 'back of the envelope' calculation one can calculate the variance of a predicted spatial integral over a disk with radius R. Let $\hat{r}(d)$ be the estimated covariance, and let $r(d)$ be the true covariance such that,

$$\hat{r}(d) = r(d) + e(d),$$

so e is the error in the covariance estimate a a function of distance. Then if we denote the disk by A, and the true spatial field with $g(\mathbf{x})$, the variance of our spatial integral under the predictive distribution is:

$$\begin{aligned}
\widehat{\mathrm{Var}}(g(A)) &= \int_A \int_A \widehat{\mathrm{Cov}}(\mathbf{u}, \mathbf{v}) \; d\mathbf{u} \; d\mathbf{v} \\
&= \int_A \int_A r(||\mathbf{u} - \mathbf{v}||) + e(||\mathbf{u} - \mathbf{v}||) \; d\mathbf{u} \; d\mathbf{v} \\
&= \mathrm{Var}(g(A)) + \int_A \int_A e(||\mathbf{u} - \mathbf{v}||) \; d\mathbf{u} \; d\mathbf{v}.
\end{aligned}$$

Let D be the random distance between any two points chosen in the disk with independent uniform distributions. Then Tuckwell (2018) shows the density of D is:

$$p_D(d) = \begin{cases} \frac{4d}{\pi R^2} \left(\arccos\left(\frac{d}{2R}\right) - \frac{d}{2R} \sqrt{1 - \left(\frac{d}{2R}\right)^2} \right), & 0 \leq d \leq 2R \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$\widehat{\mathrm{Var}}(g(A)) = \mathrm{Var}(g(A)) + \int_0^{2R} e(D) \cdot \frac{4d}{\pi R^2} \left(\arccos\left(\frac{d}{2R}\right) - \frac{d}{2R}\sqrt{1 - \left(\frac{d}{2R}\right)^2} \right) \, dD.$$

Fig. C.1 shows that the density $p_D(d)$ roughly parabolic with peak just under R (approximately $0.834R$), and has zeros at 0 and $2R$. Because of this, errors in very short and very long-range correlations are less relevant than errors in the assumed correlation function at the spatial scale near the radius of the area over which we integrate, R, when calculating predictive uncertainties. This is of course not the full story, since the covariance structure conditional on the data will not be so neatly stationary and isotropic, and will likely have shorter spatial range. At the same time, we believe this shows greater emphasis must be placed on long range spatial correlations when producing area level predictions, especially in large areas.

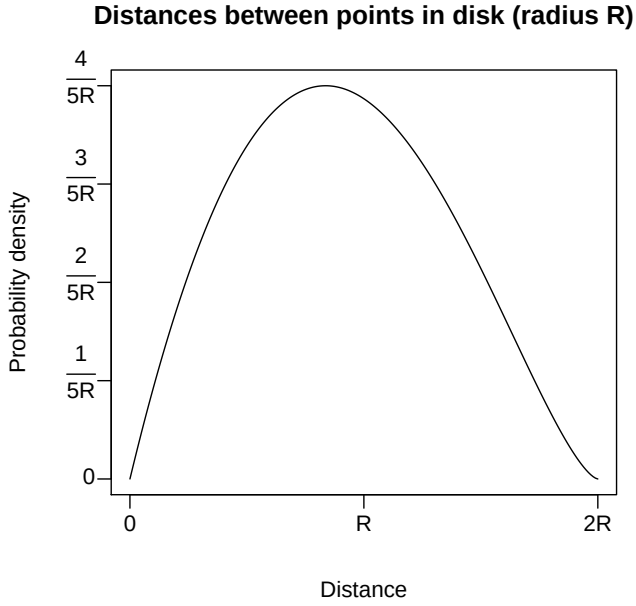


Figure C.1: The distribution of distances between points uniformly distributed on a disk of radius R.

