Estimating Under Five Mortality in Space and Time in a Developing World Context

Jon Wakefield[1,2], Geir-Arne Fuglstad[3], Andrea Riebler[3], Jessica Godwin[1], Katie Wilson[2] and Samuel J. Clark[4]

Abstract

Accurate estimates of the under-5 mortality rate (U5MR) in a developing world context are a key barometer of the health of a nation. This paper describes a new model to analyze survey data on mortality in this context. We are interested in both spatial and temporal description, that is, wishing to estimate U5MR across regions and years, and to investigate the association between the U5MR and spatially-varying covariate surfaces. We illustrate the methodology by producing yearly estimates for subnational areas in Kenya over the period 1980–2014 using data from the Demographic and Health Surveys (DHS). We use a binomial likelihood with fixed effects for the urban/rural stratification to account for the complex survey design. Smoothing is carried out using Bayesian hierarchical models with continuous spatial and temporally discrete components. A key component of the model is an offset to adjust for bias due to the effects of HIV epidemics. Substantively, there has been a sharp decline in Kenya in U5MR in the period 1980–2014, but large variability in estimated subnational rates remains. A priority for future research is understanding this variability. In exploratory work, we examine whether a variety of spatial covariate surfaces can explain the variability in U5MR. Temperature, precipitation, a measure of malaria infection prevalence, and a measure of nearness to cities were candidates for inclusion in the covariate model, but the interplay between space, time and covariates is complex.

[1] Department of Statistics, University of Washington, Seattle
[2] Department of Biostatistics, University of Washington, Seattle
[3] Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim
[4] Department of Sociology, The Ohio State University, Columbus
Email: jonno@uw.edu

1 Introduction

Currently UNICEF estimates the under-five child mortality rate (U5MR) at the national level (which is known as Admin 0), using the Bayesian B-spline bias-reduction (B3) method (Alkema et al. 2014; Alkema and New 2014). However, subnational variation is of great interest, and has been highlighted as such in the Sustainable Development Goals (SDGs). SDG 3.2 states, "By 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to at least as low as 12 per 1,000 live births and under-5 mortality to at least as low as 25 per 1,000 live births". From https://sustainabledevelopment.un.org/post2015/transformingourworld, with reference to review processes, paragraph 74.g states, "They will be rigorous and based on evidence, informed by country-led evaluations and data which is high-quality, accessible, timely, reliable and disaggregated by income, sex, age, race, ethnicity, migration status, disability and geographic location and other characteristics relevant in national contexts."

In much of the developing world, there is limited or deficient vital registration, and estimates of U5MR are based mostly on survey and census data. In this paper, we carry out detailed analyses of such data from Kenya. Many health policies and interventions in Kenya are implemented at the Admin 1 level, which consists of 47 counties (Barasa et al. 2017), and hence it is the spatial aggregation that provides our target of inference. To estimate U5MR, we use data from the Demographic and Health Surveys (DHS). The DHS Program began in 1984 and has carried out more than 300 surveys in over 90 countries. Typically stratified, cluster sampling is carried out with information collected on population, health, HIV and nutrition. We have also carried out a detailed analysis for Malawi, using the methodology developed in this paper, but due to space limitations, we focus on Kenya, with results for Malawi being relegated to the Supplementary Materials.

We briefly review previous approaches to producing sub-national U5MR estimates. Adopting demographic notation, we define $_nq_x = \Pr(\text{ death in } [x, x+n] \mid \text{ survival to } x)$; so that U5MR corresponds to $_5q_0$, where we are using this notation on a yearly scale. Later in the paper, when defining the discrete hazards model, we shall use a monthly time scale. Note that, strictly speaking, U5MR is a probability rather than a rate. Dwyer-Lindgren et al. (2014) compare various spatial models for U5MR modeling in Zambia using DHS data. In their approach, the logit of the U5MR is modeled as normally distributed, but with a single common variance across all studies, which is clearly inappropriate since it does not acknowledge the differing effective sample sizes in each area. Computation was carried out using the integrated nested Laplace approximation (INLA) of Rue et al. (2009). Mercer et al. (2015) analyzed DHS data from 22 regions in Tanzania and assumed a likelihood in which the logit of the weighted (design) estimator was assumed to be normally distributed with variance given by the design variance. A discrete space, discrete time (5-year intervals) interaction model (Knorr-Held 2000) was used to smooth the mean of this distribution, with implementation via INLA. Pezzulo et al. (2017) model $_4q_1$ across 27 countries in sub-Saharan Africa, at the Admin 1 level. Estimation was based on the most recent DHS with the log weighted U5MR estimators assumed to be normally distributed with spatial smoothing being carried out via the model of Leroux et al. (1999). Extensive covariate modeling was carried out with potential variables being averaged within areas, and also allowing interactions by large regions (with three regions in total). As with all approaches that include covariates at the area level, the associations at the area-level cannot be transferred to the individual-level as this opens up the possibility of the ecological fallacy (Wakefield 2008).

In other contexts, methods for small-area estimation (Rao and Molina 2015) using spatial smoothing models have been proposed by a number of authors including Congdon and Lloyd (2010), You and Zhou (2011), Porter et al. (2014), Chen et al. (2014), Vandendijck et al. (2016) and Watjou et al. (2017). Notably, these approaches all utilize spatial models at the area level, whereas the model we propose models space continuously.

Burke et al. (2016) follow a different approach to modeling U5MR across sub-Saharan Africa. Kernel density estimation (KDE) is carried out with surfaces produced at a geographical scale of approximately 10km×10km. This approach follows Larmarange and Bendaud (2014) who used the same method in the context of HIV prevalence estimation. Inference, including producing uncertainty surfaces, is difficult to obtain with KDE and the approach has been found to be inferior, when considering prediction at unsampled locations, to Bayesian geostatistical modeling (Hallett et al. 2016).

More recently, Golding et al. (2017) carried out subnational estimation of U5MR for sub-Saharan Africa, with a continuously indexed spatial model. Four separate models were fitted to the age groups 0–1 months, 1–11 months, 12–35 months, 36–59 months, with the subsequent estimates being combined to give the U5MR. This combination is done by taking draws from the posteriors assuming they are independent, which is not correct, since they are based on the same children. Data from a variety of sources are included in the analysis including both full birth history (FBH) and summary birth history (SBH) data. FBH data include information for all children on the times of birth and death, if the latter occurs before the time of the survey, and these are the data we utilize from the DHS. SBH data consist of the number of children ever born, and the number who have died, along with the age of the mother. The FBH data are modeled as binomial with no explicit correction for the survey design. The SBH data are also assumed to be binomially distributed, with an artificial response and denominator created through an elaborate procedure with a heuristic justification. A space-time smoothing model is specified via the stochastic partial differential equations (SPDEs) formulation of Lindgren et al. (2011). The same space-time covariance parameters are assumed for the whole of Africa. Covariates are also modeled, and we give further details of the approach followed in Section 4. There is no adjustment for mothers lost to HIV, which can lead to serious underestimation in countries (such as Kenya and Malawi) with HIV epidemics. Estimates in each spatial grid cell are adjusted so that the national total agrees with the Global Burden of Disease (GBD) estimates. The most recent GBD (GBD 2016 Mortality Collaborators 2017) produced national estimates for 195 countries and territories over the period 1970–2016. Some of the constituent data in the study of Golding et al. (2017) do not contain GPS locations, but rather the administrative region within which the clusters were sampled. In this case, Golding et al. (2017, Supplementary Materials, Section 8) assign the data to a set of points selected within the area, where the points are obtained through k-means clustering. This approach is, at best, an approximation, since one needs to take a mixture over the likelihoods at each potential location, see Wilson and Wakefield (2017).

In this paper we develop a new continuous space/discrete time model that acknowledges the complex design by including urban/rural stratum effects. It was necessary to develop this model, because the approach of Mercer et al. (2015) requires design-based (weighted) estimates of the U5MR, with an associated standard error, for each time period and area, and as the time intervals become small, and/or the number of areas become large, the estimates and standard errors become unstable. In particular, for the Kenya data, it was not possible to implement the Mercer et al. (2015) method on a yearly scale with 47 counties. The rest of this paper is structured as follows. In Section 2 we describe the data that we use for analysis. Section 3 develops the method and gives the results for constructing the space-time child

mortality surface, while Section 4 considers covariate modeling. Section 5 concludes the paper with a discussion of ways in which we would like to extend the model.

2   Data

2.1   *Survey Data*

To estimate child mortality in Kenya, we use data from three DHS conducted in 2003, 2008–2009 and 2014. Both the 2003 and 2008–2009 Kenya DHS were designed to give reliable estimates for the 8 provinces, and for urban and rural regions separately. To this end, the sample was stratified by 8 provinces crossed with an urban/rural designation to yield 15 strata (Nairobi is solely urban). In each of these surveys the first sampling stage selected 400 enumeration areas (EAs) from a sampling frame constructed from the 1999 Census. In the second stage for both the 2003 and 2008–2009 surveys, 10,000 households were selected within the sampled EAs. The 2014 Kenya DHS was designed to make estimates of demographic indicators at the 47 county level, so it was stratified by the 47 counties crossed with urban/rural indicators. This yields 92 strata since Nairobi and Mombasa are both entirely urban. The first sampling stage of the 2014 survey produced 1,584 EAs that gave data that could be used, across the 92 strata, using a sampling frame developed from the 2009 Census. In the second stage, 40,300 households were sampled from the selected EAs. All households within the same EA are aggregated to a a single point location. Figure 1 shows the cluster locations for the three surveys along with the boundaries of the 47 counties. For confidentiality reasons, the GPS coordinates of the cluster centers are randomly displaced. Urban/rural cluster locations are displayed by up to 2km/5km; the locations of a further 1% random sample of rural clusters are displaced by up to 10 kilometers. We see that the distribution of the sampling locations is far from uniform, reflecting population density. Reported response rates for households and women are high. Such data are potentially subject to various biases, e.g., recall bias, as the birth histories may go back many years if the woman surveyed is old. Though we have data from only three survey waves, the retrospective birth history gives us data on births over the period 1980–2014.

To estimate U5MR we use the portion of the survey devoted to retrospective birth histories. Women who slept in the house the night before, and are aged 15–49 are asked to enumerate all births with dates of birth, and for children who have died, dates of death. Birth histories are converted into person months for each child in the dataset. Using a discrete hazards model, each person month yields a Bernoulli (binary) random variable, survived/dead. Hence, we implement a discrete time event history analysis. It is important to note that each unique case can result in at most one death. We would like to investigate temporal trends in U5MR (at the yearly scale), and the subnational variability in these trends across the 47 counties. Kenya provides a good test example due to the large number of clusters (1,584) sampled in the 2014 DHS. The Supplementary Materials contain extensive details on the numbers of deaths by period and county.