C.2 ELK Sparse Matrix Computations

The computational performance of our implementation of ELK within *inla* is almost entirely determined by how quickly the sparse precision matrix of the basic coefficients $\mathbf{c}$ can be generated. As such, we precompute any information for this task that will improve the performance. Recall that thus basis coefficients for each layer follow independent SAR models with mean zero Gaussian distribution, $\mathbf{c}_l \sim \text{MVN}(\mathbf{0}, \alpha_l \sigma_{\text{S}}^2 \mathbf{B}_l^{-1} \mathbf{B}_l^{-T})$, with,

$$\mathbf{B}_{l,i,j} = \begin{cases} 4 + \kappa_l^2, & i = j \\ -1, & i \in N_l(j) \\ 0, & \text{otherwise,} \end{cases}$$

where $N_l(j)$ is this set of indices of lattice knots in layer l neighboring lattice knot i. The precision matrix for layer l, $\mathbf{Q}_l$, can therefore be represented as,

$$\mathbf{Q}_l = \frac{\omega_l}{\alpha_l \sigma_{\mathrm{S}}^2} \left(\kappa_l^4 \mathbf{I}_{m(l)} - \kappa_l^2 (\mathbf{D}^l + (\mathbf{D}^l)^T) + (\mathbf{D}^l)^T \mathbf{D}^l \right),$$

for matrices,

$$\begin{aligned} \mathbf{D}^l &= \mathbf{D}_x^l + \mathbf{D}_y^l \\ \mathbf{D}_x^l &= \mathbf{I}_{m_y(l)} \otimes \boldsymbol{\nabla}^2_{m_x(l)} \\ \mathbf{D}_y^l &= \mathbf{I}_{m_x(l)} \otimes \boldsymbol{\nabla}^2_{m_y(l)}, \end{aligned}$$

where $m_x(l)$ and $m_y(l)$ are the number of basis functions in the horizontal and vertical directions of layer l, $\mathbf{I}_{m_x(l)}$ and $\mathbf{I}_{m_y(l)}$ are $m_x(l) \times m_x(l)$ and $m_y(l) \times m_y(l)$ identity matrices respectively, and '$\otimes$' denotes the Kronecker product. Note that the variance normalization factor ω_l is a function of κ_l, although we leave out this dependence in the notation for simplicity. We can therefore precompute $\mathbf{D}^l + (\mathbf{D}^l)^T$ and $\mathbf{D}^l)^T \mathbf{D}^l$ in order to calculate $\mathbf{Q}_l$ as quickly as possible for each chosen value of κ_l.

Since there is no exact closed form solution for the functions $f_l : \kappa_l \mapsto \omega_l,\ l = 1, \ldots, L$, they are approximated using monotonic smoothing splines (Hyman, 1983) fit on a log-log scale over a set of reasonable effective ranges for each layer. Throughout this analysis, the effective ranges used for fitting f_1 vary from a fifth of the first layer lattice width to the diameter of the spatial domain, and the effective ranges used when fitting subsequent f_l shrink proportionally with the corresponding lattice widths. Hence, if w is the diameter of the spatial domain, then each f_l is fit with effective ranges varying in the interval $\left(\frac{\delta_l}{5},\ \frac{\delta_1}{\delta_l} \cdot \frac{w}{5} \right)$. We find the splines are nearly linear, so estimates of f_l are very accurate even somewhat outside of the interval used for fitting.

C.3 Fuzzy Coverage and Interval Width for Count Data

For observation i, let $Q^i_{\alpha/2}$ and $Q^i_{1-\alpha/2}$ be the discrete $\alpha/1$ and $1-\alpha$ quantiles of a predictive distribution for empirical proportion y_i so that $p_l \equiv P(y_i < Q^i_{\alpha/2}) \leq \alpha/2$ and $p_u \equiv P(y_i > Q^i_{1-\alpha/2}) \leq \alpha/2$. Then $P(Q^i_{\alpha/2} \leq y_i \leq Q^i_{1-\alpha/2}) \geq 1-\alpha$. We will show how to calculate coverage using fuzzy coverage intervals in order to achieve coverage closer to the nominal rate.

Rather than using a fixed uncertainty interval when performing hypothesis tests for discrete data, randomized tests involving randomized uncertainty intervals are the uniformly most powerful (UMP) one tailed and UMP unbiased (UMPU) two-tailed tests (Lehmann and Romano, 2005, Chapters 3 and 4). In a randomized test, we could randomly reject that y_i is in our interval if y_i is equal to $Q^i_{\alpha/2}$ or $Q^i_{1-\alpha/2}$ in such a way as to obtain equal tail rejection probabilities and achieve $1-\alpha$ coverage. In the lower tail case, we have:

$$\begin{aligned}\alpha/2 &= P(\text{reject } y_i \text{ at lower tail}) \\ &= P(y_i < Q^i_{\alpha/2}) + P(\text{reject } y_i \text{ at lower tail},\ y_i = Q^i_{\alpha/2}).\end{aligned}$$

This implies we can choose $P(\text{reject } y_i \mid y_i = Q^i_{\alpha/2})$ and $P(\text{reject } y_i \mid y_i = Q^i_{1-\alpha/2})$ in the following way in order to achieve the correct coverage, assuming the predictive distribution is correct:

$$\alpha/2 = p_l + P(\text{reject } y_i \mid y_i = Q^i_{\alpha/2}) \cdot P(y_i = Q^i_{\alpha/2}) \tag{C.1}$$

$$\alpha/2 = p_u + P(\text{reject } y_i \mid y_i = Q^i_{1-\alpha/2}) \cdot P(y_i = Q^i_{1-\alpha/2}). \tag{C.2}$$

Since the resulting coverage intervals are random, different statisticians may randomly report different results. To eliminate this possibility, we will follow the proposal of Geyer and Meeden (2005) to use fuzzy set theory to compute fuzzy intervals. To do this, we calculate

the membership function for U^i, the *fuzzy* uncertainty interval for the ith observation, as:

$$I_{U^i}(y_i) = \begin{cases} 1, & Q^i_{\alpha/2} < y_i < Q^i_{1-\alpha/2} \\ P(\text{reject } y_i \mid y_i = Q^i_{\alpha/2}), & y_i = Q^i_{\alpha/2} \\ P(\text{reject } y_i \mid y_i = Q^i_{1-\alpha/2}), & y_i = Q^i_{1-\alpha/2} \\ 0, & \text{otherwise,} \end{cases} \tag{C.3}$$

where $P(\text{reject } y_i \mid y_i = Q^i_{\alpha/2})$ and $P(\text{reject } y_i \mid y_i = Q^i_{\alpha/2})$ are calculated from Eqs. (C.1-C.2). We then calculate coverage for a single observation as the membership function of the fuzzy interval,

$$\text{Cvg}(y_i) = I_{U^i}(y_i),$$

in order to achieve the nominal coverage deterministically.

To calculate fuzzy credible interval width, we modify the standard width calculations by accounting for the rejection probabilities in Eqs. (C.1-C.2):

$$\text{Width}(y_i) = Q^i_{1-\alpha/2} - Q^i_{\alpha/2} - \frac{1}{N_i}\left[P(\text{reject } y_i \mid y_i = Q^i_{\alpha/2}) + P(\text{reject } y_i \mid y_i = Q^i_{1-\alpha/2})\right],$$

where $\frac{1}{N_i}$ is the width of the discrete steps the interval width could increase by if fuzzy intervals were not used.

C.4 Assessing Performance Under Multiscale Dependence: Additional Results

In addition to the results for the simulation study shown in the main text, we also calculated scores for the predictions integrated over the nine grid cells throughout the domain, averaged over the 100 realizations for the single central grid cell without observations, and for the eight other grid cells containing observations. These average scores are respectively given in Tables C.1 and C.2.

	RMSE	CRPS	80% Cvg	CI Width
SPDE	*0.37*	*0.21*	*60*	**0.61**
LK	0.31	0.18	75	0.68
ELK-F	0.33	0.19	67	0.64
ELK-T	**0.27**	**0.16**	**76**	0.66

Table C.1: Scoring rules for predictions of integrals of the latent field over the center most of the nine cells in the 3×3 regular grid. The scoring rules are averaged over 100 simulated realizations and are calculated for each of the considered models. *Italics* indicate worse performance, **boldface** indicates better performance.

	RMSE	CRPS	80% Cvg	CI Width
SPDE	*0.066*	*0.037*	*77*	*0.16*
LK	0.062	0.035	*77*	**0.15**
ELK-F	0.063	0.035	78	**0.15**
ELK-T	**0.061**	**0.034**	**79**	**0.15**

Table C.2: Scoring rules for predictions of integrals of the latent field over the outer most eight of the nine cells in the 3×3 regular grid. The scoring rules are averaged over 100 simulated realizations and are calculated for each of the considered models. *Italics* indicate worse performance, **boldface** indicates better performance.