2.2   *HIV Adjustment*

Kenya has had a relatively high prevalence of HIV, and this can lead to serious bias in estimates of U5MR, particularly before antiretroviral therapy (ART) treatment became widely available. Pre-treatment HIV positive women had a high risk of dying, and such women who had given birth were therefore less likely to appear in surveys. The children of HIV positive women are also more likely to die before age 5
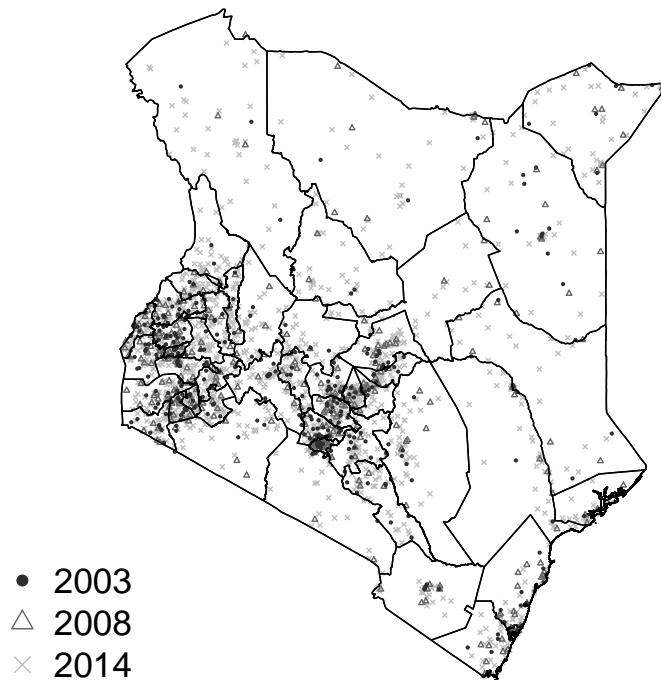
Figure 1. Cluster locations in the three DHS that we consider, with boundaries of the 47 counties.

compared to those born to HIV negative women, and therefore we expect to underestimate U5MR if we do not adjust for the missing women, i.e., the missing data are non-ignorable.

Estimates of bias may be obtained using the cohort component projection model of Walker et al. (2012). Under this model, for a particular survey, year and province (of which there are eight), the number of births is estimated, and these are attributed to HIV-negative and HIV-positive women, using estimates of the number of women in need of services to prevent mother-to-child transmission. The children born are then further subdivided into those that will and those that will not become infected with HIV, and survival probabilities of these children are then estimated, to produce a bias ratio. Let $_5q_{0l,k}(t)$ represent the true U5MR and $_5q_{0l,k}^{\star}(t)$ the biased (unadjusted for HIV) U5MR in survey k, province l and year t.

The Walker et al. (2012) method gives an estimate of,

$$\text{BIAS}_{l,k}(t) = \frac{{}_5\mathsf{q}_{0l,k}^{\star}(t)}{{}_5\mathsf{q}_{0l,k}(t)} \leq 1. \tag{1}$$

Figure 2 shows the bias ratios plotted against year for each of the three surveys, and for the 8 provinces of Kenya for which we have available data; we would prefer to have estimates at the 47 county level, but the constituent data are not available. The 47 counties are nested within the 8 provinces, which eases the application of the adjustment. We see that the ratios of reported to true rates decrease as the HIV epidemic takes hold, and then increase with the uptake of ART. Figure 3 shows maps of the ratios in 1995 (as an example year), and large between-province differences are apparent. The ratios will clearly make a significant impact on our estimates, and are included in an offset in the model we describe in Section 3. A current weakness of our approach is that we do not account for the uncertainty in the manner by which the ratios were estimated.
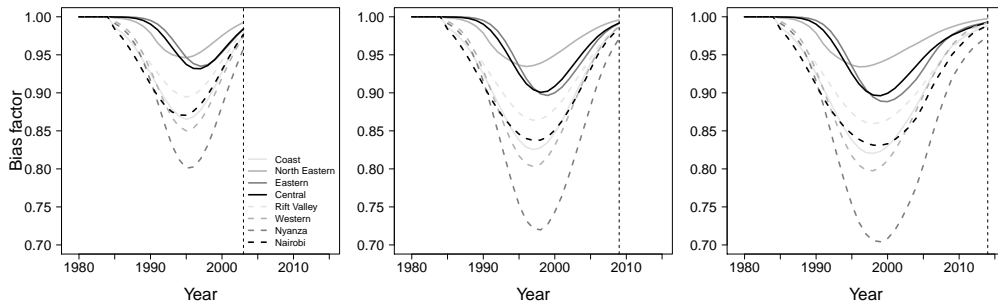

Figure 2. HIV adjustment ratios of reported U5MRs to "true" U5MRs, that is (1), by survey, over time (left is 2003, middle is 2008–2009, right is 2014), and in eight provinces. Ratios were calculated using the method of Walker et al. (2012).

3 Constructing a Space-Time Surface

3.1 The Space-Time Model

Survey data come from and describe a finite population. The DHS provides sampling weights for each individual that account for the selection probability and non-response. Skinner and Wakefield (2017) review the design and analysis of survey data. The design-based (or randomization) approach to inference is to place inference in the context of repeated sampling from the fixed finite population. The word *fixed* is key here, the data are not viewed as random, rather the indices of the units (households, in this context) within the population that are sampled are the random variables. Weighted (often referred to as direct) estimators (Horvitz and Thompson 1952) provide a design-consistent approach to estimation, but the sparsity of data in both time and space, are problematic since a greater proportion of cells with zero deaths in some age groups occur when we drill down to finer spatio-temporal units. Even with small numbers of deaths, variance estimates are unstable. The Supplementary Materials contain information
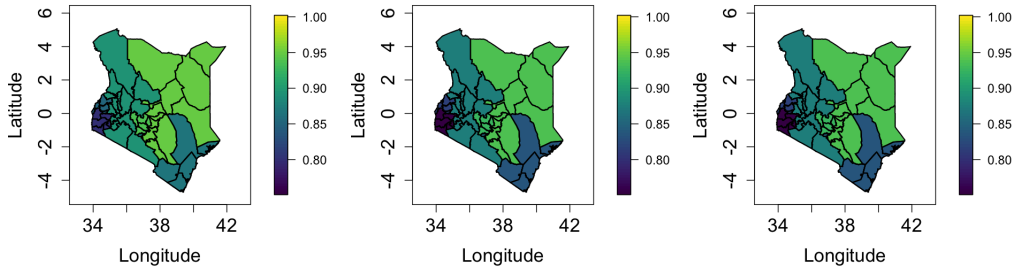
Figure 3. Maps of HIV adjustment ratios of reported U5MRs to "true" U5MRs, that is (1), by survey, in 1995. The 3 columns represent the adjustments from the 2003, 2008–2009, 2014 surveys. Ratios were obtained using the method of Walker et al. (2012).

on the standard errors of the direct estimates, by county, as a function of time period. This is a small area estimation problem and at the scale for which inference is desired, smoothing in space and time is required.

As an alternative to design-based inference, a more traditional statistical approach may be employed in which a probability model for the observations is assumed, and the mean model contains terms that reflect the design, with a carefully chosen variance model. This approach is known as *model-based inference*; Wakefield et al. (2016) compare the two approaches via simulation in a spatial context. In general, when a model-based approach is followed, the design must be acknowledged when inference is performed, otherwise biased estimates with an incorrect measure of uncertainty will be produced. As an extreme example, in the DHS, sampling is stratified by urban/rural and if in a particular county (which has both urban and rural clusters) only urban clusters were selected then ignoring this aspect will lead to bias in the estimation of the county level estimate, if U5MR is associated with urban/rural.

As in Mercer et al. (2015) we assume a discrete hazards model, with six hazards for each of the (monthly) age bands: [0,1), [1,12), [12,24), [24,36), [36,48), [48,60). Detailed argument in, for example, Allison (2014) show that the contributions for a generic child correspond to the product of up to 60 Bernoulli likelihoods with $Y_{m,k}(\boldsymbol{s}_j, t)$ being a binary indicator of survival in month m, $m = 0, \ldots, 59$, for a child in survey k, in a household sampled at location $\boldsymbol{s}_j$ in year t, $t = 1980, \ldots, 2014$, and for $j = 1, \ldots, N_k$ cluster locations in survey k. For a month beginning at m, the hazard within the next month, in survey k, and at location $\boldsymbol{s}_j$ and time t, is $_1q^\star_{m,k}$. This is the hazard that is relevant in the absence of HIV bias. Note that we have switched our demographic notation to a monthly scale. The likelihood for survival from month m to $m + 1$ in survey k and at location $\boldsymbol{s}_j$ in year t is,

$$Y_{m,k}(\boldsymbol{s}_j, t)|_1q^\star_{m,k}(\boldsymbol{s}_j, t) \sim \text{Bernoulli} \left[\,_1q^\star_{m,k}(\boldsymbol{s}_j, t) \, \right].$$

Notice that the potentially HIV biased outcomes are Bernoulli with probability of death given by the biased hazards $_1q^\star_{m,k}$. We let $a[m]$ link the month m to the six age bands a that we allow to have distinct

hazards, i.e.,

$$a[m] = \begin{cases} 1 & \text{if } m = 0, \\ 2 & \text{if } m = 1, \ldots, 11, \\ 3 & \text{if } m = 12, \ldots, 23, \\ 4 & \text{if } m = 24, \ldots, 35, \\ 5 & \text{if } m = 36, \ldots, 47, \\ 6 & \text{if } m = 48, \ldots, 59. \end{cases}$$

Then the latent logistic model we use is,

$$\text{logit}\left[\,_1 q^\star_{m,k}(\boldsymbol{s}_j, t)\,\right] = \log\left[\,\text{BIAS}_{l[\boldsymbol{s}_j],k}(t)\,\right] + \beta_{a[m]}(\boldsymbol{s}_j, t) + \eta_j + \upsilon_k + \epsilon_t \tag{2}$$

$$\beta_{a[m]}(\boldsymbol{s}_j, t) = \beta_{a[m]} + \delta_{\text{str}[\boldsymbol{s}_j]} + \phi_a(t) + u(\boldsymbol{s}_j, t). \tag{3}$$

This form consists of a collection of terms that are used for prediction, $\beta_{a[m]}(\boldsymbol{s}_j, t)$, and random effects to acknowledge the cluster sampling, survey and independent temporal effects, and an offset that adjusts for the bias due to HIV epidemics, given in (1). We now describe each of the components. More details on the HIV bias offset are given in the Supplementary Materials but, as discussed in Section 2.2, the adjustment is carried out at the province level, indexed by l, with $l[\boldsymbol{s}_j]$ corresponding to the province in which the cluster at $\boldsymbol{s}_j$ is located. The random cluster effects $\eta_j \sim_{iid} \text{N}(0, \sigma_\eta^2)$ acknowledge the cluster design and allow for dependence amongst mothers within households and between mothers in households in the same cluster (at location $\boldsymbol{s}_j$). This dependence will induce excess-binomial variation. The survey random effects $\upsilon_k \sim_{iid} \text{N}(0, \sigma_\upsilon^2)$ allow for systematic biases in each of the three surveys (though of course this is relative to the average of the three surveys, and does not correct for any overall bias in the three surveys combined). The temporal terms $\epsilon_t \sim_{iid} \text{N}(0, \sigma_\epsilon^2)$ allow for yearly perturbations that have no structure in time. Each of the six age bands, has its own intercept $\beta_{a[m]}$. The surveys are each stratified on an urban/rural indicator and on either 8 (years 2003 and 2008–2009) or 47 (year 2014) areas. The area-level stratification is strongly confounded with space and so we do not include a fixed effect for these strata, rather we assume the spatial field accounts for any such differences at a relatively large scale. The urban/rural classification changes far more quickly around urban centers, and for this reason we include a strata fixed effect $\delta_{\text{str}[\boldsymbol{s}_j]}$; within the DHS data there is an urban/rural indicator for each cluster location $\boldsymbol{s}_j$, which allows us to fit this model. The temporal terms $\phi_a(t)$ are random walks of order 2 (RW2), with one each for months [0,1) and [1,12) and then a third for the remaining period of [12,60) months. We decided on these splits based on initial analyses and on the known demographic pattern in which the majority of U5MR deaths occur in the first year of life. For each of the three RW2 models, for reasons of parsimony, the same precisions were used (we investigated the use of different precision parameters for the three age groups, but there was little difference in the resulting inference), i.e., the distribution is RW2(σ_ϕ^2) for all three age bands. Sharing the precision parameter forces the same smoothness in the temporal evolution for the logit of the hazard in each age group, but the temporal trends are independent between age groups, conditional on the precision parameter. The RW2s have sum-to-zero constraints to make them identifiable when combined with the age-group specific intercepts, $\beta_{a[m]}$. The most complex term to explain is the space-time interaction $u(\boldsymbol{s}, t)$; before describing the model we use, we give a brief description of separable processes.