C.5 Prevalence of Secondary Education in Kenya: Survey Design and Additional Results

The KHDS follows a typical DHS design: it is a stratified, two-stage design, where the first stage consists of selecting enumeration areas (EAs) from each stratum with probability proportional to size (PPS) sampling, where the 'size' used to calculate sampling probabilities is based on the number of households in each EA. The second stage consists of selecting 25 households randomly within each EA, (Kenya National Bureau of Statistics, Ministry of

Health/Kenya, National AIDS Control Council/Kenya, Kenya Medical Research Institute, and National Council For Population And Development/Kenya, 2015; ICF International, 2012). Strata are based on the 47 counties crossed with official urban/rural designations, where two counties, Nairobi and Mombasa, are entirely urban, making 92 strata in total. 1,612 clusters are sampled from the 96,251 EAs in Kenya that are based on the 2009 Kenya Population and Housing Census (Kenya National Bureau Of Statistics, 2014).

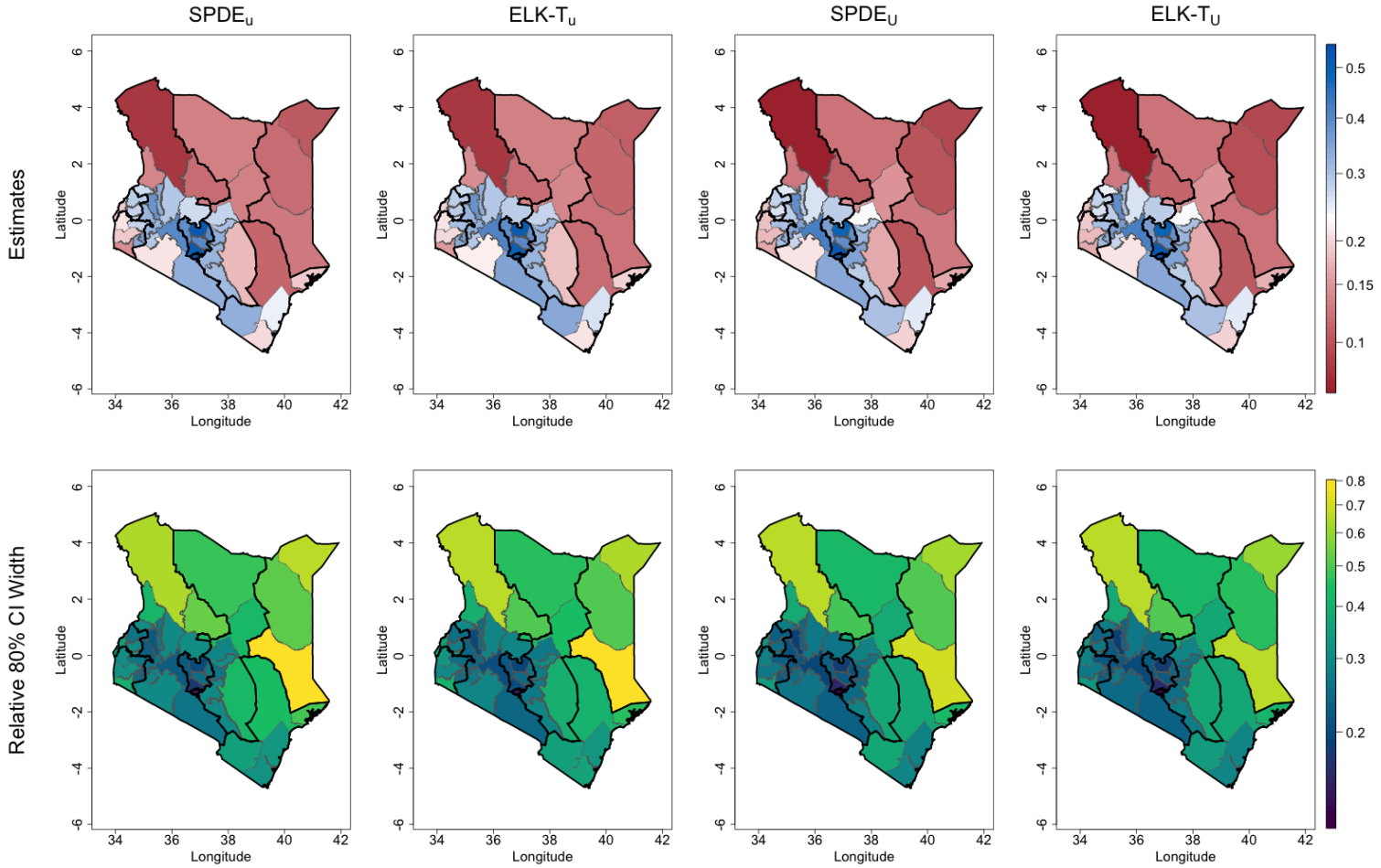


Figure C.2: Central county level predictions (top row) and relative 80% credible interval widths (bottom row) of secondary education prevalence for young women in Kenya in 2014. Models with subscript 'u' and 'U' respectively do and do not include urban effects. Province and county borders are shown as black and grey lines respectively.