A separable spatio-temporal process has a covariance function that is a combination of a spatial dependence structure, c_S, and a temporal dependence structure, c_T, through

$$c_{ST}(\,(\boldsymbol{s}_1, t_1), (\boldsymbol{s}_2, t_2)\,) = c_S(\boldsymbol{s}_1, \boldsymbol{s}_2) \times c_T(t_1, t_2), \qquad \text{for all } t_1, t_2, \boldsymbol{s}_1 \text{ and } \boldsymbol{s}_2.$$

The multiplicative structure is beneficial because it is easy to construct valid spatio-temporal covariance functions by combining valid spatial and temporal covariance functions. We want the spatial component of the separable spatio-temporal effect to have a Matérn covariance function,

$$c_S(\boldsymbol{s}_1, \boldsymbol{s}_2) = \sigma_S^2 \frac{2^{1-\nu_S}}{\Gamma(\nu_S)} \left(\sqrt{8\nu_S} \frac{||\boldsymbol{s}_2 - \boldsymbol{s}_1||}{\rho_S} \right) K_{\nu_S} \left(\sqrt{8\nu_S} \frac{||\boldsymbol{s}_2 - \boldsymbol{s}_1||}{\rho_S} \right),$$

where ρ_S is the spatial range corresponding to the distance at which the correlation is approximately 0.1, σ_S is the marginal standard deviation, ν_S is the smoothness, and K_{ν_S} is a modified Bessel function of the second kind, order ν_S. In our model, the Matérn spatial structure is approximated via a SPDE, and combined with an AR(1) process in time. Inference is done using INLA with samples drawn from the approximate posterior for inference on functions of interest. The process is written as $u(\boldsymbol{s}, t)$ and is a combination of a temporal structure c_T and a spatial structure, c_S which translates to,

$$\Sigma_{ST} = \Sigma_T \otimes \Sigma_S,$$

if the process is observed on $(\boldsymbol{s}, t) \in \{\boldsymbol{s}_1, \ldots, \boldsymbol{s}_N\} \times \{1, 2, \ldots, T\}$ (in which case Σ_S is $N \times N$, Σ_T is $T \times T$ and Σ_{ST} is $NT \times NT$).

The hazard for each age group is expected to vary spatially, but due to data sparsity the data will not support separate spatial main effects for each of the six age bands. A parsimonious model would include a shared spatial main effect for all age groups, but since a spatio-temporal interaction is necessary to account for the yearly changes in the spatial pattern, we do not include a separate spatial main effect. It is too expensive to apply the necessary temporal sum-to-zero constraints that would be required to give identifiable spatial main effects alongside a spatio-temporal interaction. Therefore, the shared spatio-temporal interaction is handled with a separable spatio-temporal model that combines an AR(1) structure with the Matérn covariance function, with the smoothness parameter fixed. The resulting spatio-temporal covariance function can be explained through a constructive example which gives some intuition on the space-time interaction. A stable AR(1) process with marginal variance 1 can be generated by

$$a_{t+1} = \rho a_t + \epsilon_t, \qquad t = 2, 3, \ldots, T,$$

where $\epsilon_t \sim_{iid} N(0, 1 - \rho^2)$, for $t = 2, \ldots, T$, and $a_1 \sim N(0, 1)$. The temporal process can be made spatio-temporal by replacing the starting condition and the innovations with spatial Matérn fields, to give

$$a_{t+1}(\boldsymbol{s}) = \rho a_t(\boldsymbol{s}) + \epsilon_t(\boldsymbol{s}), \qquad t = 2, 3, \ldots, T,$$

for all $\boldsymbol{s} \in \mathbb{R}^2$, where $\epsilon_t \sim N(0, (1 - \rho^2)c_S(\cdot))$, and $a_1 \sim N(0, c_S(\cdot))$, where c_S is the stationary Matérn covariance function. Hence, a proportion ρ^2 of the marginal variance is explained by the previous time step and a proportion $1 - \rho^2$ is arising from a new realization of a spatial field.

The joint identifiability of the three temporal trends and the spatio-temporal interaction can be achieved through integrate-to-zero constraints for each year. This integration is carried out with respect to the

spatially varying population density $d(\boldsymbol{s})$:

$$\int u(\boldsymbol{s},t)d(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s} = 0, \qquad t = 1980,\ldots,2014,$$

where $u(\boldsymbol{s},t)$ is the separable spatio-temporal process, and $d(\boldsymbol{s})$ is the population density for 2014. These yearly integrate-to-zero constraints give a weighted spatial average of the spatio-temporal effect that is constantly equal to zero and also mean that the temporal change in the weighted spatial average of the logits of the hazards of each age group is explained by the corresponding temporal main effects. In particular, the RW2 trends are approximately interpretable as the change in the national level with time. Further details on the integrate-to-zero constraint are given in the Supplementary Materials.

This spatio-temporal effect on a temporal resolution of 35 years is too computationally expensive to include in the SPDE implementation of the Bayesian model, but since we want the spatio-temporal process to change gradually in time, it is possible to use an approximation that changes piecewise linearly in time; a similar approach was taken in Blangiardo and Cameletti (2013, Chapter 8). We decrease the resolution of the spatio-temporal process to 8 time steps by defining $\widetilde{u}_h(\boldsymbol{s})$ for knot locations $h = 1,2,\ldots,8$, corresponding to years $1980, 1985,\ldots,2015$, and defining

$$u(\boldsymbol{s},t) = (1 - \alpha_h(t))\widetilde{u}_h(\boldsymbol{s}) + \alpha_h(t)\widetilde{u}_{h+1}(\boldsymbol{s}), \qquad \text{for } 1975 + 5h \le t < 1980 + 5h,$$

where $\alpha_h(t) = t/5 - \mathrm{floor}(t/5)$ gives the factor required for linear interpolation between the two knot locations. The number and placement of knots is context specific and is chosen to make the computation manageable. Note that if the integrate-to-zero constraint is satisfied for $\widetilde{u}_h(\boldsymbol{s})$ for $h = 1,2,\ldots,8$, the integrate-to-zero constraint is also satisfied for linear combinations $u(\boldsymbol{s},t)$ for $t = 1980, 1981,\ldots,2015$.

Each of the precisions for the independent and identically distributed effects, σ_η^{-2}, σ_υ^{-2}, σ_ϵ^{-2}, have Gamma$(0.5, 5 \times 10^{-4})$ priors (which give 5%, 50%, 95% quantiles for the standard deviations of 0.016, 0.047, 0.52). The spatial part of the spatio-temporal interaction has fixed smoothness $\nu_\mathrm{S} = 1$ and a "penalized complexity" prior (Fuglstad et al. 2018; Simpson et al. 2017) for the spatial range ρ_S and the marginal standard deviation σ_S, where the hyperparameters are selected so that $\Pr(\rho_\mathrm{S} < 0.5) = 5\%$ and $\Pr(\sigma_\mathrm{S} > 3) = 5\%$. The remaining parameters have the default priors in INLA; the autocorrelation parameter of the AR(1) in the temporal part of the spatio-temporal interaction has the prior $\log((1+\rho)/(1-\rho)) \sim \mathrm{N}(0, 0.15)$ and the marginal variance of the RW2 has the prior $\sigma_\phi^{-2} \sim$ Gamma$(1, 5 \times 10^{-5})$. The 5% quantile of the prior for spatial range is 0.5 degrees, which corresponds to 6% of the spatial extent of Kenya in the north-south direction. This allows spatial variation on the country scale, but also allows the resolution of the SPDE model to be chosen to be low enough to make the complex spatio-temporal model computationally feasible.

For predictions, the cluster, survey and temporal independent and identically distributed effects in (2) are not included so that the only contribution is $\beta_a(\boldsymbol{s}_j, t)$. The survey random effects υ_k are bias terms and their non-inclusion is uncontroversial. The independent temporal terms ϵ_t represent one-off "shocks" and it is not so clear whether or not they should be included, since they may correspond to true adjustments due to particular conditions in year t (in which case we would include), or to measurement problems in that year (in which case we would not include). On examination of predictions under both scenarios, we decided to not include since the predictions including ϵ_t were very jagged. The predicted U5MR at

location s and at time t is,

$$\text{U5MR}(s, t) \quad = \quad 1 - \prod_{a=1}^{6} \left[\frac{1}{1 + \exp[\beta_a(s, t)]} \right]^{z[a]},$$

where $z[a] = 1, 11, 12, 12, 12, 12$, for $a = 1, \ldots, 6$ and with $\beta_a(s, t)$ given by (3).

The data and the fitted model are on a continuous spatial scale, but the aim is to produce values on a discrete scale using the 47 administrative regions. To construct the predictive spatial surfaces over time we use the posterior of the spatially-temporally varying U5MR and the population density $d(s)$. We obtained the latter from `worldpop.org` (Linard et al. 2012). We would prefer to use births density, but such data are difficult to obtain; we examined a surface of estimated live births for one year that was available (WorldPop 2017), and inference using this birth surface showed little difference to inference using the population density surface. We define the U5MR of region i by

$$\text{U5MR}_i(t) = \frac{\int_{R_i} \text{U5MR}(s, t) d(s) \, \mathrm{d}s}{\int_{R_i} d(s) \, \mathrm{d}s}, \quad i = 1, 2, \ldots, 47, \tag{4}$$

where R_i denotes administrative region i. This averaging gives zero weight to areas with no population, even though the continuous surface is defined at such points. We also need to assign each location to urban/rural, since we have a fixed effect in the model corresponding to this dichotomy. For this purpose, we used the urbanicity map described in Pesaresi et al. (2016).

3.2 Constructing a Space-Time Surface Results

We begin by reporting inference on some of the key components of the model, before reporting on substantive summaries. We also fitted a model with no HIV bias adjustment, and the left panel of Figure 4 shows the posterior medians of the RW2 median fits for each of the [0,1), [1,12), [12,60) age groups (specifically, $\exp[\phi_a(t)]$ reveals how hazard odds ratios evolve by year, t), along with 95% point-wise credible interval envelopes. We emphasize that these are hazards *odds ratios* and so the three curves are not comparable, since they are are relative measures. The right panel of Figure 4 shows the HIV adjusted version of this plot and the effect of the epidemic is clear to see in all three age groups. It is clearly important to include an HIV adjustment. We see that over 1980–2014, the temporal trend decreases for all three age groups. While the [0,1) age group shows a very shallow decreasing slope from the late 1990s, a much steeper decrease can be seen for the other two age groups from around 1995, with the most prominent drop being for the [12,60) month age group. There are many potential reasons for this, see Liu et al. (2017) for a discussion of the specific causes that contribute to under-5 mortality in neonatal and non-neonatal children.

Table 1 gives posterior summaries of key parameters in the space-time model. The standard deviations are not all comparable since for the RW2 the standard deviation is conditional while the other (IID and spatio-temporal) terms are marginal. The spatio-temporal standard deviation is relatively large indicating that there are strong spatial effects for the Kenya data; the median of the range parameter is $1.77°$, which is quite large (about a fifth the size of the study region). There is also strong year-to-year correlation in the AR(1) model. The hazard odds is estimated as 8% greater in rural versus urban locations, all else being equal.
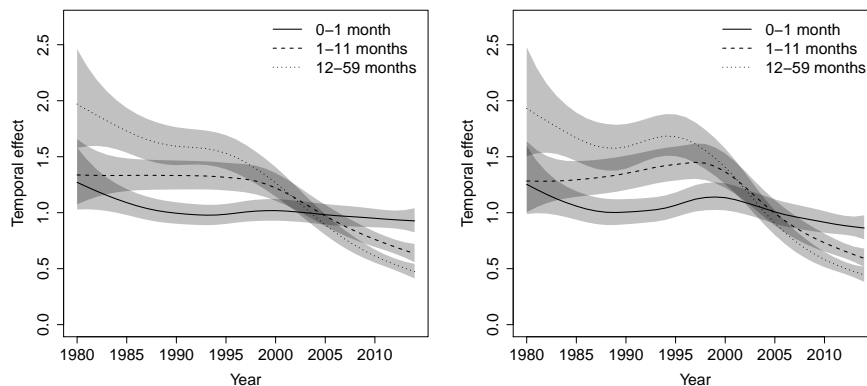
Figure 4. Left: Median RW2 model temporal trends (left) HIV adjusted time trends (right) for the three age bands. Both with 95% pointwise credible intervals. The trends are on the odds ratio scale.

Parameter	2.5%	50%	97.5%
Standard deviation for RW2 time	0.0089	0.017	0.032
Standard deviation for IID-time	0.024	0.050	0.11
Range for spatio-temporal effect	1.29	1.73	2.40
Standard deviation for spatio-temporal effect	0.48	0.57	0.69
AR(1) parameter for spatio-temporal effect	0.78	0.86	0.93
Standard deviation for IID-cluster	0.32	0.36	0.39
Standard deviation for IID-survey	0.019	0.044	0.11
Effect of rural versus urban	1.01	1.08	1.16

Table 1. Posterior quantiles for model parameters.

Figure 5 shows a comparison between the modeled U5MR and weighted estimates at the 47 county level, and aggregated over 5 years (aggregation over years is required, otherwise the direct estimates are unstable). The weighted estimates in a particular area and time period are based on data from all surveys that were collected in those areas/time periods; the way we combine the data from different surveys and make the HIV adjustment, is described in the Supplementary Materials. We see some attenuation of the modeled estimates due to shrinkage, as expected. In the Supplementary Materials we include more detailed plots and show the uncertainty in the modeled and weighted estimators. These plots show that, again as expected, the modeled estimates have much greater precision.

As mentioned in Section 1, we wish to make inference at the spatial level at which policy interventions occur. For Kenya, this is at the 47 county level, and Figure 6 shows a sequence of 9 maps of U5MR for the years 1980, 1985, ..., 2015, 2020 (we have 35 yearly estimates, but for space reasons we look at estimates 5 years apart). The last two of these years are obtained by forecasting from the model. On these plots, the hatching shows the size of the standard deviation relative to the value of the estimate measured
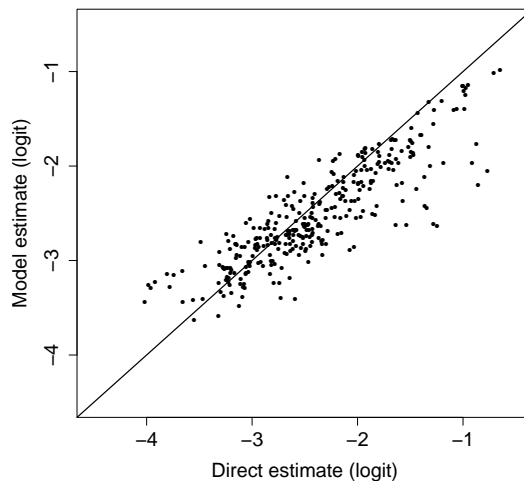
Figure 5. Modeled estimates versus weighted (direct) estimates on the logit scale.

in percent, i.e., std.dev./median $\times\ 100\%$. The dramatic decrease over time in U5MR since around 1995 is apparent, though strong subnational variation persists. The Supplementary Materials contain maps of the uncertainty.

Figure 7 shows the posterior medians of the spatio-temporal terms $\exp[u(s,t)]$ for the years 1980, 1985, 1990, . . . , 2015, 2020. The last two of these years are obtained by predicting forward the space-time field. From 1980 onwards strong spatial effects can be seen in the counties Turkana and West Pokot in the north west part, the province Nyanza in the middle west part and the counties Kilifi, Tana River and Garissa in the south east part of Kenya. While the highs in the north west and south east have almost disappeared by 2004, a higher effect in the counties Migori and Homa Bay of the Nyanza province persist and, without interventions, one would expect these trends to continue until 2020. In the period 1990 to 1995 higher effects can also be seen in the north east.

While it might appear that the spatio-temporal variability is decreasing over time it should be emphasized that there is still strong variability present across the map in recent periods and also in the future. To illustrate this we computed the 95% and 5% quantiles of the posterior medians across pixel values for each of the nine maps. Figure 8 summarizes the spatial heterogeneity over time. In 1980, the 95% quantile was 2.2 and the 5% quantile was 0.63 leading to a ratio of 3.4. While the 5% quantile decreases until 2005 and then increases again, the 95% quantile decreases almost constantly. The ratio of 95% to 5% points increases until 1995 with a value of 4.4 and then decreases. However, in 2010 the ratio is still 3.5, with ratios of 3.0 and 2.6 in the (predicted) years of 2015 and 2020, respectively. In summary, there remains strong subnational heterogeneity in U5MR in Kenya; further discussion will be given in Section 3.3.
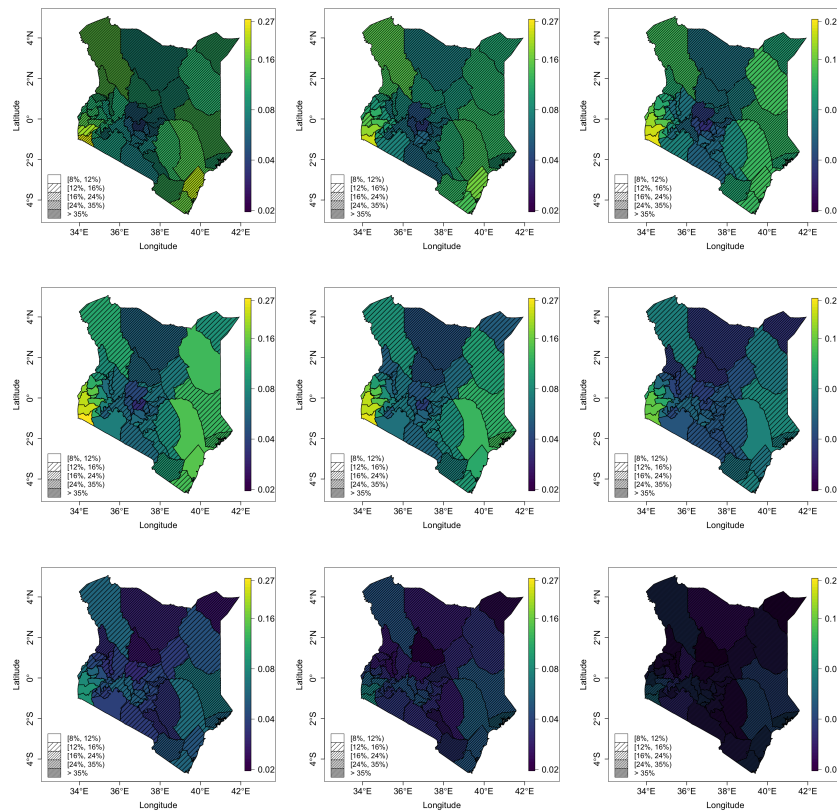
Figure 6. Maps of the posterior median estimates of U5MR at the county level, with uncertainty represented by hatching. Top row: 1980, 1985, 1990. Middle row: 1995, 2000, 2005. Bottom row: 2010, 2015, 2020.

The Millennium Development Goals (MDG) aimed for a drop of 67% in U5MR between 1990 and 2015. In the left hand panel of Figure 9 we map the posterior median of the percentage drop at the county level. Counties in the central part of Kenya experienced very small decreases only. In the right hand panel we plot the posterior probability that each county achieved this aim and we see that very few attained a 67% drop. Over the country as a whole, the posterior median drop was 55% with 95% credible interval of (45%, 61%), and a 0% probability that the 67% drop was achieved.