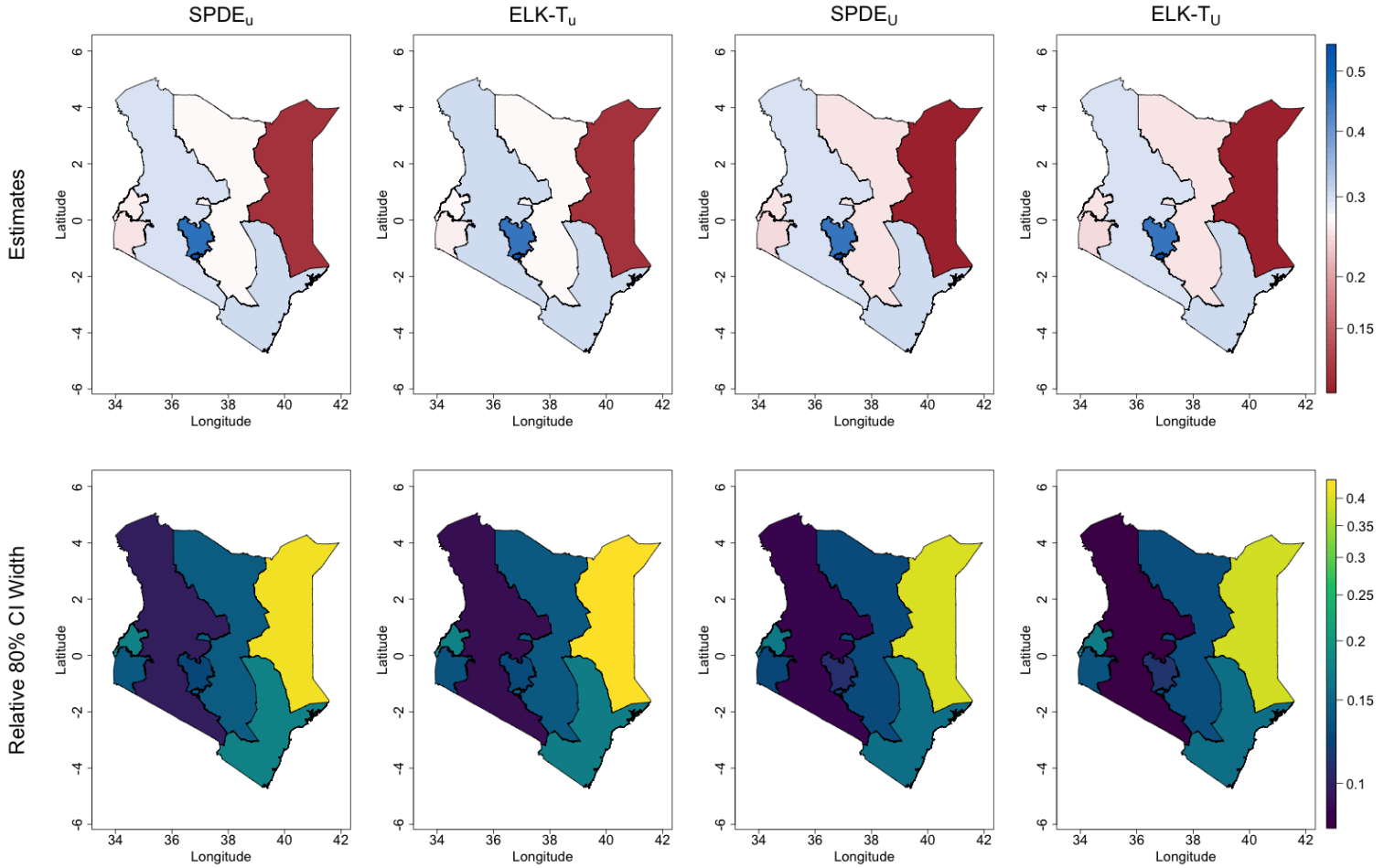


Figure C.3: Central province level predictions (top row) and relative 80% credible interval widths (bottom row) of secondary education prevalence for young women in Kenya in 2014. Models with subscript 'u' and 'U' respectively do and do not include urban effects.

Parameter	Est	SD	Q10	Q50	Q90
***SPDE**$_U$*					
Intercept	-2.392	0.262	-2.381	-2.724	-2.073
Urban	0.927	0.067	0.927	0.841	1.012
Total Var	1.244	0.233	0.941	1.194	1.609
Spatial Var	0.824	0.227	0.555	0.775	1.167
Cluster Var	0.421	0.052	0.342	0.419	0.510
Total SD	1.111	0.103	0.970	1.092	1.268
Spatial SD	0.899	0.124	0.745	0.880	1.080
Cluster SD	0.647	0.040	0.585	0.647	0.714
Range (km)	203	43	154	196	267
***ELK-T**$_U$*					
Intercept	-2.433	0.316	-2.836	-2.416	-2.055
Urban	0.930	0.067	0.844	0.930	1.016
Total Var	1.298	0.290	0.977	1.206	1.705
Spatial Var	0.870	0.277	0.556	0.780	1.283
Cluster Var	0.428	0.053	0.354	0.426	0.487
Total SD	1.132	0.126	0.988	1.098	1.306
Spatial SD	0.921	0.146	0.746	0.883	1.133
Cluster SD	0.653	0.041	0.595	0.653	0.698
Range_1 (km)	571	451	315	375	1365
Range_2 (km)	135	51	78	118	199
α_1	0.480	0.209	0.241	0.538	0.786
α_2	0.520	0.209	0.214	0.462	0.759

Table C.3: Parameter posterior estimates, standard deviations, and 80% CIs for the given models fit to secondary education completion KDHS data for women aged 20-29 in Kenya in 2014.

	$SPDE_U$			$ELK\text{-}T_U$		