To examine the accuracy of the space-time smoothing model, we held out some of the data and then predicted the U5MR at these left-out points, using weighted and smoothed estimates. Specifically, we calculated estimates of U5MR for all counties and periods from the model using all the 2003 and 2008–2009 DHS, along with 397 clusters from the 2014 DHS. We then calculated weighted estimates of U5MR using the remaining 1,187 clusters, and these are treated as the target, since they are based on a relatively large sample. Due to stability of the weighted estimates we look only at the periods 1990–1994, 1995–1999, 2000–2004, 2005–2009 and 2010–2014, and form estimates for each of the 47 counties in these
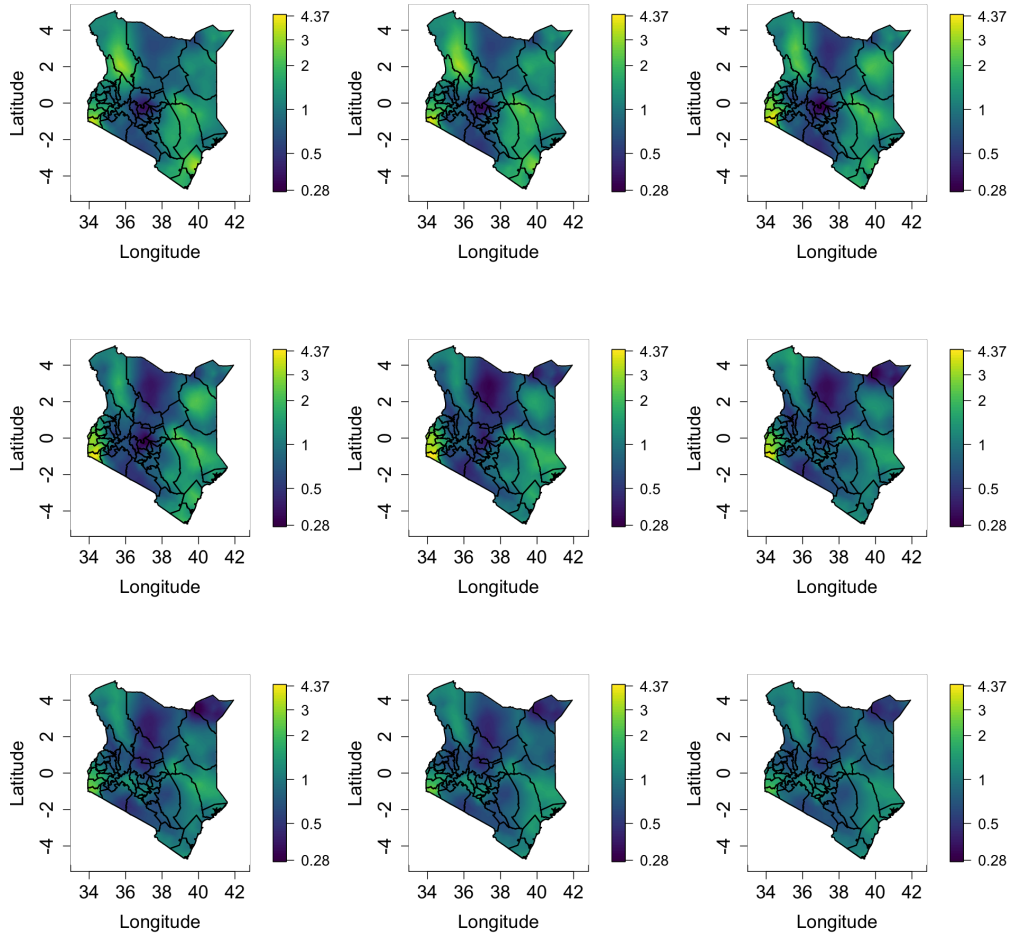
Figure 7. Maps of the spatio-temporal odds surface, $\exp[\,u(\boldsymbol{s}, t)\,]$. Top row: 1980, 1985, 1990. Middle row: 1995, 2000, 2005. Bottom row: 2010, 2015, 2020.

periods. For county i and period p, we let $Y_{ip}^{(1)}$ denote the weighted estimator (on the logit scale) and $Y_{ip}^{(2)}$ the smoothed estimator from our continuously-indexed spatial model.

We also calculate predictions using a model that is the discrete spatial analog of the continuously-indexed spatial model described in Section 3.1. Hence, the likelihood is a product of Benoulli's with a HIV adjustment, six age-specific intercepts, independent and identically distributed random effects for cluster, survey and time, a fixed effect for urban/rural, three RW2 models for yearly time for the three age bins and a space-time interaction model that replaces the SPDE (continuous space) model with an
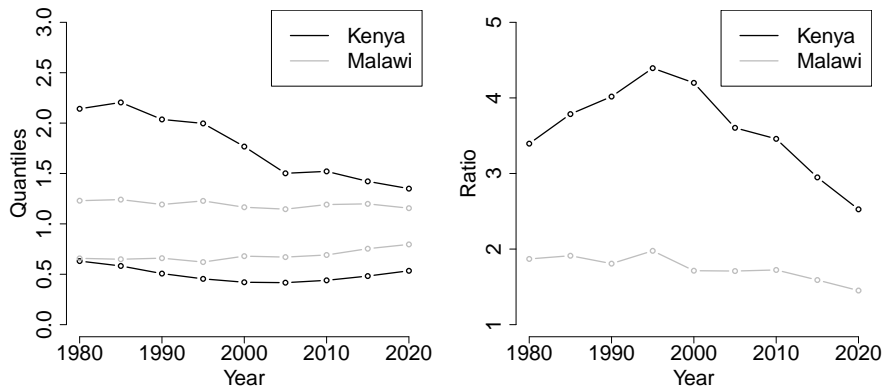
Figure 8. Left plot: 5% and 95% quantiles of pixel map of the posterior medians of the spatio-temporal effect. Right plot: ratio of 95% to 5% quantiles. The values are computed for years $1980, 1985, \ldots, 2020$.
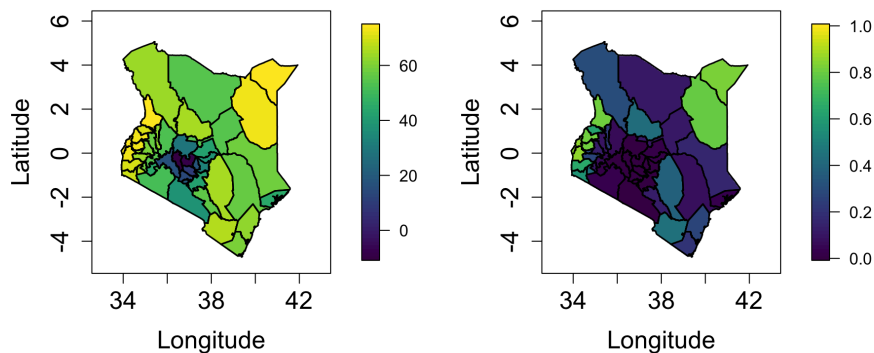

Figure 9. Left plot: Posterior median of $100 \times [\, \mathsf{U5MR}_i(1990) - \mathsf{U5MR}_i(2015)\,]/\mathsf{U5MR}_i(1990)$; that is, the percentage drop in U5MR for each county over the period 1990–2015. Right plot: Posterior probability that county i achieved a 67% drop over 1990–2015, $i = 1, \ldots, 47$.

ICAR model. With respect to the latter, we therefore have an ICAR spatial model (Besag et al. 1991) at the first time point and this then contributes to the next time point, via an AR(1) model, with the addition of a new ICAR contribution. This space-time interaction model is defined on 8 rather than 35 time steps, as in Section 3.1. The estimators from this model will be denoted $Y_{ip}^{(3)}$.

The three sets of estimates are compared with the weighted estimates of the logit of U5MR from the 1,187 clusters, y_{ip}. As a summary of the accuracy we calculate,

$$\text{MSE}_p^{(m)} = \frac{1}{47} \sum_{i=1}^{47} \left(Y_{ip}^{(m)} - y_{ip} \right)^2, \tag{5}$$

for $p = \{1990\text{--}1994, 1995\text{--}1999, 2000\text{--}2004, 2005\text{--}2009\}$ and $m = 1, 2, 3$. Table 2 presents the MSEs. For 4 out of 5 of the periods considered the weighted (direct) logit estimates could not be calculated in some counties (the numbers are listed in the caption), due to zero estimates. We see that in all cases the Bayesian spatial models have far superior performance in terms of MSE. The discrete and continuous models that we have developed in this paper gave the same MSEs in all periods, up to two decimal places. However, these summaries hide differences in the underlying estimates and the continuous smoothing model can give estimates at lower geographies than county, if required, though as emphasized above, such estimates should be judged cautiously.

Period	Weighted	Continuous Space	Discrete Space
1990–1994	49	29	29
1995–1999	46	21	21
2000–2004	40	22	22
2005–2009	41	20	20
2009–2014	37	15	15

Table 2. Mean-squared errors $(\times 10^2)$ comparing weighted and spatially and temporally smoothed estimates, via (5). Over the five periods considered, from earliest to latest, 8, 2, 1, 0, 1 counties were excluded from the relevant calculation due to unusable direct estimates.

3.3 Analyzing Multiple Countries

The spatio-temporal model can be applied separately to multiple countries to obtain estimates for each country, and to perform within-country and between-country comparisons. We demonstrate by using Malawi and its four DHS from 2000, 2004, 2010 and 2015. The details are given in the Supplementary Material, and show that, as for Kenya, the MSE of the estimates from the continuously-indexed spatial model are considerably lower than for the direct estimates; there is also a modest improvement over the BYM formulation. The spatio-temporal component can be used to examine the temporal evolution of spatial inequality across and between countries. As reported in Table 2, we compute the 5% and 95% quantiles of the posterior medians of the spatio-temporal effects at the pixels within Malawi. Figure 8 demonstrates that Kenya has stronger spatial inequality than Malawi and that there is larger temporal change in the spatial inequality for Kenya than for Malawi.

4 Exploratory Covariate Modeling

4.1 The Covariate Model

In this section we carry out an exploratory investigation into whether any of the spatial variability we see in Kenya can be attributed to a variety of covariates that we have acquired. Mosley and Chen (1984)

attempted to bring together medical and social sciences research, in order to provide a framework for child survival. A key element of this framework is the identification of a set of proximate determinants that directly influence the U5MR. Mosley and Chen (1984) list five categories of proximate determinants: maternal factors (age, parity, birth interval), environmental contamination (air, food/water/fingers, skin/soil/inanimate objects, insect vectors), nutrient deficiency (calories, protein, micronutrients), injury (accidental, intentional), and personal illness control (personal preventive measures, medical treatment). Socioeconomic determinants influence these proximate determinants.

At this point, we comment briefly on the roles and limitations of different kinds of spatial modeling in this context. We can distinguish between *individual* and *ecological* modeling. In the former, one may directly estimate the associations with proximate determinants. In an ecological setting, we are in a very different situation as there is no individual adjustment for these determinants, but instead we introduce area (or cluster) level variables which are proxies for proximate or socioeconomic variables. In an ecological study for a complex outcome such as U5MR, one will not have a hope of getting close to mimicking individual-level associations, due to ecological bias (Wakefield 2008), but if the areas are not too large, and if the input variables are well measured, then one may find variables that can aid in predicting area-level U5MR. It is this latter setting that we are in. If we wish to obtain predictions for unobserved locations on the basis of a covariate model, then those covariates must be available

In a comprehensive analysis of DHS data from 10 West African countries, Balk et al. (2004) carried out individual level modeling and fitted a range of models that included child and mother demographics, household characteristics and spatial characteristics that included urban/rural, population density, rainfall, distance to coast and a farming variable. Models were fitted for both $_1q_0$ and $_4q_1$ but these models could not be used for prediction, since the variables were not universally available spatially. Distance to coast was strongly associated with U5MR for both $_1q_0$ and $_4q_1$. Tottrup et al. (2009) also carried out a district-level analysis of U5MR in Tanzania, using census data and various spatial covariates including vegetation greeness, elevation, proportion of maternal orphans. Variables in a linear model for U5MR were selected using stepwise methods. A linear model with constant variance is not appealing as a model for U5MR.

Before outlining our approach to covariate modeling, we provide a brief literature review of suggestions for building covariate models in the setting considered here. Gething et al. (2015) describe the use of DHS data to construct surfaces of: access to HIV testing in women, stunting in children, anemia prevalence in children and access to improved sanitation. For each outcome and each country the following procedure was carried out. A collection of 17 covariates were examined. Initially, simple linear regression was used taking three versions (the original, the square and the square root) of each of the 17 variables. Cross-validation was then used to reduce these to a subset of 17 terms. Two-way interactions for these 17 were added to the collection to give $289 = 17 \times 17$ additional terms. This complete set was reduced to 20, again via cross-validation. Then the resultant potential $2^{20} - 1$ models, that were combinations of these 20 terms, were compared.

Bhatt et al. (2017) use an approach known as stacked generalization (Wolpert 1992) in which multiple predicting algorithms are weighted to produce a final prediction. This approach is closely related to the more general super-learner approach (Van der Laan et al. 2007). This approach has optimality properties for prediction but has a lack of interpretability, and the model is not suitable for predictions into the future. There is also no way that uncertainty in the estimation procedure can be incorporated into interval estimates for the surface. A similar approach was used by Golding et al. (2017).

The variables that we selected for examination were: access (estimated travel time to cities with at least 50,000 people; Nelson 2008), aridity (Zomer et al. 2007, 2008), precipitation (Fick and Hijmans 2017), temperature (Fick and Hijmans 2017), enhanced vegetation index (EVI; Didan 2015) and the *Plasmodium falciparum* parasite rate (*Pf*PR) in children (Bhatt et al. 2015). The rationale for including access is that it is thought to be related to availability of health services and improved public health infrastructure (for example, clean drinking water). Population density can play a role in the transmission of infectious diseases and is also related to access of resources (Root 1997). The climatic variables (aridity, precipitation, temperature, and vegetation) may affect vector-borne disease transmission and food production, which influences malnutrition. Malaria transmission has been previously shown to explain mortality especially in Eastern Africa (Root 1999).

Further details on these covariates can be found in the Supplementary Materials, including the sources and the spatial resolution. For the purposes of exploration, we model access, aridity, temperature, and precipitation as time-invariant; plots of these variables can be found in the left column of Figure 10. Data on *Pf*PR, population, and vegetation were obtained for the years 2000–2014 and subsequently averaged within each of the three 5-year periods (2000–2004, 2005–2009 and 2010–2014) to obtain values for each period; these data are also displayed Figure 10.

In order to determine which covariates are predictive of U5MR, we will use a simplified version of the model described in Section 3.1, in which we replace the yearly model with a model over 5-year periods $p = \{2000\text{–}2004, 2005\text{–}2009, 2010\text{–}2014\}$. The model is,

$$\beta_{a[m],k}(\boldsymbol{s}_j, p) = \beta_{a[m]} + \delta_{\text{str}[\boldsymbol{s}_j]} + \gamma_p + \eta_j + \upsilon_k + \text{Other Variables}, \tag{6}$$

where $\beta_{a[m]}$ are three age-specific intercepts (0–1 months, 1–12 months, and 12–60 months), $\delta_{\text{str}[\boldsymbol{s}_j]}$ are stratum (fixed) effects, γ_p is a temporal random effect (assumed common to all age groups) and is modeled using a RW1 (rather than a RW2, since we have three periods only), $\eta_j \sim_{iid} N(0, \sigma_\eta^2)$ are cluster random effects and $\upsilon_k \sim_{iid} N(0, \sigma_\upsilon^2)$ are survey random effects. We used three age-specific intercepts, rather than the six we used in Section 3, in order to reduce computation, since in this exercise hundreds of models are being fitted. In comparisons to be presented in Section 4.2 we compare six different approaches/models: M_1 refers to the direct estimates, M_2–M_6 correspond to choosing the "Other Variables" in (6) to be a period-invariant spatial surface (M_2), a period-invariant spatial surface and covariates (M_3), a period-varying spatial surface (M_4), a period-varying spatial surface and covariates (M_5), and covariates only (M_6). To summarize:

$$S(\boldsymbol{s}_j) \qquad\qquad M_2$$
$$\boldsymbol{\beta x}(\boldsymbol{s}_j, p) + S(\boldsymbol{s}_j) \qquad\qquad M_3$$
$$S_p(\boldsymbol{s}_j) \qquad\qquad M_4$$
$$\boldsymbol{\beta x}(\boldsymbol{s}_j, p) + S_p(\boldsymbol{s}_j) \qquad\qquad M_5$$
$$\boldsymbol{\beta x}(\boldsymbol{s}_j, p) \qquad\qquad M_6$$

where $\boldsymbol{x}(\boldsymbol{s}_j, p)$ are the spatial covariates at location $\boldsymbol{s}$ and in period p, $S(\boldsymbol{s}_j)$ is a spatial random effect at a cluster with location $\boldsymbol{s}_j$, and $S_p(\boldsymbol{s}_j)$ is a spatial random effect at cluster with location $\boldsymbol{s}_j$ in period p. The spatial model is, as before, a Gaussian Markov random field with Mátern covariance function (fitted using the SPDE approach) and, for simplicity, we assume it has the same structure for every age
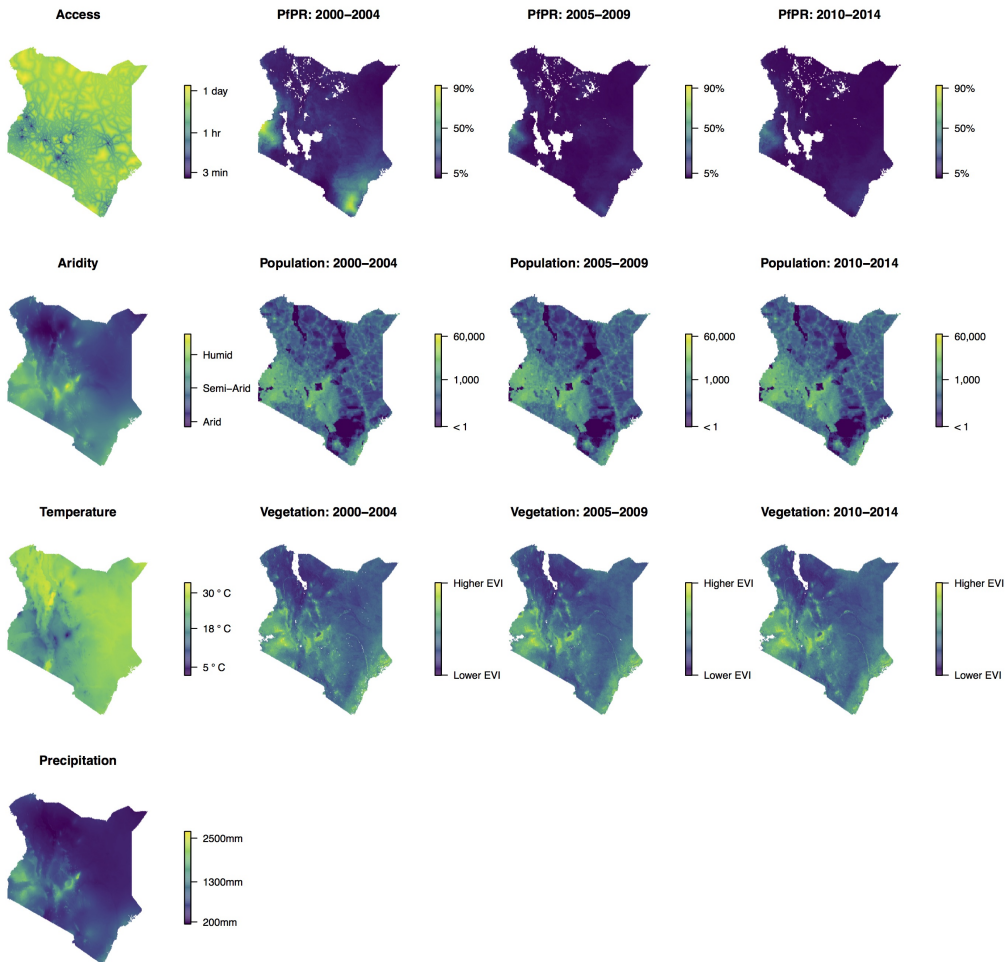
Figure 10. Left column: maps of time-invariant spatial covariates in Kenya. Columns 2–4 represent 5-year maps for PfPR (row 1), population (row 2) and vegetation (row 3). Access, aridity, and population have been log-transformed for presentation purposes. The units for population are number of people per 5km × 5km area. All time points are on the same scale for each variable.

group. For M_2 and M_3 we assume it is the same for every period and for M_4 and M_5 we only assume the spatial range and standard deviation is the same across all periods. We divide the data into training and test sets. In the training set we build the models and in the test set we compare their performance. We split the 2014 DHS into two, roughly equal-sized, groups. We use 799 clusters from the 2014 DHS as our test set (for comparison purposes). The other clusters in the 2014 DHS along with data from the 2003 and 2008–2009 DHS will be used for training the model, resulting in 1,581 clusters being used.

To emphasize, the spatial models, M_2 and M_4, are fit just once, while M_3, M_5, and M_6 are fit multiple times, for each combination of covariates. For these models, we assess their performance using the DIC (Spiegelhalter et al. 1994), CPO (Held et al. 2010), and WAIC (Watanabe 2013) criteria. As a result of this exercise performed on the training clusters, we determine the best models in each of the M_3, M_5 and M_6 collections to be used to compare with the direct estimates M_1 and the spatial only models, M_2 and M_4. We will have a total of six final comparisons (with all estimates based on the training data): M_1 direct estimates, M_2 a model with a "fixed" spatial random effect, M_3 a model with a "fixed" spatial random effect and the "best" collection of covariates, M_4 a model with a period-varying spatial random effect, M_5 a model with a period-varying spatial random effect and the "best" collection of covariates, and M_6 a model with the "best" collection of covariates when no spatial effects are added.

Under M_j we have an estimator of the logit of U5MR for each area i and period p, $Y_{ip}^{(j)}$. Under model M_1, the direct estimator has normal distribution $N(\widehat{Y}_{ip}^{(1)}, V_{ip}^{(1)})$, and under M_2–M_6, we have posterior distributions with posterior means $\widehat{Y}_{ip}^{(j)}$ and posterior variances $V_{ip}^{(j)}$, $j = 2, \ldots, 6$. Then, with the "truth" (direct estimate from test data) y_{ip},

$$\mathrm{MSE}_{ip}^{(j)} = \mathrm{E}\left[\left(Y_{ip}^{(j)} - y_{ip}\right)^2\right] = \mathrm{E}[\widehat{Y}_{ip}^{(j)} - y_{ip}]^2 + \mathrm{var}(Y_{ip}^{(j)}).$$

The best approach is that which minimizes the MSE.

4.2 Exploratory Covariate Modeling Results

The DIC, CPO and WAIC scores for all possible covariate combinations for models M_3, M_5, and M_6 are reproduced in the Supplementary Materials. There is good agreement between the three different assessments of model fit for M_3 and M_5. For M_3 ("fixed" spatial effect with covariates), the best model was that which included access, precipitation and *Pf*PR. For M_5 (period-varying spatial effect with covariates), the model that included temperature, *Pf*PR and access performed best. For M_6 (covariates only), WAIC and DIC suggested the model that included temperature, precipitation and *Pf*PR and aridity, while CPO suggested the model with temperature, precipitation and *Pf*PR only.

The MSE and constituent squared bias and variance are shown in Figure 11. We see that M_2, M_3, M_4, M_5, and M_6 perform much better than M_1, with M_5 performing (marginally) better than the others. In the summary figures we report, we take the version of M_6 that was suggested by WAIC and DIC, though results are similar for the CPO version, with the model including aridity have a slightly lower MSE. We see in Figure 12 that the predicted surfaces are almost identical under models M_2–M_6. Somewhat surprisingly, the spatial standard deviation and range parameters did not change much with the addition of covariates when the spatial surface was fixed; there was a larger change when the spatial surface was different for each period (see Supplementary Materials). The strength of the associations between the covariates and outcome did not change greatly when spatial components were added, except for precipitation where the association tended to be positive when no spatial effect was included and negative when a time-invariant spatial effect was added (see Supplementary Materials). As expected, there is always a strong positive association between the logit of U5MR and log *Pf*PR.