Province	Est	Q10	Q90	Est	Q10	Q90
Central	0.4578	0.4331	0.4826	0.4566	0.4326	0.4813
Coast	0.3030	0.2791	0.3276	0.3047	0.2814	0.3287
Eastern	0.2542	0.2377	0.2706	0.2556	0.2381	0.2734
Nairobi	0.5403	0.5082	0.5727	0.5391	0.5066	0.5706
North Eastern	0.1059	0.0856	0.1281	0.1046	0.0845	0.1261
Nyanza	0.2429	0.2277	0.2578	0.2432	0.2278	0.2598
Rift Valley	0.2935	0.2808	0.3066	0.2940	0.2809	0.3065
Western	0.2492	0.2283	0.2696	0.2489	0.2291	0.2684

Table C.4: Province predictions and 80% CIs for prevalence of secondary education completion for women aged 20-29 in Kenya in 2014.

Table C.5: County level predictions and 80% CIs for prevalence of secondary education completion for women aged 20-29 in Kenya in 2014.

	$SPDE_U$			$ELK\text{-}T_U$		
County	Est	Q10	Q90	Est	Q10	Q90
Baringo	0.2615	0.2267	0.2977	0.2607	0.2256	0.2983
Bomet	0.2768	0.2418	0.3128	0.2765	0.2408	0.3131
Bungoma	0.2734	0.2422	0.3063	0.2748	0.2440	0.3082
Busia	0.1810	0.1513	0.2103	0.1806	0.1524	0.2122
Elgeyo Marakwet	0.2935	0.2591	0.3292	0.2953	0.2609	0.3292
Embu	0.3183	0.2800	0.3568	0.3205	0.2829	0.3590
Garissa	0.1250	0.0829	0.1719	0.1225	0.0824	0.1692
Homa Bay	0.1795	0.1552	0.2032	0.1803	0.1558	0.2067
Isiolo	0.1467	0.1195	0.1757	0.1483	0.1216	0.1765
Kajiado	0.3372	0.2972	0.3755	0.3394	0.3012	0.3770
Kakamega	0.2573	0.2263	0.2916	0.2560	0.2248	0.2882
Kericho	0.3510	0.3127	0.3891	0.3513	0.3142	0.3889
Kiambu	0.5182	0.4776	0.5577	0.5164	0.4742	0.5565