We conclude from this exercise that the covariates we have investigated add little predictive power at the 47 county level to the space-time models. This is disappointing, and consistent with the results
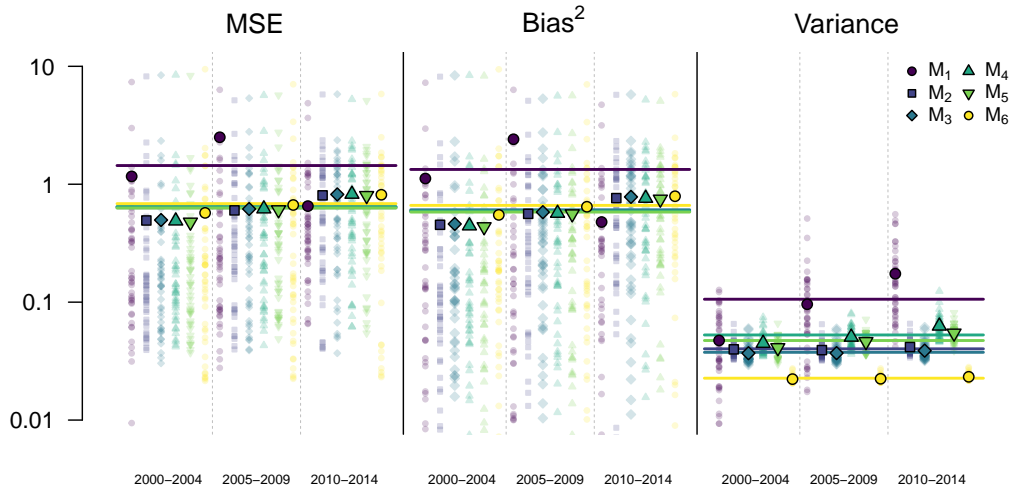
Figure 11. Plot of MSE broken down into Bias^2 and variance terms for the logit of U5MR. Color coded by model. Horizontal lines indicate the value average over all years. Larger, darker points indicate the average of the 47 admin regions. Note, the y-axis has been transformed and truncated so not all individual values are shown.

of Golding et al. (2017), but we believe that continued examination of spatial covariates is warranted, though the quality and relevance of potential covariates should be critically evaluated.

5 Discussion

In this paper we have developed a continuous space/discrete time model for investigating the dynamics of U5MR in a developing world setting. We have illustrated that the model improves on the use of weighted estimates, and can provide reliable inference at the required geographical scale. As a further illustration of the model's applicability we have included in the Supplementary Materials a parallel analysis of data in Malawi, and find similar behavior of the model, and in particular its superiority (in terms of MSE) to the use of weighted estimates. The potential for between-country comparisons based on the estimated model components is an advantage of the modeling framework that we have described in the paper. In particular, spatial inequality can be examined via the estimated spatio-temporal component; differences in temporal change from the estimated RW2 component; unexplained variation from the nugget variance, etc. Figure 8 gives a hint of the between-country comparisons that can be made, here showing the across-country spatial inequality for both Kenya and Malawi.

However, there are a number of aspects that we aim to improve upon in future work. An adjustment for HIV epidemics is crucial, given the extent of the epidemic in Kenya (and in many other countries), and we would like to acknowledge the uncertainty in the bias correction, and also obtain corrections at a finer
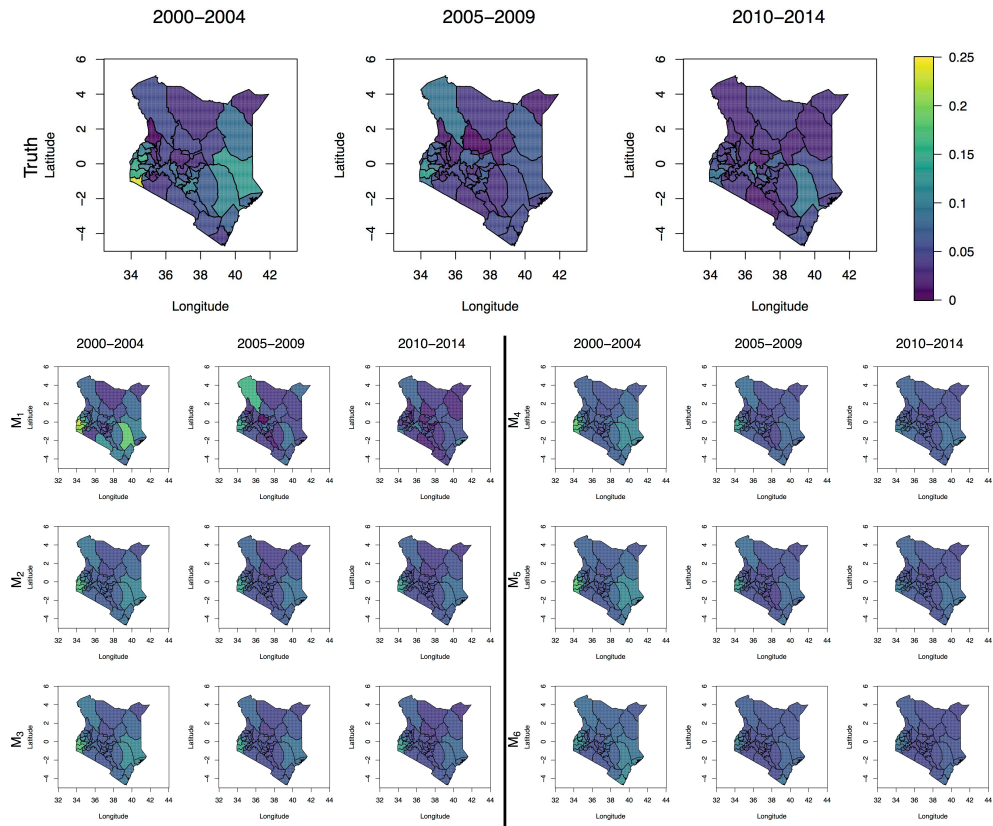
Figure 12. Regional predicted U5MR. Top row is the "truth", i.e., direct estimates based on the 799 test locations in the 2014 survey. Model M_1 are the direct estimates based on the other clusters. Model M_2 is the fixed spatial only model (no covariates). Model M_3 is the fixed spatial model with covariates (access, precipitation and *Pf*PR). Model M_4 is the time-varying spatial only model (no covariates). Model M_5 is the time-varying spatial only model with covariates (access, temperature, and *Pf*PR). Finally, model M_6 is the only covariates model (aridity, temperature, precipitation, and *Pf*PR).

geographical granularity. A source of potential bias that we want to investigate is migration, since earlier births in particular may have occurred at different locations to those at which the survey was carried out.

The age pattern of human mortality between ages 0 and 5 years follows a regular, decreasing pattern across a wide range of overall levels. Net of level, this age pattern can be characterized by the ratio of mortality at each age compared to a reference age. Our model has six age-specific intercepts and a random walk of second order to model the time trend, with one each for [0,1) and [1,12) months and then a third for the remaining period of [12,60) months. An alternative might be to use a Lee-Carter (Lee and Carter 1992) like approach that includes one component representing the average shape of the

age profile, and a second component representing a time varying trend, which is multiplied by the age-specific mortality change from the average age profile. For applications of this model in a similar context, see for example Sharrow et al. (2014) and Alexander et al. (2017). One computational challenge of the Lee-Carter approach is the multiplicative structure of two random effects which is hard to incorporate into INLA.

Our models estimate mortality in six independent age groups, and it is possible that the age pattern that results from combining the estimates from the six models does not follow any of the regularly observed age patterns of human child mortality. In our analysis, this was not a problem (see Supplementary Materials), but we are currently working on a flexible model of the age pattern of mortality that can enforce this constraint.

We would like to include other data sources, for both Kenya and other countries. Early DHS do not contain the GPS coordinates of the sampled clusters, but rather the administrative areas within which sampling took place. We plan to extend methods presented in Wilson and Wakefield (2017) to model the location of the unknown sampling point. As described in Section 1, we have utilized so-called full birth history data in which the births and deaths of each child are available. Summary birth history consist of only the number of children ever born and the number who died, by age of mother. These data are easier to collect and are available in a large number of surveys and censuses. The incorporation of such data into a model-based framework is a priority for future work.

In this work we have used a continuous spatial model, whereas our major interest was to inspect results on the discrete scale for the 47 administrative regions. To this end, we may view the continuous spatial prior as a means by which we induce a prior for the collection of 47 discrete areas. Our model can produce estimates at much finer geographical scales, but an important question is how reliable would estimates be at such scales? One way of answering this is by comparison with direct estimates, but at a fine scale, such estimates have large variance. Without such a comparison, using estimates at a fine scale is inherently hazardous, and we would only carry out such an endeavor in exploratory analyses.

To produce estimates at the 47 area level, we integrated over the spatial field and included the population density to produce the results at the county level. An obvious question that arises is: what advantages are there with this approach as compared to using a discrete spatial model, such as the ICAR model (Besag et al. 1991), directly? One advantage of the continuous model is that we get a smoothed estimated field giving an indication of the U5MR at a finer resolution, though as just pointed out, caution in such surfaces is required. Other advantages of the continuous spatial model are the ability to avoid ecological bias when modeling covariates, and its ability to naturally incorporate data measured at different spatial resolutions, in particular the model can account for boundary changes in a very clean way.

Another advantage of our model is that when using a continuous random field we do not need to specify a neighborhood structure. The 47 administrative regions of Kenya vary widely in shape and size, and therefore in the number of neighbors, so that it is not clear how to define a sensible neighborhood dependence structure. Part of our future research will continue to investigate how discrete spatial models would perform in this setting. In this context, we are particularly interested in the performance of the recently proposed model of Riebler et al. (2016) and in a comparison of the results to the continuous model presented here. It would also be interesting to compare the spatial model that we have developed with other possibilities including lattice kriging (Nychka et al. 2015), fixed rank kriging (Cressie and Johannesson 2008) and predictive processes (Banerjee et al. 2008). See Bradley et al. (2016) and Heaton

et al. (2017) for recent reviews and comparisons of these approaches (and others), with an emphasis on big data. Multiscale models have also been recently proposed (Fonseca and Ferreira 2017).

There are several limitations to the covariate modeling carried out in Section 4. For one, we use geographically-referenced covariates rather than household or individual level variables since we were interested in predicting U5MR at locations without outcome data, and so we restricted ourselves to covariates that were available at the pixel level for the whole of Kenya. Therefore, we do not directly model several variables that are known to have an impact on childhood mortality such as characteristics of the child/birth (e.g., gender, single versus multiple birth, birth order) maternal demographics (e.g., age, education), biological factors (e.g., vaccination rates, disease prevalence), and household characteristics (e.g., toilet facilities, access to water). We also assumed a common covariate model for all age hazards when there is evidence (Balk et al. 2004) that both the covariates and the strengths of the association depend on the age of the child. In future work, we will carry out individual-level modeling and refine the model. It will not be possible to use such a model for prediction, but it will be of great interest to see if spatial characteristics can improve on a model that includes child, mother and household variables, for different ages.