Table C.5: County level predictions and 80% CIs for prevalence of secondary education completion for women aged 20-29 in Kenya in 2014. *(continued)*

	$SPDE_U$			$ELK\text{-}T_U$		
County	Est	Q10	Q90	Est	Q10	Q90
Kilifi	0.2461	0.2113	0.2816	0.2471	0.2119	0.2841
Kirinyaga	0.3883	0.3461	0.4285	0.3879	0.3458	0.4331
Kisii	0.3039	0.2715	0.3396	0.3055	0.2709	0.3417
Kisumu	0.3057	0.2714	0.3416	0.3041	0.2705	0.3384
Kitui	0.1631	0.1335	0.1942	0.1652	0.1352	0.1958
Kwale	0.1978	0.1696	0.2278	0.1999	0.1726	0.2298
Laikipia	0.2856	0.2480	0.3227	0.2862	0.2519	0.3237
Lamu	0.1647	0.1297	0.2011	0.1684	0.1317	0.2102
Machakos	0.3514	0.3086	0.3951	0.3515	0.3073	0.3977
Makueni	0.2821	0.2421	0.3226	0.2845	0.2448	0.3267
Mandera	0.0945	0.0659	0.1242	0.0943	0.0655	0.1256
Marsabit	0.1238	0.0981	0.1520	0.1223	0.0955	0.1514
Meru	0.2330	0.2001	0.2661	0.2345	0.2016	0.2682
Migori	0.1676	0.1398	0.1970	0.1689	0.1399	0.1973
Mombasa	0.4215	0.3761	0.4680	0.4241	0.3809	0.4665
Murang'a	0.3897	0.3465	0.4326	0.3859	0.3419	0.4289
Nairobi	0.5403	0.5082	0.5727	0.5391	0.5066	0.5706
Nakuru	0.4185	0.3780	0.4587	0.4196	0.3797	0.4607
Nandi	0.2799	0.2504	0.3097	0.2777	0.2487	0.3093
Narok	0.2091	0.1817	0.2364	0.2094	0.1837	0.2370
Nyamira	0.3278	0.2894	0.3658	0.3267	0.2855	0.3659
Nyandarua	0.3586	0.3187	0.4004	0.3599	0.3208	0.3986
Nyeri	0.4863	0.4413	0.5294	0.4878	0.4427	0.5330
Samburu	0.1101	0.0847	0.1383	0.1130	0.0831	0.1438
Siaya	0.1871	0.1609	0.2149	0.1876	0.1618	0.2150
Taita Taveta	0.3038	0.2509	0.3631	0.3040	0.2499	0.3599
Tana River	0.1026	0.0845	0.1233	0.1030	0.0842	0.1230
Tharaka-Nithi	0.3024	0.2638	0.3435	0.3037	0.2643	0.3428
Trans-Nzoia	0.2465	0.2136	0.2826	0.2473	0.2137	0.2810
Turkana	0.0707	0.0495	0.0965	0.0714	0.0504	0.0963

Table C.5: County level predictions and 80% CIs for prevalence of secondary education completion for women aged 20-29 in Kenya in 2014. *(continued)*

	$\boldsymbol{SPDE_U}$			$\boldsymbol{ELK\text{-}T_U}$		
County	Est	Q10	Q90	Est	Q10	Q90
Uasin Gishu	0.3883	0.3516	0.4239	0.3885	0.3506	0.4248
Vihiga	0.2699	0.2341	0.3053	0.2694	0.2347	0.3060
Wajir	0.0996	0.0755	0.1254	0.0982	0.0745	0.1241
West Pokot	0.1278	0.1032	0.1525	0.1291	0.1052	0.1551