Though spatial surfaces do exist for some of the above variables (e.g., measles vaccination coverage: Takahashi et al. 2017) or surfaces could be developed based on DHS data (Gething et al. 2015), there is greater uncertainty associated with these variables, which can lead to misleading inference (Foster et al. 2012). We therefore limited the number of heavily-modeled covariates in our model. Additionally, many of these factors are associated with variables already included in our model.

The computations were run on a computing server with 32 Intel Xeon 2.7 GHz CPUs available. The full Bayesian model required around 14 hours for estimation and 19.5 hours for predictions. An empirical Bayes version of the model required around 2.5 hours for estimation and 10 hours for predictions. Code to run the models described here can be found at `http://faculty.washington.edu/jonno/software.html`.

Acknowledgements

References

Alexander, M., E. Zagheni, and M. Barbieri (2017). A flexible Bayesian model for estimating subnational mortality. *Demography 54*(6), 2025–2041.

Alkema, L. and J. New (2014). Global estimation of child mortality using a Bayesian B-spline bias-reduction model. *The Annals of Applied Statistics 8*, 2122–2149.

Alkema, L., J. R. New, J. Pedersen, D. You, et al. (2014). Child mortality estimation 2013: an overview of updates in estimation methods by the United Nations Inter-Agency Group for Child Mortality Estimation. *PLoS One 9*, e101112.

Allison, P. (2014). *Event History and Survival Analysis, Second Edition*, Volume 46. SAGE publications.

Balk, D., T. Pullum, A. Storeygard, F. Greenwell, and M. Neuman (2004). A spatial analysis of childhood mortality in West Africa. *Population, Space and Place 10*, 175–216.

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

Barasa, E., A. Manyara, S. Molyneux, and B. Tsofa (2017). Recentralization within decentralization: county hospital autonomy under devolution in kenya. *PloS One 12*, e0182440.

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

Bhatt, S., E. Cameron, S. Flaxman, D. Weiss, D. Smith, and P. Gething (2017). Improved prediction accuracy for disease risk mapping using Gaussian process stacked generalization. *Journal of The Royal Society Interface 14*, 20170520.

Bhatt, S., D. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, K. Battle, C. Moyes, A. Henry, P. Eckhoff, et al. (2015). The effect of malaria control on plasmodium falciparum in Africa between 2000 and 2015. *Nature 526*, 207–211.

Blangiardo, M. and M. Cameletti (2015). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley and Sons.

Bradley, J. R., N. Cressie, T. Shi, et al. (2016). A comparison of spatial predictors when datasets could be very large. *Statistics Surveys 10*, 100–131.

Burke, M., S. Heft-Neal, and E. Bendavid (2016). Sources of variation in under-5 mortality across sub-Saharan Africa: a spatial analysis. *The Lancet Global Health 4*, e936–e945.

Chen, C., J. Wakefield, and T. Lumley (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 P. Lloyd (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 G. Johannesson (2008). Fixed rank kriging for very large spatial data sets. *Journal of the Royal Statistical Society: Series B 70*, 209–226.

Didan, K. (2015). MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006. NASA EOSDIS Land Processes DAAC. https://doi.org/10.5067/modis/mod13a3.006.

Dwyer-Lindgren, L., F. Kakungu, P. Hangoma, M. Ng, H. Wang, A. D. Flaxman, F. Masiye, and E. Gakidou (2014). Estimation of district-level under-5 mortality in Zambia using birth history data, 1980–2010. *Spatial and Spatio-Temporal Epidemiology 11*, 89–107.

Fick, S. E. and R. J. Hijmans (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology 37*, 4302–4315.

Fonseca, T. C. and M. A. Ferreira (2017). Dynamic multiscale spatiotemporal models for Poisson data. *Journal of the American Statistical Association 112*, 215–234.

Foster, S., H. Shimadzu, and R. Darnell (2012). Uncertainty in spatially predicted covariates: is it ignorable? *Journal of the Royal Statistical Society: Series C 61*, 637–652.

Fuglstad, G.-A., D. Simpson, F. Lindgren, and H. Rue (2018). Constructing priors that penalize the complexity of Gaussian random fields. *Journal of the American Statistical Association*. In press.

GBD 2016 Mortality Collaborators (2017). Global, regional, and national under-5 mortality, adult mortality, age-specific mortality, and life expectancy, 1970–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet 390*, 1084–1150.

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

Golding, N., R. Burstein, J. Longbottom, A. Browne, N. Fullman, A. Osgood-Zimmerman, L. Earl, S. Bhatt, E. Cameron, D. Casey, L. Dwyer-Lindgren, T. Farag, A. Flaxman, M. Fraser, P. Gething, H. Gibson, N. Graetz, L. Krause, X. Kulikoff, S. Lim, B. Mappin, C. Morozoff, R. Reiner, A. Sligar, D. Smith, H. Wang, D. Weiss, C. Murray, C. Moyes, and S. Hay (2017). Mapping under-5 and neonatal mortality in Africa, 2000–15: a baseline analysis for the Sustainable Development Goals. *The Lancet*. Available online, September 25th, 2017.

Hallett, T., S.-J. Anderson, C. A. Asante, N. Bartlett, V. Bendaud, S. Bhatt, C. Burgert, D. F. Cuadros, J. Dzangare, D. Fecht, et al. (2016). Evaluation of geospatial methods to generate subnational HIV prevalence estimates for local level planning. *AIDS 30*, 1467–1474.

Heaton, M. J., A. Datta, A. Finley, R. Furrer, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, F. Lindgren, et al. (2017). Methods for analyzing large spatial data: A review and comparison. *arXiv preprint arXiv:1710.05013*.

Held, L., B. Schrödle, and H. Rue (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. In T. Kneib and G. Tutz (Eds.), *Statistical Modeling and Regression Structures – Festschrift in Honour of Ludwig Fahrmeir*, pp. 91–110. Physica-Verlag.

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

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine 19*, 2555–2567.

Larmarange, J. and V. Bendaud (2014). HIV estimates at second subnational level from national population-based surveys. *AIDS 28*, S469–S476.

Lee, R. and L. Carter (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association 87*, 659–671.

Leroux, B., X. Lei, and N. Breslow (1999). Estimation of disease rates in small areas: A new mixed model for spatial dependence. In M. Halloran and D. Berry (Eds.), *Statistical Models in Epidemiology, the Environment and Clinical Trials*, pp. 179–192. New York: Springer.

Linard, C., M. Gilbert, R. W. Snow, A. M. Noor, and A. J. Tatem (2012). Population distribution, settlement patterns and accessibility across africa in 2010. *PloS One 7*, e31743.

Lindgren, F., H. Rue, and J. Lindström (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.

Liu, L., S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black (2017). Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals. *The Lancet 388*, 3027–3035.

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

Mosley, W. H. and L. C. Chen (1984). An analytical framework for the study of child survival in developing countries. *Population and Development Review 10*, 25–45.

Nelson, A. (2008). Estimated travel time to the nearest city of 50,000 or more people in year 2000. Technical report, Global Environment Monitoring Unit - Joint Research Centre of the European Commission, Ispra, Italy.

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

Pesaresi, M., D. Ehrlich, S. Ferri, A. Florczyk, S. Freire, M. Halkia, A. Julea, T. Kemper, P. Soille, and V. Syrris (2016). Operating procedure for the production of the global human settlement layer from landsat data of the epochs 1975, 1990, 2000, and 2014. *Publ. Off. Eur. Union*.

Pezzulo, C., E. Utazi, T. B. A. Sorichetta, A. Tatem, J. Yourkavitch, T. Pullum, and C. Burgert-Brucker (2017). Subnational modelling of child mortality and its drivers across 27 countries in Sub-Saharan Africa. Technical report, Paper presented at PAA Meeting.

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

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

Riebler, A., S. H. Sørbye, D. Simpson, and H. Rue (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Statistical Methods in Medical Research 25*(4), 1145–1165.

Root, G. (1997). Population density and spatial differentials in child mortality in zimbabwe. *Social Science & Medicine 44*, 413–421.

Root, G. P. (1999). Disease environments and subnational patterns of under-five mortality in sub-saharan africa. *Population, Space and Place 5*, 117–132.

Rue, H., S. Martino, and N. Chopin (2009). 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.

Sharrow, D. J., S. J. Clark, and A. E. Raftery (2014). Modeling age-specific mortality for countries with generalized hiv epidemics. *PloS One 9*, e96447.

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

Skinner, C. and J. Wakefield (2017). Introduction to the design and analysis of complex survey data. *Statistical Science 32*, 165–175.

Spiegelhalter, D., L. Freedman, and M. Parmar (1994). Bayesian approaches to randomized trials (with discussion). *Journal of the Royal Statistical Society, Series A 157*, 357–416.

Takahashi, S., C. J. E. Metcalf, M. J. Ferrari, A. Tatem, and J. Lessler (2017). The geography of measles vaccination in the african great lakes region. *Nature Communications 8*.

Tottrup, C., B. Tersbol, W. Lindeboom, and D. Meyrowitsch (2009). Putting child mortality on a map: towards an understanding of inequity in health. *Tropical Medicine & International Health 14*(6), 653–662.

Van der Laan, M. J., E. C. Polley, and A. E. Hubbard (2007). Super learner. *Statistical Applications in Genetics and Molecular Biology 6*.

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

Wakefield, J. (2008). Ecologic studies revisited. *Annual Review of Public Health 29*, 75–90.

Wakefield, J., D. Simpson, and J. Godwin (2016). Comment: Getting into space with a weight problem. Discussion of, "Model-based geostatistics for prevalence mapping in low-resource settings", by P.J. Diggle and E. Giorgi. *Journal of the American Statistical Association 111*, 1111–1119.

Walker, N., K. Hill, and F. Zhao (2012). Child mortality estimation: methods used to adjust for bias due to AIDS in estimating trends in under-five mortality. *PLoS Med 9*, e1001298.

Watanabe, S. (2013). A widely applicable Bayesian information criterion. *Journal of Machine Learning Research 14*, 867–897.

Watjou, K., C. Faes, A. Lawson, R. Kirby, M. Aregay, R. Carroll, and Y. Vandendijck (2017). Spatial small area smoothing models for handling survey data with nonresponse. *Statistics in medicine 36*(23), 3708–3745.

Wilson, K. and J. Wakefield (2017). Pointless continuous spatial surface reconstruction. *arXiv:1709.09659*.

Wolpert, D. (1992). Stacked generalization. *Neural Networks 5*, 241–259.

WorldPop (2017). Kenya 1km births, version 2. Technical report, University of Southampton, DOI: 10.5258/SOTON/WP00349.

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

Zomer, R. J., D. A. Bossio, A. Trabucco, L. Yuanjie, D. C. Gupta, and V. P. Singh (2007). Trees and water: smallholder agroforestry on irrigated lands in Northern India. Technical report, International Water Management Institute.

Zomer, R. J., A. Trabucco, D. A. Bossio, and L. V. Verchot (2008). Climate change mitigation: A spatial analysis of global land suitability for clean development mechanism afforestation and reforestation. *Agriculture, Ecosystems and Environment 126*, 67–80.